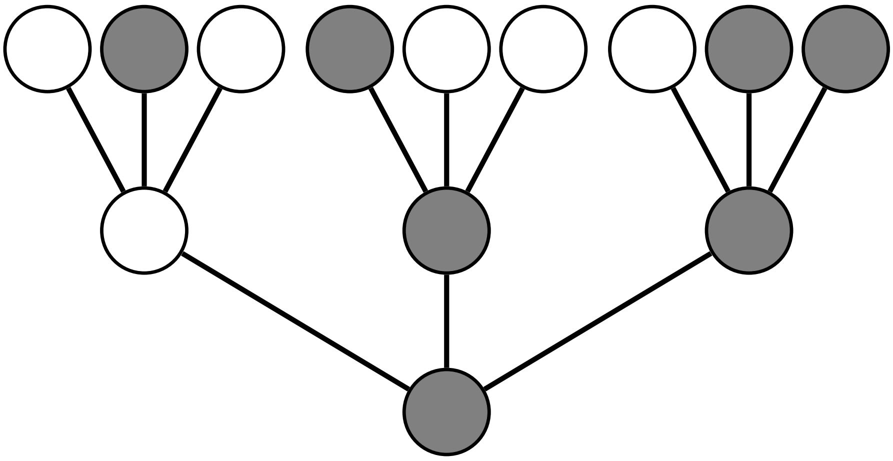
Ancestral Lines in Deterministic and Stochastic Mutation-Selection Models

---

Dissertation

zur Erlangung des akademischen Grades
Doktor der Mathematik (Dr. math.)

vorgelegt an der Fakultät für Mathematik der Universität Bielefeld



eingereicht von

M. Sc. Sebastian J. Hummel

im Mai 2019

Abstract

Mutation-selection models are traditionally either deterministic or stochastic. Deterministic approaches assume that the size of the population is such that a law of large numbers applies so that random fluctuations may be neglected. The resulting models are (ordinary or partial) differential equations or (discrete-time) dynamical systems, which describe the evolution in the usual forward direction of time. In contrast, stochastic approaches take into account the fluctuations due to random reproduction; the resulting stochastic processes have a firm place in probability theory. Here, the corresponding ancestral processes, which describe the ancestry of a sample of individuals from a population at the present, play an eminent role in the analysis. Deterministic models of population genetics and their stochastic counterparts have largely led separate lives. It is the purpose of this thesis to bring these two areas of research closer together by extending the backward point of view, so far reserved for stochastic models of population genetics, to deterministic mutation-selection equations. The corresponding ancestral processes describe the history of a finite sample of individuals and remain random; although the type-frequency process of the entire population evolves deterministically. Tailored versions of the genealogical processes yield stochastic representations of the solutions of the deterministic equations. The analysis sheds new light on the deterministic dynamic and its long-term behaviour. Special emphasis is placed on the connection between bifurcation phenomena and ancestral structures. The genealogical approach allows the notion of a (random) ancestral type also in the deterministic setting and provides the framework to determine its distribution. We illustrate the underlying ideas by applying them to a special case of frequency-dependent selection. The ancestral processes for such models are largely unexplored territory. We first establish appropriate structures and then make them tractable by applying our aforementioned concepts. The tailored processes allow an explanation of the richer bifurcation structure by genealogical means and lead to expressions for the ancestral type distribution.

Acknowledgement

I am extremely grateful to Ellen Baake for supervising my doctorate. Her guidance, encouragement, and continuous support have been essential. My deepest gratitude extends to Fernando Cordero who patiently shared his knowledge and many ideas with me. It has been wonderful he has shared the office with me throughout my PhD studies. I had the great pleasure to participate in many fruitful discussions with Anton Wakolbinger. Furthermore, he provided helpful comments on Chapter 3 and pointed out the proof of Lemma 4.45. I am grateful to Jan Swart for helpful discussions and many insightful comments that improved the presentation of this thesis. I very much appreciate discussing parts of this thesis with Ute Lenz and Anja Sturm. I thank Nina Gantert for supervising my Bachelor’s and Master’s thesis and for introducing me to probability theory. It has been a joyful and enriching experience to discuss mathematics with Adrián González Casanova and Emmanuel Schertzer. Many thanks to Mareike Esser for sharing her thesis-template with me. I thank Frederic Alberti, Fernando Cordero, Luigi Esercito, Sebastian Probst, and Patrik Hummel for proofreading parts of this thesis. All remaining errors are mine. I thankfully acknowledge the financial support from the SPP-1590 ‘Probabilistic Structures in Evolution’ and the SFB 1283 ‘Taming uncertainty and profiting from randomness and low regularity in analysis, stochastics and their applications’. I am grateful to all people associated with the SPP-1590 for many inspiring meetings and workshops. I thank all current and former members of the ‘Biomathematics and Theoretical Bioinformatics’ group for the pleasant working atmosphere and the great company in the coffee breaks. I am indebted to my friends for the distractions and joys they have provided throughout.

I am deeply grateful to Nairy, Günter, Chiara, Timon, Felix, Patrik, and Gudrun for their encouragement, support, and everything they have taught me.

Accepter l'absurdité de tout ce qui nous entoure est une étape, une expérience nécessaire: ce ne doit pas devenir une impasse. Elle suscite une révolte qui peut devenir féconde.

Albert Camus (Interview, *Les Nouvelles littéraires*, No. 954, 11/1945)

Contents

List of Figures

Abbreviations

ASG ancestral selection graph

IVP initial value problem

ODE ordinary differential equation

pLD-ASG pruned lookdown ancestral selection graph

smss strong mutation–strong selection, i.e. mutation and selection rates are independent of the population size

smsssi strong mutation–strong selection–strong interaction, i.e. mutation, selection, and interaction rates are independent of the population size

1

Introduction

Models of population genetics provide a unified mathematical framework for Darwinian evolution and Mendelian inheritance. The former means that heritable traits that increase reproductive success spread in a population [Dar59]. The latter explains the preservation of variation and inheritance of the traits [Men66]. Evolutionary forces captured by classical population models include mutation, selection, recombination, random variation, and migration. How these forces shape a population is one of the fundamental questions in evolutionary biology. Depending on the kind of questions to be addressed, usually only a subset of these forces is combined into a model.

The models that allow the study of the interplay between mutation and selection form a major subgroup. By and large, the field is divided into two major lines of research, devoted to *deterministic* and *stochastic* models, respectively. Deterministic mutation-selection equations describe the action of mutation and selection on the genetic composition of an effectively infinite population. The first version goes back to Crow and Kimura [CK56]. Deterministic mutation-selection equations are formulated in terms of discrete- or continuous-time dynamical systems, and they are treated forward in time throughout, via the well-developed methods of dynamical systems; a comprehensive overview of the research is provided in the monograph by Bürger [Bür00]. Stochastic mutation-selection models additionally capture the fluctuations due to random reproduction over long time scales. These fluctuations are absent in the deterministic dynamics. The stochastic models have their roots in the seminal work of Fisher [Fis30], Wright [Wri31], Malécot [Mal48], Feller [Fel51], and Moran [Mor58]. Overviews of the area can be found in the monographs by Ewens [Ewe04], Durrett [Dur08], and Etheridge [Eth11]. It is the purpose of this thesis to bring these two areas of research closer together by extending the backward point of view, so far reserved for stochastic models of population genetics, to deterministic mutation-selection equations.

Background

The *Moran model* is one of the classic continuous-time stochastic mutation-selection models [Mor58]. It describes the evolution of a finite panmictic population subject to *mutation*, *selection*, and *genetic drift*. (A panmictic population is a well-mixed population in which all individuals are potential partners. Genetic drift is common terminology for random variation; not to be confused with the drift of a stochastic process.) Reproduction events are always coupled with death events so that the population size remains constant. In the simplest form, individuals are haploid, which means that each individual carries only a single copy of its genetic information, and the population is subdivided into only two types.

In the finite setting of the Moran model, it is possible to represent the entire population and their interactions via a random graph in the spirit of the celebrated *graphical representation* that goes back to Harris [Har78]. This random graph is constructed in a Poissonian manner, which allows the use both for the forward and the backward direction of time. In the forward direction, time evolves with the population, whereas in the backward direction, time evolves with the population's ancestry. Historically, most of the analyses have been carried out for time evolving in the forward direction [CK56; Mor58]. The first one to rigorously introduce a genealogical perspective was Kingman [Kin82a; Kin82b; Kin82c] in his seminal work. He analyses a very large population in terms of the ancestral structure of a sample of individuals. This ancestral structure is tree-like and is called the *coalescent*. Since its introduction this approach gained great popularity because it is easy to simulate [GT94] and has a natural connection to applications. There one usually deals with a sample of the population and tries to infer its formation. Thus, it is a major advantage of the genealogical perspective to naturally encapsulate the relevant information of the ancestral structure of a given sample. Wakeley [Wak09] provides a general overview.

Unfortunately, the derivation of explicit results for finite population models becomes cumbersome for more complicated dynamics. This is why one classically considers a large population limit in an appropriate time- and parameter rescaling, which is usually more tractable. Models that differ only in the details of their evolutionary mechanisms often lead to the same limit models. Depending on the scaling, the limit processes are either stochastic or deterministic. Stochastic models still comprise random fluctuations in the type distribution that are absent in the deterministic models. Stochastic models are usually characterised via a *stochastic differential equation* or the infinitesimal generator of the transition semigroup of a Markov process. Deterministic models are characterised via an *ordinary differential equation* (ODE). For the Moran model with mutation and selection, arguably, the most classical large population limit is the one in which time and parameters are both appropriately rescaled to obtain a diffusion process. This process is then a *Wright-Fisher diffusion* with mutation and selection, and various explicit results are available. It is an appropriate approximation if mutation and selection are weak, which means that both are rare compared to the population size. Another large population limit arises if neither parameters nor time are rescaled. This leads to the classical deterministic *mutation-selection equation* [CK56]. This limit is suitable in a *strong mutation–strong selection* framework, which means that both forces are independent of the population size.

The graphical representation does not straightforwardly carry over to these large population limits. This makes it difficult to properly define an ancestral picture directly in the limits. The most notable finite population particle representations that carry over to representations for a wide class of appropriately scaled large population limits (e. g. the Wright-Fisher diffusion) are the *lookdown-constructions* of Donnelly and Kurtz [DK99b]. A crucial idea in these constructions is the ordering of the individuals according to their persistence in the population.

A particle representation and, in particular, the inherent retrospective view naturally give rise to processes that are dual to the respective forward process. Loosely speaking, a *dual* *process* is a process that can be used to express certain functions of the process to which it is dual. A *duality relation* establishes the formal connection. Ideally, the dual process is simpler so that the functions of the original process are determined more easily. Jansen and Kurt [JK14] provide an overview of the notions of duality for Markov processes.

The most classical of such relations in the context of population genetics is between the neutral (i. e. only subject to genetic drift) Wright-Fisher diffusion and the line-counting process of Kingman's coalescent. There are dual processes suited for more complex forward dynamics. In the diffusion limit of the Moran model with (frequency-independent) selection mechanism, the *ancestral selection graph* (ASG) augments Kingman's coalescent with virtual lineages [KN97; NK97], which correspond to lineages of potential ancestors. The line-counting process of the ASG is dual to the Wright-Fisher diffusion with selection. A priori, the ASG is untyped. This means that, since the types in the sample from which we construct the potential genealogy are unknown, the types along the lines in the ASG are unknown as well. The ASG can be used to estimate relevant quantities like sampling probabilities [KN97, Sect. 5], time to a common ancestor of a sample [KN97, Sect. 3], or time to fixation [Man09, Sect. 5](see also [PP13]). A similar duality arises if one conditions on the types in the sample and then traces back the sample in a typed way [DG14; EG09; Shi81].

We will see that some suitably tailored ancestral processes for the Moran model with mutation and selection actually carry over also to the large population limit that leads to the mutation-selection equations. They remain random even in the deterministic limit and hence lead to stochastic representations of the solutions of the deterministic equations. In particular, they shed new light on the deterministic dynamics and its long-term behaviour.

The Moran dynamic without mutation leads to a population that contains only a single type after some finite time. The *fixation probability* of a type is the probability that this type is the one to remain. In contrast, if mutation acts on the population, several types may coexist in the long run. In this case, the *stationary type distribution* reasonably describes the population. If in addition selection favours one type, both the fixation probability and the stationary type distribution will differ from the neutral case (i. e. without mutation and selection). Going backwards in the graphical representation naturally provides a way to characterise fixation probabilities in terms of the model-inherent genealogies [Cor17a, Lem. 4.5]. It turns out that this approach also leads to a representation of the stationary type distribution. The *equilibria* of the deterministic mutation-selection equations correspond to the stationary type distributions. The long-term behaviour of the type distribution deterministically depends on the initial type frequency in the population. For certain parameter values of the mutation-selection equation, a variation of the parameters dramatically changes the qualitative structure of the long-term behaviour of the type distributions. These parameter values correspond to the *bifurcation points* of this dynamical system.

Given the model parameters, the retrospective view allows not only to predict the type distribution in the population in the future, but also the population's progeny structure. The type of the ancestor in the past of an individual sampled at present is called the individual's *ancestral type*. Note that, due to mutations, descendant and ancestor are not necessarily of the same type. The distribution of the ancestral type at a fixed time in the past is called the *ancestral type distribution* at that time. Under the dynamics of the Moran model, all individuals share a single *common ancestor* [Cor17a, Lem. 4.3]. Fearnhead [Fea02, Thm. 3] was the first to rigorously analyse the distribution of the type of this common ancestor in a population at equilibrium in the diffusion limit with mutation and selection. His approach relies on the ideas of the virtual lineages in the ASG. It leads to a representation of the common ancestor type distribution in terms of a power series, whose coefficients satisfy a recursion that by now is classical and goes under the name *Fearnhead-recursion.* Taylor [Tay07] generalised Fearnhead's result to other forms of selection. His approach is based on the structured coalescent, which essentially codes the (typed) coalescent along with the random background in which it evolves. It was introduced by Kaplan et al. [KDH88] and rigorously defined by Barton et al. [BES04]. Kluth et al. [KHB13, Thm. 2] obtained the Fearnhead-recursion for the finite Moran model with selection and mutation (see also Cordero [Cor17a, Prop. 4.7]). Lenz et al. [Len+15, Prop. 6] complemented the analysis by providing a probabilistic interpretation of the recursion in terms of tail probabilities of a pruned ASG that is constructed in a 'lookdown'-manner. Their ideas lead to the *pruned lookdown ancestral selection graph* (pLD-ASG).

In the Moran model with mutation and selection, suitable ancestral processes lead to expressions for fixation probabilities, the stationary type distribution, and the ancestral type distribution. We will first identify these processes and then exploit that the processes and their relation to the forward model translate to the large population limits. Hence, they can be used for the analysis of the limit models.

In simple population models, selection is usually modelled as being independent of the type frequencies in the population [Mor58]. Yet, in many biological systems a type's frequency influences its reproductive success [AC74]. Let us mention Müllerian mimicry as a classical example for *frequency-dependent selection.* It describes the mimicry of warning signals (like the yellow-black pattern in bees) of prey that is in fact distasteful to the predator [Mül78; Mül79]. The more frequent the pattern in the population, the faster the predator learns to avoid the prey that possesses the warning signal, and the safer the prey is from the predator.

A special case of frequency-dependent selection is *pairwise interaction* where the types' fitnesses depend on the type of a uniformly chosen individual in the population (see Chapters 2 and 4 for details). In the diffusive and deterministic setting, classical diffusion theory covers also models with interactive selection and it leads to expressions for fixation probabilities and stationary distributions [Eth11, Ch. 3]. In a general framework, for example in the case of jump-diffusion processes, the classical tools from diffusion theory are not available. In special cases the analysis of such models is still possible via the genealogical perspective [Fou13; GS18]. But general frequency-dependent selection is not covered by current coalescent theory. In particular, general selection models so far resist an analysis. One might hope that this will change once appropriate structures are identified.

Outline

In this thesis, we work out and compare similarities in the analyses of stochastic and deterministic population models in terms of the backward point of view. Our starting point is the Moran model with mutation, selection, and pairwise interaction. In the non-interactive case, we mainly rely on the ideas of Krone and Neuhauser [KN97] to characterise the type distribution in terms of the number of lines in an ASG that we *prune* or *kill* upon mutations and that we call *killed ASG.* The type distribution of the forward process is connected to the ancestral process via duality. This also provides the link between the long-term behaviour of the two processes. We use a concept of Lenz et al. [Len+15] to characterise the ancestral type distribution. Their idea to *order* and *prune* lines in the ancestral process yields a representation for the ancestral type distribution.

The concepts developed for the (non-interactive) Moran model translate to the strong mutation–strong selection (smss) limit. In particular, they allow a probabilistic interpretation of the (deterministic) smss-limit via the (still stochastic) ancestral picture. Also in this setting, the type distribution of the forward process in the distant future is connected to the long-term behaviour of this ancestral process. Our probabilistic approach reveals the genealogical structure behind the bifurcation phenomena of the ODE. The ancestral structure allows the notion of an ancestral type also in the smss-framework. We characterise its distribution in terms of a pLD-ASG in this setting.

We illustrate the flexibility of the underlying ideas by applying them to the case of strong pairwise interaction, which is usually difficult to treat by genealogical means. Also in this framework, we equip the mutation-selection equation with pairwise interaction (or equivalently the diploid mutation-selection equation; details in Chapter 4) with a probabilistic interpretation. This leads to a *stratified ASG*. This ancestral process has a tree structure, but is not directly susceptible to treatment with tools from the theory of branching processes. Nevertheless, we can modify the ideas of the classic results to our setup and establish results from the frequency-independent selection case in this more general setting.

At this point, let us mention related work of Mach et al. [MSS18b], who study a large class of ODEs that arise as a mean-field limit of (stochastic) interacting particle systems on the complete graph. Their study relies crucially on recursive tree processes. As an example Mach et al. [MSS18b] (see also [Mac17]) treat the cooperative branching model with deaths, which corresponds to a special case of our mutation-selection model with pairwise interaction. The analyses via recursive tree processes and our ancestral process provide complementary insight into the behaviour of the model. We will spell out the similarities and differences of the two approaches in the course of Chapter 4.

The structure of the thesis is as follows. We lay down the foundations in Chapter 2 and recapitulate the *Moran model with selection, mutation, and pairwise interaction*. We start out by recalling the graphical representation and classic results of the prospective approach like expressions for the stationary distribution and absorption probabilities. We then move on to treat the retrospective approach. First, we state the ASG in this general setup. In the remainder of the chapter, we restrict ourselves to selection that is frequency-*independent*. We reobtain the type distribution of the Moran model in terms of a killed ASG. A duality relation provides the formal link. At last, we recapitulate the construction of the pLD-ASG for finite populations of Cordero [Cor17a] and derive the Fearnhead-recursion in this setting. A duality relation between the pLD-ASG and a Markov chain which absorbs at the boundaries leads to a connection between the absorption probabilities of that Markov chain and the common ancestor type distribution.

Chapter 3 is devoted to the *smss-limit of the Moran model*. Throughout the chapter, we treat only the case of frequency-independent selection. Again, we start out by summarising known results that are derived via the forward direction of time. Switching to the backward perspective enables us to work along the lines of Chapter 2. A duality relation connects a killed ASG in the smss-limit with the smss-limit of the Moran model. We express the type distribution of the forward picture via the killed ASG and in this way relate the long-term behaviour in the two perspectives. The framework allows the notion of an ancestral type. A pLD-ASG leads to a characterisation of its distribution. We establish a duality relation between the pLD-ASG and a piecewise-deterministic Markov process that absorbs at the boundaries. This leads to a connection between the ancestral type distribution and the absorption probabilities of that piecewise-deterministic Markov process.

Chapter 4 treats the *strong mutation–strong selection–strong interaction* (smsssi) *limit of the Moran model.* The general procedures and techniques from Chapter 2 (and Ch. 3) are not straightforwardly applicable because the ancestral processes in Chapter 2 (and Ch. 3) are only defined for the case without interaction. Hence, we first adapt the backward processes to this new set up. Reducing, pruning, and stratifying the ASG leads to the aforementioned stratified ASG. Second, we recover the type composition given by the mutation-selection equation with interaction in terms of this backward process. At last, we use a sequence of stratified ASGs, which we call forest of stratified ASGs, to characterise the ancestral type distribution.

In Section 5, we summarise the thesis and place it into the context of current research in mathematical population genetics. We also point out directions for research that is beyond the scope of this thesis.

Chapters 3 and 4 are based on [BCH18a] and [BCH18b]. Both articles are joint work with Prof. Dr. Ellen Baake and Dr. Fernando Cordero.

2 Moran model – forward and backward in time

The Moran model is one of the most classic continuous-time stochastic models for the evolution of a finite population. Our setup is for haploid individuals subject to genetic drift, mutation, selection, and pairwise interaction, which is a special case of frequency-dependent selection.

The structure of the present chapter is as follows. We start with the description of the graphical representation and recapitulate the main properties, fixation probabilities, and the stationary type distribution of the Moran model in Section 2.1 from the classical forward point of view. For the sake of completeness, we provide the classic proofs of these results. From Section 2.2 onwards, we consider the backward direction of time. The tools and techniques developed in the subsequent sections are the basis for the analysis of the deterministic-mutation selection equations, which arise as the large population limit of the Moran model and which we consider later in Chapters 3 and 4. We start out by recapitulating the construction of the ancestral selection graph (ASG) in the general setup. In the remainder of the chapter, i.e. from Section 2.3 onwards, we consider *only* the case without interaction. It turns out that in this setting only the number of lines in the ASG is relevant for our purposes. Mutations lead to further reductions of the ASG. All these simplifications give rise to the process we call the *killed ASG*. We connect its line-counting process to the Moran model via a duality relation (Section 2.3). This allows a representation of the factorial moments of the number of unfit individuals in the Moran model in terms of the killed ASG. We recover fixation probabilities and the stationary distribution by genealogical means. In the Moran model, all individuals at present share a common ancestor in the sufficiently distant past [Cor17a, Lem. 4.3]. In the diffusion limit of the Moran model, the classic representation of the common ancestor type distribution at equilibrium in terms of a series goes back to Fearnhead [Fea02, Thm. 3]. The series coefficients satisfy the so-called Fearnhead-recursion. Lenz et al. [Len+15] introduced an ancestral structure, which is called the pruned lookdown ASG (pLD-ASG), to provide a probabilistic interpretation of this recursion, which was originally obtained by analytic means. Cordero [Cor17a, Sect. 4] adapted the construction to the finite Moran model and its smss-limit and obtains a characterisation of the common ancestor type distribution in this setting. In Section 2.4, we recapitulate his construction and reobtain Fearnhead's recursion. Furthermore, we derive properties of the common ancestor type distribution at equilibrium and extend previous results to the finite time horizon. A duality relation between the line-counting process of the pLD-ASG and a Markov chain which absorbs at the boundaries complements the analysis.

2.1 Moran model with mutation, selection, and pairwise interaction

The two-type Moran model with mutation, selection, and pairwise interaction describes the evolution of a haploid population of a finite number N of individuals in continuous time. Each individual has a type, which is either 0 or 1. We refer to type 0 as the *fit* or *beneficial* type, whereas type 1 is *unfit* or *deleterious*. When an individual reproduces, its single offspring inherits the parent's type and replaces a uniformly chosen individual, thereby keeping the population size constant. All individuals reproduce at rate 1. The selective advantage of type 0 is reflected by a larger reproduction rate. It has two contributions: a part that is independent of the current type distribution and occurs at rate $s \geq 0$ per fit individual; and a part that depends on the frequency of the fit type and is encoded by the parameter γ. More precisely, a fit individual chooses at rate $\gamma \geq 0$ uniformly a partner from the population. If the partner is fit (unfit), then the type-0 individual does (not) reproduce. Each individual mutates at rate $u \geq 0$; the type after the event is i with probability $\nu_i,\ i \in \{0,1\}$, where $\nu_0, \nu_1 \geq 0$ and $\nu_0 + \nu_1 = 1$.

Let $Y_t^{(N)}$ be the (random) number of type-1 individuals at time t in a population of size N. The process $Y^{(N)} = (Y_t^{(N)})_{t \geq 0}$ is a continuous-time Markov chain with transition rates

$$q_{Y^{(N)}}(k, k+1) = k \, \frac{N-k}{N} + (N-k)u\nu_1, \tag{2.1}$$

$$q_{Y^{(N)}}(k, k-1) = k \, \frac{N-k}{N} \left(1 + s + \gamma \, \frac{N-k}{N}\right) + ku\nu_0, \tag{2.2}$$

where $k \in [N]_0 := [N] \cup \{0\}$ with $[N] := \{1, \ldots, N\}$. There are no other transitions. Since the dependence on the population size only becomes important in Chapter 3 and 4, when we consider the large population limit, we omit the superscript for the remainder of this chapter to ease the notation, i.e. we write Y instead of $Y^{(N)}$.

The Moran model with pairwise interaction has a well-known graphical representation as an interacting particle system, see Fig. 2.1. Here, individuals are represented by pieces of horizontal lines. Time runs from left to right in the figure. Reproduction events are depicted by arrows between the lines. If a parent (at the tail of an arrow) places offspring via the arrow, the offspring inherits the parent's type and replaces the individual at the tip. If an individual places offspring via an arrow, we say that the individual *uses* the arrow. We decompose reproduction events into neutral, selective, and interactive ones. Neutral arrows appear at rate $1/N$ per ordered pair of lines; selective arrows appear at rate s/N per ordered pair. Interactive arrows occur at rate γ/N per ordered pair of lines and are always accompanied by a checking arrow whose tip shares the tip of the interactive arrow; but whose tail is connected to a uniformly chosen line. That is, these arrow pairs occur at rate γ/N^2 per triple of lines. All types of arrows (including the interactive/checking pairs) are laid down via Poisson point processes independently of each other. The rules for their use are as follows. All individuals use the neutral arrows. In addition, the fit individuals use the selective arrows. Interactive arrows are used by a fit individual if there is a fit individual at the tail of the associated checking arrow.

Mutation events are depicted by crosses and circles on the lines. A circle (cross) indicates

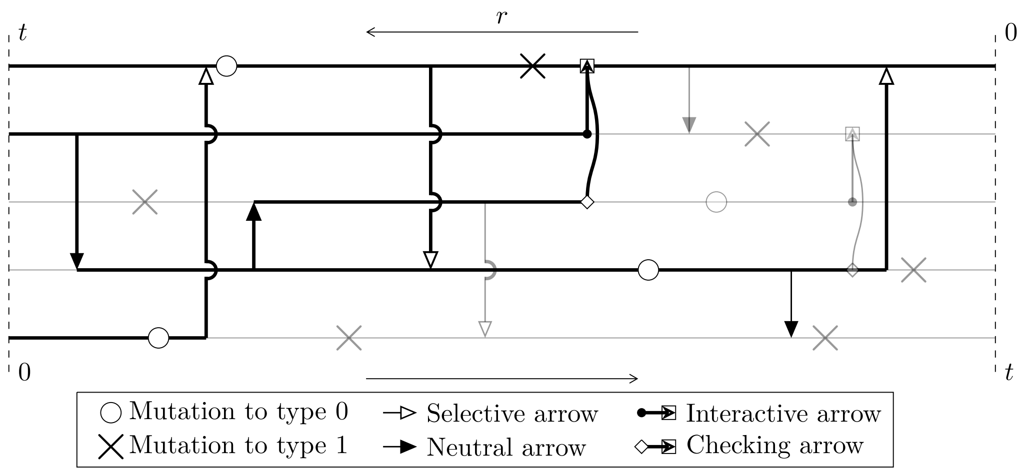


Figure 2.1. A realisation of the Moran interacting particle system (thin lines) for a population of size $N = 5$ and the embedded ASG (bold lines) for a sample of size 1. Time runs forward in the Moran model ($\rightarrow$) and backward in the ASG ($\leftarrow$). An arrowhead inscribed into a square marks the joint tip of an interactive and a checking arrow.

a mutation to type 0 (type 1), which means that the type on the line is 0 (is 1) after the mutation. This occurs at rate $u\nu_0$ (at rate $u\nu_1$) on every line, again by way of independent Poisson point processes. Note that it is no restriction to describe mutation in this parent-independent way; indeed, in the two-type case, the rates can always be parametrised in this way. Given a realisation of the particle system and an initial type configuration (that is, a type assigned to each line at $t = 0$), we can read off the types on the lines at all later times $t > 0$. The distribution of the initial types and the law of the graphical elements (arrows, circles, and crosses) are independent of each other.

Remark 2.1. The Moran model with pairwise interaction can be translated into the cooperative branching process on a complete graph [Mac17, Ch. I.1.2.3, Ch. I.2.1] (see also [MSS18b]) by interchanging the roles of type 0 and 1 and by setting $s = \nu_0 = 0$ and $u = 1$. A deleterious mutation then corresponds to a death event and an interactive arrow to a cooperative branching event. Other variants of such dynamics may be found in [Neu94; Nob92; SS15]. $\diamondsuit$

Remark 2.2. Graphical representations can be constructed in various ways for a comprehensive class of interacting particle systems, see e. g. [SS18, Sect. 5.2]. $\diamondsuit$

The graphical representation naturally induces a graph structure, which we implicitly use throughout the thesis. For the sake of completeness, we present its construction in detail in the next section.

2.1.1 Graphical representation as a directed graph with labels

The graphical representation of the Moran model is depicted in Fig. 2.1. We now turn this representation into an infinite uncountable directed graph $G = (V, E)$ with a function l that labels the vertices, where

$$V = \mathbb{R} \times [N], \quad E \subseteq V \times V, \quad \text{and} \quad l : V \to \{\blacktriangleright, \triangleright, \boxed{\triangleright}, \bullet\!-, \diamond\!-, \circ, \times, \emptyset\}.$$

For a vertex $(t, i) \in V$, we refer to $t \in \mathbb{R}$ as its time coordinate and to $i \in [N]$ as its site. The time coordinate and the site correspond to the horizontal and vertical position of the vertex in Fig. 2.1, respectively. The label of vertex (t, i) is denoted by $l(t, i)$.
Let us explain, how we set the edges and labels. A neutral, selective, and interactive arrow translates to a directed edge from the tail vertex to the tip vertex. The tip vertex of a neutral, selective, and interactive arrow obtains the label $\blacktriangleright$, $\triangleright$, and $\boxed{\triangleright}$, respectively. The tail vertex of an interactive and checking arrow obtains the label $\bullet\!-$ and $\diamond\!-$, respectively. A beneficial and deleterious mutation leads to the vertex label $\circ$ and $\times$, respectively. All remaining vertices obtain the label $\emptyset$. Time-consecutive vertices on the same site are connected if the vertex with larger time coordinate is not labelled $\blacktriangleright$, corresponding to a vertex with an incoming neutral arrow, and no vertex with an incoming edge from another site or with a mutation lies between the vertices. A vertex with label $\blacktriangleright$ has only one incoming edge and this edge originates from a vertex of another site. This leads to the following definition.

Definition 2.1 (Moran model graph)**.** Fix $N \in \mathbb{N}$. Consider the following five independent families of independent homogeneous Poisson processes on the real line

$$\begin{aligned}
&K^{\blacktriangleright}_{i,j} \text{ with rate } \frac{1}{N}, &&K^{\triangleright}_{i,j} \text{ with rate } \frac{s}{N}, &&K^{\boxed{\triangleright}}_{i,j,k} \text{ with rate } \frac{\gamma}{N^2},\\
&K^{\times}_{i} \text{ with rate } u\nu_1, &&K^{\circ}_{i} \text{ with rate } u\nu_1, &&i, j, k \in [N],\ i \neq j.
\end{aligned}$$

The *Moran model graph (of size N)* is the directed graph $G = (V, E)$ with label function l. For $t \in \mathbb{R}$ and $i \in [N]$, we have $(t, i) \in V$. For an arrival time t of

- $K^{\blacktriangleright}_{i,j}$, set $l(t, j) = \blacktriangleright$,
- $K^{\triangleright}_{i,j}$, set $l(t, j) = \triangleright$,
- $K^{\boxed{\triangleright}}_{i,j,k}$, set $l(t, j) = \boxed{\triangleright}$, $l(t, i) = \bullet\!-$, and $l(t, k) = \diamond\!-$.
- K°_{i}, set $l(t, i) = \circ$,
- $K^{\times}_{i}$, set $l(t, i) = \times$,

All remaining vertices obtain label $\emptyset$. The set of edges is as follows.

- For an arrival time t of $K^{\blacktriangleright}_{i,j}$, $((t, i), (t, j)) \in E$.
- For an arrival time t of $K^{\triangleright}_{i,j}$, $((t, i), (t, j)) \in E$.
- For an arrival time t of $K^{\boxed{\triangleright}}_{i,j,k}$, $((t, i), (t, j)) \in E$ and $((t, k), (t, j)) \in E$.
- For $(t_1, i), (t_2, i) \in V$ with $t_1 < t_2$, $l(t_2, i) \neq \blacktriangleright$, and $l(r, i) \notin \{\blacktriangleright, \triangleright, \boxed{\triangleright}, \circ, \times\}$ for all $r \in (t_1, t_2)$, $((t_1, i), (t_2, i)) \in E$.

These are all vertices and edges.

Remark 2.3. In the construction of the graph, start with the deterministic set of vertices $\mathbb{R} \times [N]$. On the basis of the Poisson processes assign the labels. Given the Poisson processes and labels, the edge set is then deterministic. $\diamondsuit$

We often consider the restriction of the Moran model graph to a finite horizon.

Definition 2.2. Fix $t_1 < t_2$ and a Moran model graph $G = (V, E)$ with label function l. The *Moran model graph in* $[t_1, t_2]$ is the directed graph $G_{[t_1,t_2]} := (V_{[t_1,t_2]}, E_{[t_1,t_2]})$, where $V_{[t_1,t_2]} = \{(t, i) \in V : t \in [t_1, t_2], i \in [N]\}$ and $E_{[t_1,t_2]} = E \cap (V_{[t_1,t_2]} \times V_{[t_1,t_2]})$, together with the label function l restricted to the vertices in $V_{[t_1,t_2]}$.

For an edge (v, w) of a directed graph with labels, we call v an in-neighbour of w and w an out-neighbour of v. Note that, by the above construction, every vertex in a graphical representation with label ► has exactly one in-neighbour. We now formalise the idea of propagating types through the graphical representation. Without loss of generality, we consider time intervals of the form $[0, t]$. We then write $G_t = (V_t, E_t)$ for the Moran model graph $G_{[0,t]} = (V_{[0,t]}, E_{[0,t]})$.

Definition 2.3 (Type propagation)**.** Consider a Moran model graph $G_t = (V_t, E_t)$ with label function l and $c \in \{0, 1\}^N$. We refer to c as the initial type configuration. The *type function* $v_c : V_t \to \{0, 1\}$ associated to G_t and c is defined as follows.
For two vertices $(r_1, i), (r_2, i) \in V_t$, we say that (r_1, i) transmits the type to the left of (r_2, i) if there is at least one directed path from (r_1, i) to (r_2, i) and there is no directed path from (r_1, i) to (r_2, i) that contains a vertex with label {►, ▻, ⍄, ○, ×} (except the start- and endpoint of the paths).
Set $v_c(0, i) = c_i$ for $i \in [N]$. For $(r, j) \in V_t$ with $r \in (0, t]$,

- if $l(r, j) = \circ$ (resp. $l(r, j) = \times$), then set $v_c(r, j) = 0$ (resp. $v_c(r, j) = 1$).
- if $l(r, j) =$ ► and $(r, i) \in V_t$ is the in-neighbour of (r, j), then set $v_c(r, j) = v_c(r, i)$.
- if $l(r, j) =$ ▻, $(r, i) \in V_t$ is the in-neighbour of (r, j) with $i \neq j$ and $v_c(r, i) = 0$, then set $v_c(r, j) = v_c(r, i)$.
- if $l(r, j) =$ ⍄ and for the in-neighbours $(r, i) \in V_t$ and $(r, k) \in V_t$ of (r, j) with $l(r, i) =$ •– and $l(r, k) =$ ◇–, respectively, we have $v_c(r, i) = v_c(r, k) = 0$, then set $v_c(r, j) = v_c(r, i)$.

Otherwise, set $v_c(r, j) = v_c(r', j)$, where $(r', j) \in V_t$ is the unique vertex with $l(r', j) \neq \emptyset$ or $r' = 0$ that transmits the type to the left of (r, j).

Every directed path connecting vertices on different sites contains a vertex with a label in {►, ▻, ⍄}. Hence, if a vertex transmits the type to the left of another vertex, then both vertices are on the same site and all directed paths connecting the two vertices remain on the same site.
Note that the number of vertices with labels different from $\emptyset$ is finite in any finite time interval almost surely. Furthermore, if $(r_1, i) \in V_t$ transmits the type to the left of $(r_2, i) \in V_t$, then for any $(r, i) \in V_t$ with $r \in (r_1, r_2)$ we have that (r_1, i) transmits the type to the left of (r, i). If $l(r_1, i) \neq \emptyset$ or $r_1 = 0$, then (r_1, i) is the unique vertex with this property that transmits the type to the left of (r, i). In particular, for an initial assignment of types, the type function is well-defined at all sites up to the first arrival time of one of the Poisson processes. Furthermore, the type of the vertices at the first arrival time is well-defined. An iteration of this argument shows that the type function is almost surely well-defined from time 0 to time t for any initial assignment of types and for any t finite. The notion of *type propagation* refers to the way the type function is constructed from time 0 to time t. In what follows, we are only concerned with the almost sure behaviour of the type function.

In addition to the types in the graph, we can keep track of the progeny of the individuals at time 0 by means of a more general type that we call *ancestral site*. The ancestral site of a vertex is the site of its ancestor at time 0. The ancestral site of a vertex with label –► (corresponding to an individual with incoming neutral arrow) agrees with the ancestral site of its in-neighbour. The ancestral site of a vertex with label –▻ (corresponding to an individual with incoming selective arrow) is the same ancestral site as the one of the in-neighbour at another site if this in-neighbour is fit. Otherwise, the ancestral site agrees with the ancestral site of the unique in-neighbour with label different from $\emptyset$ or time component 0, that transmits the type to its left. The ancestral site of a vertex with label →▣ (corresponding to an individual with incoming interactive arrow) is the same as the ancestral site of the in-neighbour with label •– if this in-neighbour (corresponding to an individual at the tail of an interactive arrow) is fit and the in-neighbour with label ◇– (corresponding to the individual at the tail of a checking arrow) is fit. Otherwise, the ancestral site agrees with the ancestral site of the unique in-neighbour with label different from $\emptyset$ or time component 0, that transmits the type to its left. The ancestral site of a vertex with label $\emptyset$ agrees with the ancestral site of the unique in-neighbour with label different from $\emptyset$ or time component 0, that transmits the type to its left. This leads to the following definition.

Definition 2.4 (Ancestral site propagation)**.** Consider a Moran model graph $G_t = (V_t, E_t)$ with label function l and an initial type configuration $c \in \{0,1\}^N$. Let v_c be the corresponding type-function. The ancestral site function $\eta_c : V_t \to [N]$ is defined as follows. Set $\eta_c(0,i) = i$ for $i \in [N]$. For $(r,j) \in V_t$ with $r > 0$,

- if $l(r,j) =$ –► and (r,j) has in-neighbour (r,i), then set $\eta_c(r,j) = \eta_c(r,i)$.
- if $l(r,j) =$ –▻, $(r,i) \in V_t$ is the in-neighbour of (r,j) with $i \neq j$ and $v_c(r,i) = 0$, then set $\eta_c(r,j) = \eta_c(r,i)$.
- if $l(r,j) =$ →▣ and for the in-neighbours $(r,i) \in V_t$ and $(r,k) \in V_t$ of (r,j) with $l(r,i) =$ •– and $l(r,k) =$ ◇–, respectively, we have $v_c(r,i) = v_c(r,k) = 0$, then set $\eta_c(r,j) = \eta_c(r,i)$.

Otherwise, set $\eta_c(r,j) = \eta_c(r',j)$, where $(r',j) \in V_t$ is the unique vertex with $l(r',j) \neq \emptyset$ or $r' = 0$ that transmits the type to the left of (r,j).

For any initial assignment of types and any fixed finite time interval, the ancestral site of the vertex in the interval is well-defined by the same argument that shows that the type function is well-defined. In particular, this motivates the notion of ancestral site propagation. We now formalise the concept of progeny and ancestry.

Definition 2.5 (Progeny/ancestry)**.** Consider a Moran model graph $G_t = (V_t, E_t)$ with label function l, a given initial type configuration $c \in \{0,1\}^N$, the type function v_c, and ancestral site function η_c. We say that $(r,i) \in V_t$ is in the progeny of $(0, \eta_c(r,i))$ or equivalently, $(0, \eta_c(r,i))$ is the ancestor of (r,i) at time 0. We denote the type of the ancestor of (r,i) at time 0 in the Moran model graph G_t with initial type configuration c by $J_t^c(r,i) := v_c(0, \eta_c(r,i))$.

2.1.2 Stationary distribution and fixation probabilities

In this section, we recall classic results, which are derived via the forward perspective. We first consider the case with mutations and derive the expression for the stationary type distribution. Subsequently, we assume the absence of mutations and derive the fixation probability. For the sake of completeness, we provide for all the results well known proofs.

For the Moran model, the process Y is a birth-death process on a finite state space with birth and death rates given by

$$\lambda_k := q_Y(k, k+1) \quad \text{and} \quad \mu_k := q_Y(k, k-1). \tag{2.3}$$

If $u, \nu_0, \nu_1 > 0$, Y is irreducible. In this case, there exists a unique stationary distribution π_Y with

$$\pi_Y(k) = \lim_{t\to\infty} \mathbb{P}(Y_t = k \mid Y_0 = j), \qquad \forall k,\, j \in [N]_0.$$

The expression for the stationary distribution is classical; we recall it in the following proposition.

Proposition 2.6 (Stationary distribution)**.** *Let $u, \nu_0, \nu_1 > 0$. Then,*

$$\pi_Y(k) = C_1 \prod_{i=1}^{k} \frac{\lambda_{i-1}}{\mu_i}, \qquad k \in [N]_0, \tag{2.4}$$

where $C_1 := \left(\sum_{k=0}^{N} \prod_{i=1}^{k} \frac{\lambda_{i-1}}{\mu_i}\right)^{-1}$ is a normalizing constant, which depends on all model parameters

Proof. It is readily checked that $\pi_Y(k)$ as given in (2.4) is indeed a probability distribution that satisfies the detailed balance equation, i. e. for all $k \in [N]$

$$\pi_Y(k-1)\, \lambda_{k-1} = \pi_Y(k)\, \mu_k$$

and therefore is a stationary distribution [Kel11, Thm. 1.3]. □

The non-interactive case (i. e. $\gamma = 0$) allows a more explicit expression of the stationary distribution. Denote by Γ the gamma function.

Corollary 2.7 (Stationary distribution without interaction)**.** *If $\gamma = 0$ and $u, \nu_0, \nu_1 > 0$,*

$$\pi_Y(k) = \tilde{C}_1 \frac{\binom{N}{k}}{(1+s)^k} \Gamma\Big(u\nu_1 N + k\Big) \Gamma\Big(\frac{u\nu_0 N}{1+s} + N - k\Big), \qquad k \in [N]_0, \tag{2.5}$$

with

$$\tilde{C}_1 := \left[\sum_{k=0}^{N} \frac{\binom{N}{k}}{(1+s)^k} \Gamma\Big(u\nu_1 N + k\Big) \Gamma\Big(\frac{u\nu_0 N}{1+s} + N - k\Big)\right]^{-1}. \tag{2.6}$$

Proof. The proof is a straightforward calculation. If $\gamma = 0$, the rates in (2.1) and (2.2) reduce to

$$\lambda_k = k\frac{N-k}{N} + (N-k)u\nu_1 \quad \text{and} \quad \mu_k = k\frac{N-k}{N}(1+s) + ku\nu_0.$$

Then,

$$\prod_{i=1}^{k}\frac{\lambda_{i-1}}{\mu_i} = \prod_{i=1}^{k}\frac{(N-i+1)(u\nu_1 N+i-1)}{i(u\nu_0 N+(N-i)(1+s))} = \frac{\binom{N}{k}}{(1+s)^k}\prod_{i=0}^{k-1}\frac{u\nu_1 N+i}{\frac{u\nu_0 N}{1+s}+N-1-i}.$$

Next, note that

$$\prod_{i=0}^{k-1}(u\nu_1 N+i) = \frac{\Gamma(u\nu_1 N+k)}{\Gamma(u\nu_1 N)} \quad \text{and} \quad \prod_{i=0}^{k-1}\Big(\frac{u\nu_0 N}{1+s}+N-1-i\Big) = \frac{\Gamma\Big(\frac{u\nu_0 N}{1+s}+N\Big)}{\Gamma\Big(\frac{u\nu_0 N}{1+s}+N-k\Big)}.$$

Together with the preceding calculation, we have

$$\prod_{i=1}^{k}\frac{\lambda_{i-1}}{\mu_i} = \frac{\binom{N}{k}}{(1+s)^k}\frac{\Gamma\Big(u\nu_1 N+k\Big)}{\Gamma\Big(u\nu_1 N\Big)}\frac{\Gamma\Big(\frac{u\nu_0 N}{1+s}+N-k\Big)}{\Gamma\Big(\frac{u\nu_0 N}{1+s}+N\Big)}.$$

The result is then a consequence of Proposition 2.6. To see this, note that

$$\bigg(\Gamma\Big(\frac{u\nu_0 N}{1+s}+N\Big)\,\Gamma\Big(\frac{u\nu_0 N}{1+s}+N\Big)\bigg)^{-1}$$

is also a factor in the constant appearing in Proposition 2.6 and therefore cancels. ☐

If $u=0$, Y either absorbs in 0 or N. The former (latter) corresponds to the extinction of type 1 (of type 0) and hence to the fixation of type 0 (of type 1). Let $T_k = \inf\{t>0 : Y_t = k\}$ be the first time when there are k type-1 individuals in the population. Then

$$w_Y(k) = \mathbb{P}(T_N < T_0 \mid Y_0 = k)$$

is the probability that Y hits N before 0 given that Y starts in k, i.e. it is the fixation probability of type 1 if there are initially k individuals of type 1. The expression for the absorption probability of a birth-death-process on a finite state space is well-known. We recall the result in the following proposition.

Proposition 2.8 (Fixation probability). *If $u=0$, then*

$$w_Y(k) = C_2\sum_{j=0}^{k-1}\prod_{\ell=1}^{j}\frac{\mu_\ell}{\lambda_\ell}, \qquad k\in[N]_0, \tag{2.7}$$

where $C_2 = 1/\sum_{j=0}^{N-1}\prod_{\ell=1}^{j}\frac{\mu_\ell}{\lambda_\ell}$.

Proof. The derivation of the fixation probability is also classical: A first-step decomposition of w_Y leads to the recursion

$$(\lambda_k + \mu_k) w_Y(k) = \lambda_k w_Y(k+1) + \mu_k w_Y(k-1), \qquad k \in [N-1],$$

complemented by $w_Y(0) = 0$ and $w_Y(N) = 1$. By induction, we obtain

$$w_Y(k+1) - w_Y(k) = \frac{\mu_k}{\lambda_k}\Big[w_Y(k) - w_Y(k-1)\Big] = w_Y(1)\prod_{\ell=1}^{k}\frac{\mu_\ell}{\lambda_\ell}.$$

Since $w_Y(0) = 0$,

$$w_Y(k+1) = \sum_{j=0}^{k}\Big(w_Y(j+1) - w_Y(j)\Big) = w_Y(1)\sum_{j=0}^{k}\prod_{\ell=1}^{j}\frac{\mu_\ell}{\lambda_\ell}.$$

On the other hand, $w_Y(N) = 1$, so that $w_Y(1) = 1/\sum_{j=0}^{N-1}\prod_{\ell=1}^{j}\frac{\mu_\ell}{\lambda_\ell}$. □

The case without mutation and interaction (i. e. $u = \gamma = 0$) allows a more explicit expression of the fixation probability. We recall the well-known result [Dur08, Thm. 6.1] in the following corollary.

Corollary 2.9 (Fixation probability without interaction). *If* $\gamma = u = 0$ *and* $s > 0$,

$$w_Y(k) = \frac{(1+s)^k - 1}{(1+s)^N - 1}, \qquad k \in [N]_0. \tag{2.8}$$

Proof. The result is a straightforward application of Proposition 2.8 and the following calculation

$$\sum_{j=0}^{k-1}\prod_{l=1}^{j}\frac{\mu_l}{\lambda_l} = \sum_{j=0}^{k-1}(1+s)^{-j} = \frac{(1+s)^k - 1}{s}.$$

□

Remark 2.4. Letting $s \to 0$ in Corollary 2.9, we recover the neutral case (i. e. $u = s = \gamma = 0$), where $w_Y(k) = k/N$. Indeed, taking the limit in (2.8) and applying L'Hôpital's rule yields

$$\lim_{s\to 0}\frac{(1+s)^k - 1}{(1+s)^N - 1} = \lim_{s\to 0}\frac{k(1+s)^{k-1}}{N(1+s)^{N-1}} = \frac{k}{N}.$$ ◇

In the following sections, we recover the results of the present section in terms of an ancestral process.

2.2 ASG in the Moran model with mutation, selection, and pairwise interaction

The ancestral selection graph (ASG) was introduced by Krone and Neuhauser [KN97; NK97] in order to study genealogies in population models with selection. It was extended by Neuhauser [Neu99] to a model of minority advantage, a special case of interactive selection. Following these lines, we now describe the ASG for the Moran model with mutation, selection, and pairwise interaction (cf. Fig. 2.1). For the sake of completeness, we subsequently give a formal definition of the ASG in terms of a directed graph with labelled vertices.

Our starting point is a realisation of the graphical representation of the Moran model in the time interval $[0, t]$ for some time $t > 0$, to which we refer as the present. We pick an untyped sample (that is, no types have been assigned to the individuals) at present and trace back the lines of its *potential influencers*; backward time will be denoted by r, where $r = 0$ corresponds to forward time t and $r = t$ corresponds to forward time 0. We call individuals potential influencers if their type has an influence on the type of the sampled individuals where, at this stage, we only take into account the information contained in the reproduction events, and ignore the additional information due to mutation. The arrows in the representation change the number of potential influencers and their respective locations. The ASG is composed of the lines of these potential influencers.

Assume there are currently n lines in the ASG. When a neutral arrow joins two lines in the current set, a *coalescence event* takes place, i. e. the two lines merge into the single one at the tail of the arrow and the number of lines in the graph decreases by one (see Fig. 2.2). Since neutral arrows appear at rate $1/N$ per ordered pair of lines, coalescence events occur at rate $n(n-1)/N$ in our ASG of size n. When a line in the current set is hit by an arrow that emanates from a line that is currently not in the graph, a *relocation event* occurs; i. e. the ASG continues with the *incoming branch* (the line at the tail of the arrow) and the number of lines in the graph does not change. Relocation events occur at rate $n(N-n)/N$.

Figure 2.2. Coalescence event (left) and relocation event (right).

When a selective arrow hits the current set of lines, the hit individual has two potential parents, namely the individual at the incoming branch, and the one at the *continuing branch* (the one to the left of the tip). Which of these is the true parent of the individual at the *descendant branch* (the one to the right of the tip) depends on the type at the incoming branch, but for the moment we work without types. This means that we must trace back both potential parents; we say the selective event remains *unresolved.* These events can be of two types: a *bifurcation event* (that is a binary branching; not to be confused with bifurcations in the context of dynamical systems, such as the transcritical and saddle node bifurcations in Sections 3.1 and 4.1) if the selective arrow emanates from a line outside the current set of lines, and a *simple collision event* if the selective arrow links two lines in the graph (see Fig. 2.3). The former increases the number of lines in the graph by one and, since selective arrows appear at rate s/N per ordered pair of lines, occurs at rate $sn(N-n)/N$ in our ASG of size n. The latter does not change the number of lines in the ASG and occurs at rate $sn(n-1)/N$.

Figure 2.3. Bifurcation (left) and simple collision (right).

When an interactive arrow hits a line in the ASG, the individual at the incoming branch and the individual at the checking branch are potential influencers. The true parent depends on the types of the individuals at these branches; but as before, we work without types. The resulting additional unresolved reproduction events can now be of three types: a *trifurcation* if both the incoming and the checking arrows emanate from lines currently not in the ASG; a *collision-bifurcation event* if either the incoming or the checking line, but not both of them, emanate from a line outside the current set of lines; and a *double collision event* if the incoming and the checking branches are currently in the graph (see Fig. 2.4). Since interactive arrow pairs occur at rate γ/N^2 per ordered triple of lines, in our ASG of size n a trifurcation occurs at rate $\gamma n(N-n)(N-n-1)/N^2$, a collision-bifurcation at rate $\gamma n(n-1)(N-n)/N^2$, and a double collision at rate $\gamma n(n-1)(n-2)/N^2$. In contrast to the original ASG (that is, without interaction), not all potential influencers are necessarily potential ancestors. Namely, the individual on the checking line is, in general, not ancestral; but its type may have an influence on the type of the sampled individual. We generally refer to bifurcations and trifurcations as branching events. The number of lines in the ASG decreases by one in a coalescence event, increases by one in a bifurcation or collision-bifurcation event, increases by two in a trifurcation event, and remains unchanged in simple collision, double collision and relocation events. As in the Moran model, beneficial and deleterious mutations are superimposed on the lines at rate $u\nu_0$ and $u\nu_1$, respectively. The resulting object is called the *untyped ASG*; this refers to the fact that the initial types have not yet been assigned and the consequences of mutations are still unresolved.

Figure 2.4. From left to right: trifurcation, collision-bifurcation with incoming branch already in ASG, collision-bifurcation with checking branch already in ASG, double collision.

Once the untyped ASG has been constructed, the true ancestry of the initial sample is obtained as follows. First, assign types to all lines in the ASG at forward time 0, that is, backward time t, without replacement from an exchangeable distribution at forward time 0. Then, propagate the types up to time t respecting the type propagation rules of Section 2.1: Mutation circles and crosses on the ASG turn the type on that line to a type 0 or type 1, respectively (this also includes the possibility of no type change, i. e. the mutations are silent). At every selective event, the individual at the incoming branch is the ancestor if it is of type 0; otherwise the individual at the continuing branch is the ancestor. This hierarchy will be called the *pecking order*. Note that the descendant is of type 1 if and only if both the incoming and the continuing branch are of type 1. At every interactive event, the individual at the incoming branch is the ancestor if it is of type 0 and if the individual at the checking branch is also fit. Otherwise, the individual at the continuing line is the ancestor. Under these rules, the types of the sampled individuals are recovered together with their ancestry.

The above description of the ASG in $[0,t]$ has a natural correspondence to a directed graph with labelled vertices. It arises from the Moran model graph in $[0,t]$ (see Definition 2.2) as follows. First, reverse the direction of time and the direction of the edges in the Moran model graph. Second, restrict the graph to the part that is reachable from the sampled individuals at forward time t. This leads to the untyped ASG.

Definition 2.10 (ASG for finite populations). Consider a Moran model graph in $[0,t]$ denoted by $G_t = (V_t, E_t)$ with label function l. Define $\bar{G}_t := (V_t, \bar{E}_t)$ and label function $\bar{l}$ where

$$\begin{aligned}\bar{E}_t &:= \{((r,i),(r',j)) \in V_t \times V_t : ((t-r',j),(t-r,i)) \in E_t\}\\ \bar{l}(r,i) &:= l(t-r,i) \qquad \text{for } (r,i) \in V_t.\end{aligned}$$

Fix $A \subseteq [N]$ and define $\bar{G}_t^A := (\bar{V}_t^A, \bar{E}_t^A)$ by

$$\bar{V}_t^A := \{(r,i) \in V_t : \exists \text{ directed path in } \bar{G}_t \text{ from } (0,j) \text{ to } (r,i) \text{ for some } j \in A\},$$

$\bar{E}_t^A := \bar{E}_t \cap (\bar{V}_t^A \times \bar{V}_t^A)$, and $\bar{l}^A$ is the restriction of $\bar{l}$ to the vertices in $\bar{G}_t^A$. The ASG in $[0,t]$ starting in A is then $\mathcal{A}_t^A := (\bar{G}_t^A, \bar{l}^A)$.

We refer to the vertices in $\mathcal{A}_t^A$ with time component 0 as root(s) and to the vertices with time component t as leaves. If we consider the Moran model in $[-t,0]$ rather than in $[0,t]$, the ASGs can be coupled such that for $t > 0$, we have $\mathcal{A}_r^A \subseteq \mathcal{A}_t^A$ for $r \in [0,t]$. When we describe the evolution of an ASG (and pruned- or killed versions of it, see Section 2.3.1 and 2.4.1), we mean the evolution of the ASGs on increasing time intervals coupled in this way.

Because each vertex with a given time component in an ASG corresponds to a potential influencer, and hence to an individual in the Moran model at that time, we will occasionally abuse the notation and refer to the vertices with a given time coordinate as individuals. The definition of type and ancestral site propagation, progeny, and ancestry translate also to the ASG.

Definition 2.11. [Types and ancestral sites in ASG] Consider an ASG in $[0,r]$ starting from $A \subseteq [N]$ denoted by $\mathcal{A}_r^A = (\bar{G}_r^A, \bar{l})$ with $\bar{G}_r^A = (\bar{V}_r^A, \bar{E}_r^A)$. Reverse the direction of all edges. For an initial type configuration $c \in \{0,1\}^N$, define the type and ancestral site function $\bar{v}_c$ and $\bar{\eta}_c$, respectively, as follows. Set $\bar{v}_c(r,i) = c_i$ and $\bar{\eta}_c(r,i) = i$ for $(r,i) \in \bar{V}_r^A$. Propagate types and ancestral sites as follows. For $(t,j) \in \bar{V}_r^A$ with $t \in (r,0]$,

- if $l(t,j) = \circ$ (resp. $l(t,j) = \times$), then set $\bar{v}_c(t,j) = 0$ (resp. $\bar{v}_c(t,j) = 1$) and $\bar{\eta}_c(t,j) = \bar{\eta}_c(t',j)$, where $(t',j) \in \bar{V}_r^A$ is the unique vertex with $l(t',j) \neq \emptyset$ or $t' = r$ that transmits the type to the left of (t,j).
- if $l(t,j) = \blacktriangleright$ and $(t,i) \in \bar{V}_r^A$ is the in-neighbour of (t,j), then set $\bar{v}_c(t,j) = \bar{v}_c(t,i)$ and $\bar{\eta}_c(t,j) = \bar{\eta}_c(t,i)$.
- if $l(t,j) = \triangleright$, $(t,i) \in \bar{V}_r^A$ is the in-neighbour of (t,j) with $i \neq j$ and $\bar{v}_c(t,i) = 0$, then set $\bar{v}_c(t,j) = \bar{v}_c(t,i)$ and $\bar{\eta}_c(t,j) = \bar{\eta}_c(t,i)$.

- if $l(t,j) = \rightarrow\!\boxed{\triangleright}$ and for the in-neighbours $(t,i) \in \bar{V}_r^A$ and $(t,k) \in \bar{V}_r^A$ of (t,j) with $l(t,i) = \bullet\!\!-$, $l(t,k) = \diamond\!\!-$, and $(t,i) \neq (t,j)$, we have $\bar{v}_c(t,i) = \bar{v}_c(t,k) = 0$, then set $\bar{v}_c(t,j) = \bar{v}_c(t,i)$ and $\bar{\eta}_c(t,j) = \bar{\eta}_c(t,i)$.

Otherwise, set $\bar{v}_c(t,j) = \bar{v}_c(t',j)$ and $\bar{\eta}_c(t,j) = \bar{\eta}_c(t',j)$, where $(t',j) \in \bar{V}_r^A$ is the unique vertex with $l(t',j) \neq \emptyset$ or $t' = r$ that transmits the type to the left of (t,j).
For $i \in A$ we say that $(r, \bar{\eta}_c(0,i))$ is the ancestor of $(0,i)$ and $\bar{J}_r^c(0,i) := \bar{v}_c(r, \bar{\eta}_c(0,i))$ is the type of the ancestor of $(0,i)$.

The argument following Definition 2.3 that showed that the type and ancestral site function is well-defined in the Moran model graph also applies here.
In the ASG the direction of the edges between different sites is the reverse of the direction of the edges in the underlying Moran model graph. In all our figures, the direction of arrows corresponds to the direction in the underlying Moran model graph.

2.3 Type distribution in the Moran model via the killed ASG

Our first aim is to connect the ASG rigorously to the Moran model via duality. We consider here only the case without interaction, i. e. $\gamma = 0$. To find a similar relation in the case $\gamma > 0$ is more involved and we only deal with it later in the smsssi-limit in Chapter 4. In Section 2.3.1, we identify a killed ASG as the appropriate ancestral structure. Section 2.3.2 contains the duality result that formally connects the forward and the backward perspective. In the end, we exploit the duality to derive the long term properties of the Moran model and the killed ASG (Section 2.3.3).

2.3.1 Killed ASG for finite populations

In this section we define the appropriate ancestral structure to characterise the type distribution in the Moran model.

Our starting point is the well-known duality relation between the Wright–Fisher diffusion with selection (but without mutation) and the line-counting process of the ASG (see [KN97] and [Man09, Thm. 2.1]). The duality relation translates also into the finite population setting. The main idea is that in the *absence of mutations*, a single individual at time t is of type 1 if and only if all its potential ancestors at time 0 are of type 1. This is easily verified via the pecking order (cf. Fig. 2.5). Namely, at every branching event, a type 0 on either

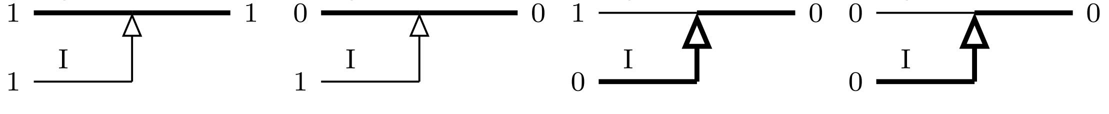


Figure 2.5. The descendant line (D) splits into the continuing line (C) and the incoming line (I). The incoming line is ancestral if and only if it is of type 0. The true ancestral line is drawn in bold.

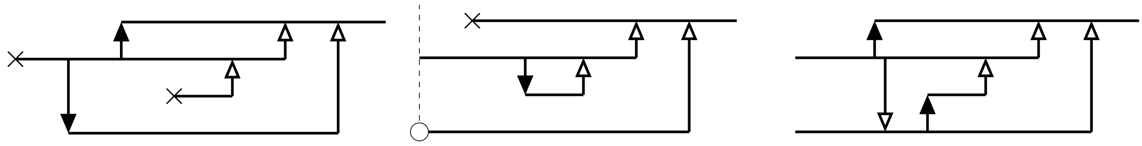


Figure 2.6. The killed ASG either absorbs in a state with 0 lines due to mutations to type 1 (left) or in a cemetery state Δ due to a mutation to type 0 (center). In the absence of mutations, the killed ASG coincides with the ASG (right).

the continuing or incoming line suffices for the descendant to be of type 0; iterating this over all branching events gives the statement. The corresponding duality in the weak mutation–weak selection limit appears already in [SU86, Lem. 2.1] (see also [AS05, Eq. (1.5)]), but without the interpretation in terms of potential ancestors.

Remark 2.5. In the case $\gamma > 0$, this statement is not true any more. The type propagation rule at trifurcations makes the analysis much more complicated. ♢

Mutations contain additional information and can allow us to determine the type of the sampled individual even before assigning types to the potential ancestors. More precisely, a mutation to type 1 determines the type of the line (to the right of the mutation) on which it occurs, so this line need not be traced back further into the past; it may be pruned. Next, the first mutation to type 0 (on any line that is still alive after the pruning) decides that the sampled individual has type 0, so that *no* potential ancestor must be considered any further and the process may be killed. We refer to the resulting structure as *killed ASG*. We now provide a definition in terms of a directed graph with labelled vertices in the spirit of Definition 2.10. The graph contains the set of vertices that are reachable from the root(s) of the ASG, after removing all outgoing edges at vertices with a mutation label.

Definition 2.12 (Killed ASG for finite populations). Let $\mathcal{A}_r^A = (\bar{G}_r^A, \bar{l}^A)$ with $\bar{G}_r^A = (\bar{V}_r^A, \bar{E}_r^A)$ be an ASG in $[0, r]$ started from $A \subseteq [N]$. Define $\bar{E}_r^{A,\star} := \{((t_1, i), (t_2, j)) \in \bar{E}_r^A : l(t_1, i) \notin \{\circ, \times\}\}$ and

$$\begin{aligned}\bar{V}_r^{A,\dagger} &:= \{(t, i) \in \bar{V}_r^A : \exists \text{ directed path in } (\bar{V}_r^A, \bar{E}_r^{A,\star}) \text{ from } (0, j) \text{ to } (t, i) \text{ for some } j \in A\},\\ \bar{E}_r^{A,\dagger} &:= \bar{E}_r^{A,\star} \cap (\bar{V}_r^{A,\dagger} \times \bar{V}_r^{A,\dagger}),\end{aligned}$$

and $\bar{l}^{A,\dagger}$ is the restriction of $\bar{l}^A$ to $\bar{V}_r^{A,\dagger}$. Let $T_\dagger$ be the time of the first beneficial mutation (after time 0) in $\bar{G}_r^{A,\dagger} = (\bar{V}_r^{A,\dagger}, \bar{E}_r^{A,\dagger})$ with label function $\bar{l}^{A,\dagger}$. Set $T_\dagger = r$ if there is no such mutation. The *killed ASG of* $\mathcal{A}_r^A$ is denoted by $\text{✂}^\dagger(\mathcal{A}_r^A)$ and it is the restriction of $\bar{G}_r^{A,\dagger}$ and $\bar{l}^{A,\dagger}$ to vertices with time component in $[0, T_\dagger]$.

Fig. 2.6 depicts some realisations of the killed ASG. The propagation of types is analogous to the ASG (see Definition 2.11).

Next, let us explain the dynamics of the killed ASG process. It starts with one line emerging from each of the n individuals in a sample. Every line branches at rate $s(N - n)/N$, due to a selective arrow from outside the current set of lines. In particular, this increases the number of lines in the graph. A selective arrow connects an ordered pair of lines in the graph at rate s/N. This does not change the current number of lines. An ordered pair of lines coalesces to a single potential ancestor at rate $1/N$, i.e. there is a neutral edge between the two vertices in the graph. The number of lines in the graph decreases by one. Every line is pruned at rate $u\nu_1$ (due to a deleterious mutation), decreasing the number of lines by one. At rate $u\nu_0$ per line, the process is killed (due to a beneficial mutation), that is, it is sent to the cemetery state Δ.

It will turn out that for our purposes, it suffices to only keep track of the number of lines in the killed ASG process. The following definition of the corresponding line-counting process arises from the above description.

Definition 2.13 (Line-counting process of killed ASG for finite populations)**.** The *line-counting process of the killed ASG* is the continuous-time Markov chain $R = (R_r)_{r\geq 0}$ with values in $[N]_{0,\Delta} := [N]_0 \cup \{\Delta\}$. The transition rates are given by

$$q_R(n,n+1) = sn\frac{N-n}{N}, \quad q_R(n,n-1) = n\frac{n-1}{N} + nu\nu_1, \quad q_R(n,\Delta) = nu\nu_0, \tag{2.9}$$

for $n \in [N]$.

The states 0 and Δ are absorbing; all other states are transient. The state 0 is reached if all lines are pruned due to deleterious mutations. The state Δ is reached upon the first beneficial mutation. Absorption in 0 (in Δ) implies that (not) all individuals in the sample are of type 1. If $u = 0$, the absorption states are not accessible and the killed ASG is never pruned or killed. R is then just the line-counting process of the ASG.

In the next section, we connect R to the Moran model via duality. This paves the way to characterise the type distribution in the Moran model by means of the ancestral structure.

2.3.2 Factorial moment duality between Moran model and killed ASG

In the present section, we derive the formal connection between the killed ASG and the Moran model. The main result is the factorial moment duality between the number of unfit individuals in the Moran model and the line-counting process of the killed ASG in Theorem 2.14.

For $k, n \in \mathbb{N}_0$, let

$$k^{\underline{n}} := k(k-1)\cdots(k-n+1)$$

be the falling factorial. Set $k^{\underline{0}} := 1$ and $k^{\underline{n}} := 0$ for $n > k \geq 0$. Define $H : [N]_0 \times [N]_{0,\Delta} \to \mathbb{R}$ as

$$H(k,n) = \frac{k^{\underline{n}}}{N^{\underline{n}}}, \qquad k \in [N]_0,\, n \in [N]_{0,\Delta}, \tag{2.10}$$

where $n^{\underline{\Delta}}/N^{\underline{\Delta}} := 0$ for all $n \in [N]_0$. If $n \in [N]_0$, then $H(k,n)$ is the probability of sampling n individuals of type 1 without replacement from a population with k type-1 individuals.

Theorem 2.14 (Duality)**.** *The Moran model $(Y_t)_{t\geq 0}$ and the line-counting process of the killed ASG $(R_t)_{t\geq 0}$ satisfy the duality relation for $t \geq 0$*

$$\mathbb{E}_k[H(Y_t,n)] = \mathbb{E}_n[H(k,R_t)], \tag{2.11}$$

for all $k \in [N]_0$ and $n \in [N]_{0,\Delta}$.

Remark 2.6. For $\nu_0 = 0$, this factorial moment duality is the analogue to the classic moment duality in the weak mutation–weak selection framework (see Remark 3.4) but in the finite population setting. ♢

The underlying idea of the duality is as follows. Sample n individuals at time t in the Moran model. To determine the probability that these n individuals are all unfit, one may proceed in two ways. Either, determine the probability that the n individuals at time t are unfit if the types are sampled without replacement from a population with Y_t type-1 individuals given that $Y_0 = k$. Alternatively, consider n (untyped) individuals at time t and trace back their associated killed ASG back to time 0. If the killed ASG is absorbed in 0 (in Δ), (not) all n individuals are unfit and the type distribution at time 0 is not relevant. If it is not absorbed, determine the probability that all the remaining lines are assigned unfit types if they are sampled without replacements from a population with k type-1 individuals.

Proof of Theorem 2.14. The infinitesimal generator of Y acts on functions $f : [N]_0 \to \mathbb{R}$ and is given by $A_Y f(k) := A_Y^d f(k) + A_Y^s f(k) + A_Y^u f(k)$, where

$$A_Y^d f(k) = k\,\frac{N-k}{N}\,[f(k+1) - f(k)] + k\,\frac{N-k}{N}\,[f(k-1) - f(k)], \tag{2.12}$$

$$A_Y^s f(k) = sk\,\frac{N-k}{N}\,[f(k-1) - f(k)], \tag{2.13}$$

$$A_Y^u f(k) = (N-k)u\nu_1\,[f(k+1) - f(k)] + ku\nu_0\,[f(k-1) - f(k)] \tag{2.14}$$

correspond to genetic drift, selection, and mutation, respectively. On the other hand, the infinitesimal generator of R acts on functions $\tilde{f} : [N]_{0,\Delta} \to \mathbb{R}$ and is given by $A_R \tilde{f}(n) := A_R^d \tilde{f}(n) + A_R^s \tilde{f}(n) + A_R^u \tilde{f}(n)$, where

$$A_R^d \tilde{f}(n) = n\,\frac{n-1}{N}\,[\tilde{f}(n-1) - \tilde{f}(n)], \tag{2.15}$$

$$A_R^s \tilde{f}(n) = sn\,\frac{N-n}{N}[\tilde{f}(n+1) - \tilde{f}(n)], \tag{2.16}$$

$$A_R^u \tilde{f}(n) = nu\nu_1\,[\tilde{f}(n-1) - \tilde{f}(n)] + nu\nu_0\,[\tilde{f}(\Delta) - \tilde{f}(n)] \tag{2.17}$$

correspond to coalescence, selection, and mutation, respectively. Since the state space of Y is finite, every function is in the domain of A_Y. In particular, $H(\cdot, n)(k)$ and $P_t^Y H(\cdot, n)(k)$ lie in the domain of A_Y, where $(P_t^Y)_{t \geq 0}$ is the transition semigroup corresponding to Y. Similarly, $H(k, \cdot)(n)$ and $P_t^R H(k, \cdot)(n)$ lie in the domain of A_R, where $(P_t^R)_{t \geq 0}$ is the transition semigroup corresponding to R. We will show that

$$A_Y H(\cdot, n)(k) = A_R H(k, \cdot)(n) \qquad \text{for } k \in [N]_0,\ n \in [N]_{0,\Delta}.$$

The result then follows from [JK14, Prop. 1.2]. For the genetic drift and coalescence part, we have

$$\begin{aligned} A_Y^d H(\cdot, n)(k) &= k\,\frac{N-k}{N}\left(\frac{[k+1]^{\underline{n}} - k^{\underline{n}}}{N^{\underline{n}}} + \frac{[k-1]^{\underline{n}} - k^{\underline{n}}}{N^{\underline{n}}}\right) \\ &= k\,\frac{N-n+1-(k-n+1)}{N^{\underline{n}}}\,n\,\frac{(n-1)}{N}\,[k-1]^{\underline{n-2}} \end{aligned}$$

$$= A_R^d H(k,\cdot)(n).$$

For the selection parts we have

$$A_Y^s H(\cdot,n)(k) = s\,\frac{n}{N}\,k\,\frac{k-n-(N-n)}{N^{\underline{n}}}\,[k-1]^{\underline{n-1}} = A_R^s H(k,\cdot)(n).$$

In the end, we match the mutation parts,

$$\begin{aligned} A_Y^u H(\cdot,n)(k) &= (N-n-(k-n))\,nu\nu_1\,\frac{k^{\underline{n-1}}}{N^{\underline{n}}} - knu\nu_0\,\frac{[k-1]^{\underline{n-1}}}{N^{\underline{n}}} \\ &= A_R^u H(\cdot,n)(k). \end{aligned}$$

□

2.3.3 Properties and applications of the killed ASG for finite populations

The duality result of the previous section allows us to connect the number of unfit individuals in the Moran model with the line-counting process of the killed ASG and vice versa. We are particularly interested in the long-term behaviour of the two processes. As a consequence of the duality, we obtain a characterisation of the stationary distribution and the fixation probability in the Moran model in terms of the killed ASG. In particular, we recover the classic results of Section 2.1.2 by genealogical means.

If $u=0$, Y absorbs in either 0 or N, whereas R has no absorbing states and coincides with the number of lines in the ASG. The line-counting process of the (killed) ASG then has a stationary distribution π_R given in the following proposition.

Proposition 2.15 (Stationary distribution of R)**.** *If $u=0$ and $s>0$, the stationary distribution of R is given by*

$$\pi_R(n) = \frac{\binom{N}{n}s^n}{(1+s)^N-1}, \qquad n \in [N]. \tag{2.18}$$

In particular, π_R follows a Binomial distribution with parameters N and $s/(1+s)$ that is conditioned to be positive.

Remark 2.7. That π_R follows a Binomial distribution with parameters N and $s/(1+s)$ that is conditioned to be positive was already noted by [Cor17a, Sect. 4.4]. ◇

Proof of Proposition 2.15. It is readily verified that the detailed balance condition holds and π_R is indeed a probability distribution. For the second statement, let B be a random variable with Binomial distribution with parameters N and $s/(1+s)$. For $n \in [N]$,

$$\mathbb{P}(B=n \mid B>0) = \frac{\mathbb{P}(B=n)}{1-\mathbb{P}(B=0)} = \frac{\binom{N}{n}\left(\frac{s}{1+s}\right)^n\left(\frac{1}{1+s}\right)^{N-n}}{1-(1+s)^{-N}} = \pi_R(n).$$

□

We denote by R_∞ a random variable on $[N]$ distributed according to π_R. As a consequence of Theorem 2.14, we can express the fixation probability of type 1 in the Moran model via the stationary distribution of the dual process.

Corollary 2.16 (Fixation probability of Y). *If $u = 0$ and $s > 0$, let $w_Y(k)$ be the fixation probability of type 1 of the Moran model started with k type-1 individuals. Let R_∞ be distributed according to π_R. Then,*

$$w_Y(k) = \mathbb{E}\left[\frac{k^{\underline{R_\infty}}}{N^{\underline{R_\infty}}}\right], \qquad k \in [N]_0. \tag{2.19}$$

Remark 2.8. Corollary 2.16 yields an alternative way to derive the expression for the fixation probability in the Moran model of Corollary 2.9. Proposition 2.15 and the fact that $k^{\underline{\ell}}/\ell^{\underline{\ell}} = \binom{k}{\ell}$ leads to

$$w_Y(k) = \sum_{\ell=1}^{N} \pi_R(\ell)\frac{k^{\underline{\ell}}}{N^{\underline{\ell}}} = \frac{1}{(1+s)^N - 1}\sum_{\ell=1}^{k}\binom{N}{\ell} s^\ell \frac{k^{\underline{\ell}}}{N^{\underline{\ell}}} = \frac{(1+s)^k - 1}{(1+s)^N - 1},$$

which coincides with the expression for the fixation probability in Corollary 2.9. ♢

Proof of Corollary 2.16. Set $n = 1$ in (2.11). Taking the limit $t \to \infty$ on both sides leads to the result. □

The corollary provides an interpretation for the fixation probability of the Moran model in terms of the ancestral process. An individual sampled in the distant future is of type 1 if and only if all lines in the stationary ancestral selection graph are of type 1.

If $u, \nu_0, \nu_1 > 0$, then R absorbs almost surely; whereas Y has no absorbing states. Instead, the latter has stationary distribution π_Y. We denote by Y_∞ a random variable on $[N]_0$ distributed according to π_Y. In contrast, R absorbs in this parameter setting in either 0 or Δ. Let $T_0 = \inf\{r > 0 : \ R_r = 0\}$ and $T_\Delta = \inf\{r > 0 : \ R_r = \Delta\}$ be the first time R hits 0 and Δ, respectively. Set

$$w_R(n) = \mathbb{P}(T_0 < T_\Delta \mid R_0 = n),$$

so that w_R coincides with the absorption probability of R in 0 if R starts in n. The following result is a direct consequence of the duality in Theorem 2.14.

Corollary 2.17 (Absorption probability of R). *If $u, \nu_0, \nu_1 > 0$,*

$$w_R(n) = \mathbb{E}\left[\frac{Y_\infty^{\underline{n}}}{N^{\underline{n}}}\right], \qquad n \in [N]_0. \tag{2.20}$$

The genealogical interpretation of the above result is as follows. R is almost surely absorbed (in $\{0, \Delta\}$) in finite time. To see this, note that a beneficial mutation, which would lead to the killing of R, occurs at rate $u\nu_0$ per line. While R is not absorbed, there is at least one such line. In particular, if we pick n individuals in the distant future we may assume that at present their associated killed ASG is already absorbed in either 0 or Δ. On the other hand, if we sample n individuals in the distant future, we sample them according to the stationary distribution. Hence, the probability that R, which is started with n lineages in the distant future, is absorbed in 0 coincides with the probability that n individuals sampled from the stationary population are unfit.

Remark 2.9. Corollary 2.17 can be used to determine the moments of Y_∞ by genealogical means. Combining a first-step decomposition of $w_R(k)$ with the corollary leads to a recursion of the falling factorial moments of Y_∞. We use such an approach later in this thesis, e.g. in Corollary 3.5. $\diamondsuit$

Proof of Corollary 2.17. Set $k = N$ in (2.11). Taking the limit $t \to \infty$ on both sides leads to the result. $\square$

The explicit expression for the stationary distribution of Y in Corollary 2.7 allows us to state an explicit expression for the absorption probability of R in 0.

Corollary 2.18. *Let $u, \nu_0, \nu_1 > 0$ and let $w_R(n)$ be the absorption probability of R in 0 if R started at n. Then, for $n \in [N]_0$,*

$$w_R(n) = \frac{\tilde{C}_1}{(1+s)^n} \sum_{k=0}^{N-n} \frac{\binom{N-n}{k}}{(1+s)^k} \Gamma\bigg(u\nu_1 N + n + k\bigg) \Gamma\bigg(\frac{u\nu_0 N}{1+s} + N - (n+k)\bigg), \tag{2.21}$$

with $\tilde{C}_1$ as given in (2.6).

Proof. The statement is a straightforward consequence of Corollary 2.17 together with Corollary 2.7. $\square$

2.4 Ancestral type distribution

So far we have been concerned with the type distribution of the population in the future. Now, we turn our attention to the type distribution of the ancestors of the individuals from today's population.

When tracing back the genealogy of the entire population, the number of ancestors of this population will decrease. Cordero [Cor17a, Lem. 4.3] proved that in the Moran model all ancestral lines eventually coalesce to a common ancestor (see also [KN97, Thm. 3.2] and [Cor17a, Sect. 3]). Hence, in the sufficiently distant past the ancestral line of any individual coincides with the ancestral line of the common ancestor. One way to see this is via the ASG. The genealogy of the entire population is contained in the ASG started from the entire population. At each coalescence event between lines of the genealogy, the number of ancestors decreases. All other events in the ASG keep the number of ancestors at most constant. If the line-counting process of the ASG reaches 1, which occurs in finite time almost surely, the number of ancestors is also 1. In particular, the ancestral type at a prior time of the roots of the ASG coincides with the type of the common ancestor at this (prior) time.

Definition 2.19 (Ancestral type distribution, common ancestor type distribution).
Consider a Moran model in the time interval $[0, r]$. Let J_r^c be the type of the ancestor at backward time 0 of a uniformly chosen individual at time $r > 0$ if the initial type

configuration is $c \in \{0,1\}^N$. The ancestral type distribution, conditional on an initial type configuration with k type-1 individuals, is defined as

$$h_r(k) := \mathbb{P}(J_r^c = 1 \mid |c| = k),$$

where $|c| = \sum_{i=1}^N c_i$ and $k \in [N]_0$.
The common ancestor type distribution, conditional on an initial type configuration with k type-1 individuals, is defined as

$$h_\infty(k) := \lim_{r \to \infty} h_r(k), \qquad k \in [N]_0,$$

if the limit exists. For $u, \nu_0, \nu_1 > 0$, the common ancestor type distribution at equilibrium is defined as

$$h_\infty(Y_\infty) := \mathbb{E}\big[h_\infty(Y_\infty)\big],$$

where Y_∞ is distributed according to π_Y.

The ancestral type distribution is difficult to determine via the forward point of view. In what follows, we still assume $\gamma = 0$, so that all lines in the ASG correspond to lines of potential ancestors. We can exploit the ideas developed in Section 2.3 to tackle the problem.

2.4.1 Pruned lookdown ASG for finite populations

In this section, we recapitulate the construction of the pruned lookdown ASG (pLD-ASG) and the representation of the ancestral type distribution in terms of its line-counting process.

The identification of the ancestor of an individual picked uniformly at present by means of the ASG is not straightforward, since we now must identify the parental branch (incoming or continuing, depending on the type) at every branching event. This requires nested case distinctions. Furthermore, some ancestral lines must be traced back beyond the first mutation. Nevertheless, mutations may, as in the killed ASG, still rule out certain potential ancestors. This leads to the notion of relevant potential ancestor.

Definition 2.20 (Relevant potential ancestors). Consider an ASG $\mathcal{A}_r^A = (\bar{G}_r^A, \bar{l}^A)$ starting from $A \subseteq [N]$ with $\bar{G}_r^A = (\bar{V}_r^A, \bar{E}_r^A)$. We call a leaf $(r,j) \in \bar{V}_r^A$ a *relevant potential ancestor* of A if there is an initial type configuration $c \in \{0,1\}^N$ such that for some $i \in A$, we have $\bar{\eta}_c(0,i) = j$.

The pLD-ASG, introduced by Lenz et al. [Len+15] in the diffusion limit of the Moran model and extended by Cordero [Cor17a, Sect. 4] to the finite population setting, keeps track of the relevant potential ancestors in an ASG started from a single individual. We first recall the idea of the pLD-ASG and then define the associated line-counting process, which will be our main tool to analyse the ancestral type distribution.

The starting point is a single individual in the Moran model. The pLD-ASG is contained in the ASG starting from this individual. Its lines correspond to the relevant potential ancestors and are assigned consecutive levels in the non-negative integers starting at level 1. The levels indicate the priority of the relevant potential ancestors to be the true ancestor if all relevant potential ancestor are fit. A lower level means a higher priority. If a line in the pLD-ASG is hit by a selective arrow from outside the current set, its level increases by

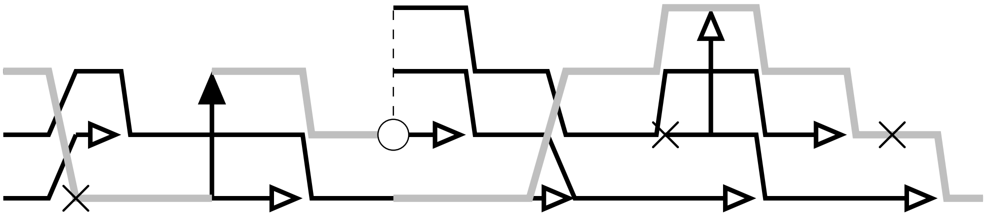


Figure 2.7. Realisation of the pLD-ASG. Grey: level of immune line.

one and, at the same time, all lines with higher levels also increase their level by one. The level formerly assigned to the line hit by the selective arrow is assigned to the incoming line. If a selective arrow joins two lines in the pLD-ASG with the line at the tail having a larger level than the line at the tip, then the line at the tail obtains the level of the line at the tip and the line at the tip increases its level by one; all lines with a level in between increase their level by one. If a selective arrow joins two lines in the pLD-ASG with the line at the tail having a smaller level than the line at the tip, both lines continue with their level. If two lines coalesce, the line with the higher level stops and only the line with the lower level continues. The lines that have a larger level than both of the lines involved in such a coalescence event decrease their level by one. This ordering of the lines is inspired by the lookdown constructions of Donnelly and Kurtz [DK99b].

Next, let us explain how mutations affect the priority levels in the graph. Mutations determine the type on a line (to the right) before any type assignment. In particular, some ancestral lines are not relevant after a mutation. Due to the specific ordering of the lines in the 'lookdown'-ASG, we can identify and then prune these lines. Doing so requires some more notation. At a given time, the line that is the ancestral line if all other relevant potential ancestral lines at that time are assigned type 1 is called the *immune line*. Such a line is always present, because it is always relevant. If a deleterious mutation occurs on the immune line, we do not prune it, but assign it the currently largest level. All lines that previously were on a higher level decrease their level by one to fill the gap. This explains the name: the immune line is immune to pruning by deleterious mutations. A line that is not the immune line will, by construction, play at some point the role of an incoming line. A deleterious mutation on such a line leads to an unfit type so that this line is not successful at the selection event. Hence, the line is not any more among the relevant potential ancestors and we remove it from the current set. The lines having a higher level decrease their level by one to fill the gap. On the other hand, a line that is not the immune line is incoming to all lines with a higher level. A beneficial mutation on such a line therefore deprives all lines with a larger level of the possibility to be relevant. Therefore, we prune all lines above the beneficial mutation. The immune line is now the line with the mutation. See Fig. 2.7 for a realisation of the pLD-ASG.
It turns out that for our purposes it will be enough to consider only the line-counting process of the pLD-ASG, which is the content of the following definition. For a rigorous definition of the pLD-ASG as a graph structure, we refer to [Cor17a, Sect. 4.6].

Definition 2.21 (Line-counting process of pLD-ASG for finite populations)**.** The line-counting process of the pLD-ASG, denoted by $L = (L_r)_{r \geq 0}$, is the continuous-time Markov chain on $[N]$ with transition rates

$$\begin{aligned} q_L(n, n+1) &= sn\,\frac{N-n}{N}, \\ q_L(n, n-1) &= n\,\frac{n-1}{N} + (n-1)u\nu_1 + u\nu_0\,\mathbb{1}_{\{n>1\}}, \\ q_L(n, j) &= u\nu_0, && j \in [n-2]. \end{aligned}$$

If $s > 0$, L is irreducible and L_r converges in distribution to its stationary distribution. In this case, we denote by L_∞ a random variable distributed according to the stationary distribution of L. If $s = 0$, then set $L_\infty = 1$ a.s. unless stated otherwise.

Remark 2.10. For later use, we do not insist on starting the pLD-ASG from a single individual; but one should keep in mind that if we start the process with $n > 1$ lines, then L does not correctly describe the number of relevant potential ancestors of a sample of n individuals. For example, assume that the first event is a beneficial mutation on the line on level 1. This induces the pruning of all other lines, which does not properley reflect the ancestry of relevant potential ancestors of the n individuals. ♢

The reason for the entire construction is the following. Consider a given realisation of the pLD-ASG from time 0 up to (backward) time r and a given assignment of types to the leaves. The true ancestral line is then the lowest line that has been assigned a type 0 in which case the individual at time 0 has a fit ancestor at time r. If no type 0 is assigned to the leaves of the pLD-ASG at time r, the true ancestral line is the immune line and the ancestor of the individual at time 0 has an ancestor of type 1 at time r. This leads to the following (known) connection, which goes back to [Cor17a, Prop. 4.6, Prop. 4.7] (see [Len+15, Thm. 4] for the corresponding earlier result in the weak mutation–weak selection limit), between the line-counting process of the pLD-ASG and the ancestral type distribution.

Theorem 2.22 (Representation of ancestral type distribution)**.** *We have*

$$h_r(k) = \mathbb{E}_1\left[\frac{k^{\underline{L_r}}}{N^{\underline{L_r}}}\right], \qquad k \in [N]_0. \tag{2.22}$$

In particular, $h_\infty(k) := \lim_{r\to\infty} h_r(k)$ *exists and is given by*

$$h_\infty(k) = \mathbb{E}_1\left[\frac{k^{\underline{L_\infty}}}{N^{\underline{L_\infty}}}\right], \qquad k \in [N]_0. \tag{2.23}$$

Proof. The proof is a direct consequence of [Cor17a, Prop. 4.6, Prop. 4.7]. □

Remark 2.11. If $u = 0$, then $L_\infty = R_\infty$. In this parameter regime one type fixates. In particular, the probability of the common ancestor of the population in the distant future to be unfit at present coincides with the fixation probability of the unfit type at present. ♢

The above representation of the ancestral type distribution is consistent with the graphical picture. The ancestral type at backward time r is 1 if and only if all L_r lines are of type 1.

The corresponding probability is given by $\mathbb{E}_1[k^{\underline{L_r}}/N^{\underline{L_r}}]$. Moreover, note that equation (2.22) can be rewritten as

$$h_r(k) = \mathbb{E}_1\left[\frac{k^{\underline{L_r}}}{N^{\underline{L_r}}}\right] = 1 - \frac{N-k}{N}\sum_{n=0}^{N-1}\mathbb{P}_1(L_r > n)\,\frac{k^{\underline{n}}}{(N-1)^{\underline{n}}}. \tag{2.24}$$

This means that we can partition the event of a beneficial ancestor according to the first level occupied by a type-0 individual. Namely,

$$\mathbb{P}_1(L_r > n)\,\frac{N-k}{N}\,\frac{k^{\underline{n}}}{(N-1)^{\underline{n}}}$$

is the probability that at least $n+1$ lines are present, the $(n+1)^{st}$ line is of type 0, and the first n lines are of type 1. Summing this probability over n gives the probability of a fit ancestor at backward time r.

Remark 2.12. Let us provide some historic remarks. Fearnhead [Fea02, Thm. 3] was the first to obtain a rigorous characterisation of the common ancestor type distribution at equilibrium in the weak mutation–weak selection limit. His approach is also based on the ancestral selection graph. He finds a power series expansion with coefficients satisfying what is now called the Fearnhead-recursion. Taylor [Tay07, Lem. 4.1] rederived Fearnhead's results by an approach via the structured coalescent of Barton et al. [BES04] (still in the weak mutation–weak selection limit). [Len+15, Thm. 5] derived the result by probabilistic means in terms of the pLD-ASG. In particular, they interpret the coefficients of Fearnhead's original representation as tail probabilities. The finite population case was first treated by Kluth et al. [KHB13]. Cordero [Cor17a] later rederived their results by means of the analogue to the pLD-ASG in the finite population setting. ♢

2.4.2 Tail probabilities of line-counting process of pLD-ASG and Siegmund duality

The representation of the ancestral type distribution in (2.24) illustrates the importance of the tail probabilities of L. We abbreviate the tail probabilities as

$$a_n := \mathbb{P}(L_\infty > n), \tag{2.25}$$

and refer to them also as Fearnhead-coefficients. In this section, we derive a recursion for these tail probabilities. This is the so-called Fearnhead's recursion in the finite population setting. Our derivation of the recursion is based on a duality result in Proposition 2.25.

Proposition 2.23 (Feanhead's recursion in the Moran model)**.** *Let a_n be defined as in* (2.25)*. They are the unique solution of the following system of equations. For $n \in [N-1]$, we have*

$$\Big(\frac{n+1}{N} + u\nu_1\Big)a_{n+1} = \Big(\frac{n+1}{N} + s\,\frac{N-n}{N} + u\Big)a_n - s\,\frac{N-n}{N}a_{n-1} \tag{2.26}$$

complemented by $a_0 = 1$ and $a_N = 0$.

Remark 2.13. Recursion (2.26) appears first in [KHB13, Thm. 2] as a finite population analogue to Fearnhead's recursion [Fea02, Thm. 3]. [Cor17a, Prop. 4.7] connects the recursion to the tail probabilities of L_∞. [CM19, Thm. 4.1] solved the recursion via a generating function approach. ♢

Our proof of Proposition 2.23 differs from the one of Cordero [Cor17a, Prop. 4.7], who also uses the pLD-ASG to prove the recursion. In the weak mutation–weak selection limit of the Moran model, Baake et al. [BLW16, Lem. 4.1] relate the tail probabilities of L_∞ to the absorption probabilities of another process. We establish such a relation for the finite population setting. Eventually, this enables us to prove Proposition 2.23. But first, let us rigorously define the said process.

Definition 2.24. Let $D = (D_t)_{t\geq 0}$ be the continuous-time Markov chain with values in $[N+1]$ and transition rates,

$$\begin{aligned} q_D(d,d+1) &= \mathbb{1}_{\{d\leq N\}}\Big[d\,\frac{d-1}{N} + (d-1)u\nu_1\Big], \\ q_D(d,d-1) &= s(d-1)\,\frac{N-(d-1)}{N}, \\ q_D(d,N+1) &= (d-1)u\nu_0. \end{aligned}$$

This process absorbs in either 1 or $N+1$.

Proposition 2.25 (Siegmund duality). *The processes D and L are Siegmund-dual, i. e. for $t \geq 0$*

$$\mathbb{P}(d \leq L_t \mid L_0 = n) = \mathbb{P}(D_t \leq n \mid D_0 = d), \qquad \forall n \in [N],\, d \in [N+1]. \tag{2.27}$$

Remark 2.14. An analogous result was proved in the weak mutation–weak selection limit by [BLW16, Lem. 4.5]. If $s = u = 0$, the relation reduces to a duality between Kingman's coalescent and the fixation line, which was introduced by Pfaffelhuber and Wakolbinger [PW06] (see also [Hén15]). ♢

Proof of Proposition 2.25. In [BLW16, Lem. 4.5] the analogous result in the weak mutation–weak selection limit is proved via Clifford-Sudburry flights [CS85]. Here, we provide a generator proof. We denote the infinitesimal generators of L and D by A_L and A_D, respectively, given by

$$\begin{aligned} A_L f(n) &= sn\,\frac{N-n}{N}\,[f(n+1)-f(n)] \\ &\quad + \Big(n\,\frac{n-1}{N} + (n-1)u\nu_1 + u\nu_0\,\mathbb{1}_{\{n>1\}}\Big)\,[f(n-1)-f(n)] \\ &\quad + \mathbb{1}_{\{n>1\}}u\nu_0\sum_{i=1}^{n-2}[f(i)-f(n)], \\ A_D\tilde{f}(d) &= \Big(d\,\frac{d-1}{N} + (d-1)u\nu_1\Big)[\tilde{f}(d+1)-\tilde{f}(d)] \\ &\quad + s(d-1)\,\frac{N-(d-1)}{N}\,[\tilde{f}(d-1)-\tilde{f}(d)] \\ &\quad + (d-1)u\nu_0\,[\tilde{f}(N+1)-\tilde{f}(d)], \end{aligned}$$

where $f : [N] \to \mathbb{R}$ and $\tilde{f} : [N+1] \to \mathbb{R}$. Set $\bar{H}(n,d) = \mathbb{1}_{\{d \leq n\}}$. Since the state space of L is finite, every function is in the domain of A_L. In particular, $\bar{H}(\cdot,d)(n)$ and $P_t^L\bar{H}(\cdot,d)(n)$ lie in the domain of A_L, where $(P_t^L)_{t\geq 0}$ is the transition semigroup corresponding to L. Similarly, $\bar{H}(n,\cdot)(d)$ and $P_t^D\bar{H}(n,\cdot)(d)$ lie in the domain of A_D, where $(P_t^D)_{t\geq 0}$ is the transition semigroup corresponding to D. We now show that $A_L\bar{H}(\cdot,d)(n) = A_D\bar{H}(n,\cdot)(d)$. The result then follows once more as an application of [JK14, Prop. 1.2]. First note that

$$\sum_{i=1}^{n-1} \mathbb{1}_{\{d \leq n < d+i\}} = \begin{cases} d-1, & \text{if } n \geq d > 1, \\ 0, & \text{else.} \end{cases}$$

Hence,

$$\begin{aligned}
&A_L\bar{H}(\cdot,d)(n) \\
&= sn\,\frac{N-n}{N}\,\mathbb{1}_{\{d=n+1\}} - \Big(n\,\frac{n-1}{N} + (n-1)u\nu_1\Big)\mathbb{1}_{\{d=n\}} - u\nu_0\sum_{i=1}^{n-1}\mathbb{1}_{\{d\leq n<d+i\}} \\
&= s(d-1)\frac{N-(d-1)}{N}\mathbb{1}_{\{d-1=n\}} - \Big(d\frac{d-1}{N} + (d-1)u\nu_1\Big)\mathbb{1}_{\{d=n\}} - (d-1)u\nu_0\mathbb{1}_{\{d\leq n\}} \\
&= A_D\bar{H}(n,\cdot)(d) \qquad \square
\end{aligned}$$

This yields another way to represent the tail-probabilities of L at stationarity.

Corollary 2.26 (Tail probabilities as absorption probabilities).

$$\mathbb{P}\Big(\lim_{t\to\infty} D_t = 1 \mid D_0 = n+1\Big) = a_n, \qquad \forall\ n \in [N]_0. \tag{2.28}$$

Proof. Setting $n = 1$ in Proposition 2.25 and taking $t \to \infty$ yields the result. □

We are now ready to provide a proof of Proposition 2.23.

Proof of Proposition 2.23. By Corollary 2.26, $a_n = \mathbb{P}(\lim_{t\to\infty} D_t = 1 \mid D_0 = n+1)$, i. e. the nth Fearnhead-coefficient corresponds to the probability of D, started in $n+1$, to absorb in 1. Clearly, $a_0 = 1$ and $a_N = 0$. For $n \in [N-1]$, a first-step decomposition of the absorption probability yields

$$\left[n\,\frac{n+1}{N} + sn\,\frac{N-n}{N} + nu\right]a_n = \left[n\,\frac{n+1}{N} + nu\nu_1\right]a_{n+1} + sn\,\frac{N-n}{N}\,a_{n-1}.$$

Dividing by n leads to the claim. Note that the system of equations (2.26) can be written in matrix form. Let $B = (b_{ij}) \in \mathbb{R}^{(N-1)\times(N-1)}$ be this matrix. B is a tridiagonal matrix that is weakly chained diagonally dominant, i.e. $|b_{ii}| \geq \sum_{j\neq i}|b_{ij}|$ and for each $i \in [N-1]$ for which this inequality is not strict there is a sequence of nonzero elements of B of the form $b_{ii_1}, b_{i_1i_2}, \ldots, b_{i_rk}$ with $b_{kk} > \sum_{j\neq k}|b_{kj}|$. It follows by [SC74] that B is nonsingular. In particular, the solution of the recursion with the boundary conditions is unique. □

2.4.3 Properties of the common ancestor type distribution

In this section, we first derive a monotonicity property of h_∞ in ν_0. Finally, we consider the common ancestor type distribution in a population at equilibrium.

The next proposition sheds light on the effect of ν_0 on $h_\infty(k)$. To stress the dependence, we (temporarily) write $h_\infty(k,\nu_0)$. The following result is heuristically proved in [Len+15, Sect. 6] in the diffusive setting via a coupling argument. We formalise this argument in our proof.

Proposition 2.27 (Monotonicity of h_∞). *Let $k \in [N-1]$. If $s,u > 0$ and $\nu_0, \bar{\nu}_0 \in [0,1]$ with $\nu_0 < \bar{\nu}_0$,*

$$h_\infty(k,\nu_0) < h_\infty(k,\bar{\nu}_0).$$

Proof. Fix $s,u > 0$, $k \in [N-1]$, and $\nu_0, \bar{\nu}_0 \in [0,1]$ with $\nu_0 < \bar{\nu}_0$. Furthermore, we set $\nu_1 = 1-\nu_0$ and $\bar{\nu}_1 = 1-\bar{\nu}_0$. Clearly, $\nu_1 > \bar{\nu}_1$. Write L and $\bar{L}$ for the line-counting processes of the pLD-ASG with the same branching parameter s, but beneficial mutation rate $u\nu_0$ and $u\bar{\nu}_0$, respectively (deleterious mutation rate $u\nu_1$ and $u\bar{\nu}_1$, respectively). The two processes can be coupled such that $\bar{L}_t \leq L_t$ for all $t > 0$ if $\bar{L}_0 \leq L_0$ as follows. Assume $L_{r\text{-}} = n$ and $\bar{L}_{r\text{-}} = \bar{n}$ with $\bar{n} \leq n$.
If L increases at time r to $n+1$ due to a selective event, then

$$\bar{L}_r = \begin{cases} \bar{n}+1, & \text{with prob. } \frac{\bar{n}(N-\bar{n})}{n(N-n)}, \\ \bar{n} & \text{with prob. } 1-\frac{\bar{n}(N-\bar{n})}{n(N-n)}. \end{cases}$$

If L decreases at time r to $n-1$ due to a neutral reproduction event, then

$$\bar{L}_r = \begin{cases} \bar{n}-1, & \text{with prob. } \frac{\bar{n}(\bar{n}-1)}{n(n-1)}, \\ \bar{n} & \text{with prob. } 1-\frac{\bar{n}(\bar{n}-1)}{n(n-1)}. \end{cases}$$

If L decreases to i due to a beneficial mutation, then $\bar{L}_r := \bar{n} \wedge i$.
If L decreases at time r to $n-1$ due to a deleterious mutation, then

$$\bar{L}_r = \begin{cases} \bar{n}-1, & \text{with prob. } \frac{\bar{\nu}_1}{\nu_1}\frac{\bar{n}-1}{n-1}, \\ i, & \text{with prob. } \frac{\nu_1-\bar{\nu}_1}{\nu_1}\frac{1}{n-1}, \quad \text{for } i \in [\bar{n}-1], \\ \bar{n} & \text{with prob. } 1-\frac{\bar{n}-1}{n-1}. \end{cases}$$

For all $k \in [N]$, by construction, $\mathbb{P}(L_r > k) \geq \mathbb{P}(\bar{L}_r > k)$. Taking the limit $r \to \infty$ leads to $a_k \geq \bar{a}_k$. Assume that $a_k = \bar{a}_k$ for all k. The tail probabilities solve Fearnhead's recursion, which, by Proposition 2.23, has a unique solution. Hence, (2.26) leads to $(\nu_1 - \bar{\nu}_1)a_n = 0$ and therefore $a_n = 0$ for $n \in [N]\backslash\{1\}$. It follows that $a_1 = s(N-1)/(2+s(N-1)+uN)$. But (2.26) with $n=2$ leads to a contraction, because $a_1 > 0$ for $s > 0$. Hence, there exists $k \in [N]$ such that $a_k > \bar{a}_k$. In particular, using (2.24), $h_\infty(k,\nu_0) < h_\infty(k,\bar{\nu}_0)$ for $k \in [N-1]$. □

Finally, we consider the common ancestor type distribution in a population at equilibrium.

Proposition 2.28. *If* $s, u, \nu_0, \nu_1 > 0$*, let* Y_∞ *be a random variable on* $[N]_0$ *with distribution* π_Y*. Then*

$$\mathbb{E}[h_\infty(Y_\infty)] = \mathbb{E}[w_R(L_\infty)]. \tag{2.29}$$

Proof. Equation (2.23), interchanging the two finite sums, and Corollary 2.17 lead to

$$\mathbb{E}[h_\infty(Y_\infty)] = \sum_{k=0}^{N}\sum_{n=1}^{N} \mathbb{P}(L_\infty = n)\mathbb{P}(Y_\infty = k)\frac{k^{\underline{n}}}{N^{\underline{n}}} = \mathbb{E}\left[w_R(L_\infty)\right]. \qquad \square$$

Proposition 2.28 is consistent with the results from Section 2.3. For an unfit common ancestor in a population at equilibrium, the killed ASG that arises from the lines of the pLD-ASG at stationarity needs to absorb in 0 so that all relevant potential ancestors are indeed unfit.

2.4.4 An absorbing Markov chain and a factorial moment duality

In this section, we provide an alternative representation of the common ancestor type distribution in terms of the absorption probability of a certain Markov chain. We connect the latter process to the line-counting process of the pLD-ASG via duality in Theorem 2.29.

In the weak mutation–weak selection limit of the Moran model, Taylor [Tay07, Eq. (11)] shows that the common ancestor type distribution, conditional on the initial type configuration, is a harmonic function for the infinitesimal generator of a jump-diffusion process. The finite population analogue of this jump-diffusion is the following continuous-time Markov chain $\tilde{Y} := (\tilde{Y}_t)_{t\geq 0}$ with values in $[N]_0$. The transition rates are given by

$$q_{\tilde{Y}}(\ell, \ell+1) = \ell\,\frac{N-\ell}{N} + (N-\ell)\,\frac{\ell}{\ell+1}\,u\nu_1, \qquad q_{\tilde{Y}}(\ell, N) = \frac{\ell}{N-(\ell-1)}\,u\nu_0,$$
$$q_{\tilde{Y}}(\ell, \ell-1) = \ell\,\frac{N-\ell}{N}\,(1+s) + \ell\,\frac{N-\ell}{N-(\ell-1)}\,u\nu_0, \qquad q_{\tilde{Y}}(\ell, 0) = \frac{N-\ell}{\ell+1}\,u\nu_1,$$

for $\ell \in [N]_0$. Note that the states 0 and N are absorbing.
Let us first connect $\tilde{Y}$ to the Moran model. Define $\widetilde{T}_0 := \inf\{t \geq 0 : \tilde{Y}_t = 0\}$ and $\widetilde{T}_N := \inf\{t \geq 0 : \tilde{Y}_t = N\}$, i.e. the first time $\tilde{Y}$ hits 0 and N, respectively. Set $\widetilde{T} := \min\{\widetilde{T}_0, \widetilde{T}_N\}$. We can couple Y and $\tilde{Y}$ on the basis of the graphical representation such that for $\tilde{Y}_0 = Y_0 \in [N]_0$, we have $Y_t = \tilde{Y}_t$ for $t \leq \widetilde{T}$. Consider an initial type configuration $c \in \{0,1\}^N$ with $|c| = Y_0 = \tilde{Y}_0$. Assume $Y_{t-} := \ell$ for some $\ell \in [N]_0$. If $\tilde{Y}_{t-} \in \{0, N\}$, then $\tilde{Y}_{t-} = \tilde{Y}_t$. Now, assume $\tilde{Y}_{t-} \notin \{0, N\}$. If Y jumps at time t to $\ell+1$ (resp. to $\ell-1$) due to a reproduction event, then set $\tilde{Y}_t = \tilde{Y}_{t-} + 1$ (resp. $\tilde{Y}_t = \tilde{Y}_{t-} - 1$). If Y jumps at time t to $\ell+1$ due to a deleterious mutation, then set

$$\tilde{Y}_t := \begin{cases} \tilde{Y}_{t-} + 1, & \text{with probability } \ell/(\ell+1), \\ 0, & \text{with probability } 1/(\ell+1). \end{cases}$$

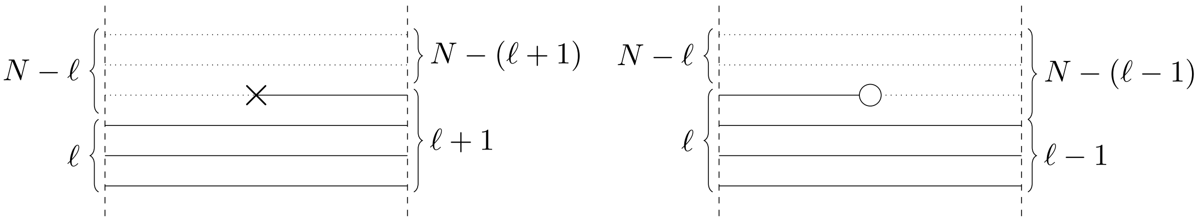


Figure 2.8. A type changing deleterious mutation (resp. beneficial mutation) on a given line changes the type distribution in the Moran model from $(N-\ell,\ell)$ to $(N-(\ell+1),\ell+1)$ (resp. to $(N-(\ell-1),\ell-1)$) and is depicted on the left (resp. right). At each type changing deleterious mutation (resp. beneficial mutation) flip a coin with frequency-dependent success probability $1/(\ell+1)$ (resp. $1/(N-\ell+1)$). If the first success is due to a deleterious mutation (beneficial mutation) $\tilde{Y}$ jumps to 0 (resp. to N). A solid (dotted) line corresponds to an unfit (fit) individual.

If Y jumps at time t to $\ell-1$ due to a beneficial mutation, then set

$$\tilde{Y}_t := \begin{cases} \tilde{Y}_{t-}-1, & \text{with probability } (N-\ell)/(N-\ell+1), \\ N, & \text{with probability } 1/(N-\ell+1). \end{cases}$$

In particular, the only events that make $\tilde{Y}$ to differ from Y are type changing mutations. See also Fig. 2.8.

Now, we strengthen the result of Taylor [Tay07, Eq. (11)] and show that $\tilde{Y}$ is dual to L.

Theorem 2.29 (Duality). *The processes $\tilde{Y}$ and L are dual with respect to the duality function $H(\ell,n)$ (from* (2.10)*), that is, for $t\geq 0$,*

$$\mathbb{E}_\ell[H(\tilde{Y}_t,n)] = \mathbb{E}_n[H(\ell,L_t)], \qquad \forall \ell\in[N]_0,\ n\in[N]. \tag{2.30}$$

Before proving this result, let us provide the following *heuristic* sketch to motivate the duality. We follow closely the interpretation of the diffusion-limit analogue of the absorbing Markov chain in terms of the structured coalescent as it is given in Lenz et al. [Len+15, Sect. 7.1] (see also [Tay07, p. 821]). This provides some intuition why the duality should hold (even though these authors do not mention a duality).

We start with an observation regarding the rates at which $\tilde{Y}$ jumps to the boundary. Let $J_r^{t+\delta}$ be the type of the ancestor at (forward) time r of an individual uniformly sampled at time $t+\delta$ in the Moran model in $[0,t+\delta]$ with initial type distribution $Y_0=l_0\in[N-1]$. Let $M_{t,\delta}^{\times}$ be the event of exactly one type changing deleterious mutation in $[t,t+\delta]$ in the Moran model (and no other event), where $\delta>0$. Denote by $(\mathcal{F}_r^Y)_{r\geq 0}$ the natural filtration of Y. Note that then

$$\mathbb{P}(J_t^{t+\delta}=0 \mid J_{t+\delta}^{t+\delta}=1,\mathcal{F}_t,M_{t,\delta}^{\times}) = \frac{\mathbb{P}(J_t^{t+\delta}=0,J_{t+\delta}^{t+\delta}=1 \mid \mathcal{F}_t,M_{t,\delta}^{\times})}{\mathbb{P}(J_{t+\delta}^{t+\delta}=1 \mid \mathcal{F}_t,M_{t,\delta}^{\times})} = \frac{1}{Y_t+1},$$

i.e. given an individual that is unfit at time $t+\delta$, $\mathcal{F}_t$, and a type changing deleterious mutation occurs in a small time horizon δ, the probability that it is the ancestral line of the picked individual that experiences a type change in $[t, t+\delta]$ is $1/(Y_t+1)$. Similarly,

$$\mathbb{P}(J_t^{t+\delta} = 1 \mid J_{t+\delta}^{t+\delta} = 1, \mathcal{F}_t, M_{t,\delta}^{\times}) = \frac{Y_t}{Y_t+1}.$$

In addition, the rate at which a type changing deleterious mutations occurs in the Moran model is $(N - Y_t)u\nu_1$. Altogether, we would expect a type changing deleterious mutation on the ancestral line at time t of an individual that is unfit at time $t+\delta$ to occur at rate $(N-Y_t)u\nu_1/(Y_t+1)$. By the same argument, we expect a type changing beneficial mutation on the ancestral line at time t of an individual that is chosen at time $t+\delta$ and that is unfit to occur at rate $Y_t u\nu_0/(N-Y_t+1)$. The probability that a type changing beneficial mutation at time t does not occur on the ancestral line of a fit individual picked at time $t+\delta$ is $(N-Y_t)/(N-Y_t+1)$. This suggests that, given $\mathcal{F}_t$, the type along an ancestral line changes according to the above rates if the type process along the ancestral line would be a Markov process.

Having this in mind, consider, for $t > 0$ and conditional on Y, the jump process $\hat{J}^t := (\hat{J}_{t-r}^t)_{r\in[0,t]}$, i.e. it starts at time t and ends at 0, with state space $\{0,1\}$ and time-inhomogeneous jump rates given by

$$q_{\hat{J}_{t-r}^t}(1,0) = \frac{N-Y_{t-r}}{Y_{t-r}+1}u\nu_1 \quad \text{and} \quad q_{\hat{J}_{t-r}^t}(0,1) = \frac{Y_{t-r}}{N-Y_{t-r}+1}u\nu_0,$$

where $r \in [0,t]$. We can construct this process on the basis of two Poisson point processes $\Pi^{0,Y}$ and $\Pi^{1,Y}$ on the positive real line with inhomogeneous arrival rates given by, respectively,

$$\frac{N-Y_\tau}{Y_\tau+1}u\nu_1 \quad \text{and} \quad \frac{Y_\tau}{N-Y_\tau+1}u\nu_0,$$

for $\tau > 0$. For any point $\tau_0 \in [0,t]$ of $\Pi^{0,Y}$, $\hat{J}^t$ jumps to 0 at time τ_0. If $\hat{J}^t$ was already in 0, it remains there. On the other hand, for any arrival time $\tau_1 \in [0,t]$ of $\Pi^{1,Y}$, $\hat{J}^t$ jumps to 1 at time τ_1; and if it already was in 1, it remains there. Now, let $\hat{T}_0$ and $\hat{T}_1$ be the arrival time of $\Pi^{0,Y}$ and $\Pi^{1,Y}$, respectively, that is closest to 0. Set $\hat{T} := \min\{\hat{T}_0, \hat{T}_1\}$. If $\hat{T} = \hat{T}_1$, then $\hat{J}_0^t = 1$; otherwise, $\hat{J}_0^t = 0$. $\hat{J}^t$ is constructed in such a way that its distribution agrees with the one that we claim is the one of J^t, i.e. the type process along the ancestral line of a picked individual at time t. In addition, given Y, points of $\Pi^{0,Y}$ and $\Pi^{1,Y}$ arrive at the same rate at which $\tilde{Y}$ jumps to N and 0, respectively. Hence, $((Y_t)_{0\leq t<\hat{T}}, N\mathbb{1}_{\{\hat{T}=\hat{T}_1\}}) \overset{d}{=} (\tilde{Y}_{0\leq t<\widetilde{T}}, \tilde{Y}_{\widetilde{T}})$. In particular, conditional on $\{\tilde{Y}_t \notin \{0,N\}\}$, the probability of an unfit ancestral type at time 0 agrees with the probability of sampling an unfit individual at time t, which is Y_t/N. The event of absorption of $\tilde{Y}$ in 0 or N has the same distribution as the last jump (before time 0) on the ancestral line of the sampled individual being to 0 or 1, respectively. This leads to the representation $h_r(\ell_0) = \mathbb{E}_{\ell_0}[\tilde{Y}_r/N]$, which is the special case of the duality for $n=1$.

Note that to turn this heuristic into a complete proof, we need to show that $(J_{t-r}^t)_{r\in[0,t]}$ is a Markov process, $(J_{t-r}^t)_{r\in[0,t]} \overset{d}{=} (\hat{J}_{t-r}^t)_{r\in[0,t]}$, and then extend the argument for $n > 1$. It what follows, we take a more analytic route to prove the duality. We require the following auxiliary lemma.

Lemma 2.30 (Auxiliary lemma). *For $N, \ell, n \in \mathbb{N}$,*

$$\sum_{i=1}^{n-1} \binom{N-i}{N-\ell} = \binom{N}{N-\ell+1} - \binom{N-n+1}{N-\ell+1}. \tag{2.31}$$

Proof. It is classical that for $k, n \in \mathbb{N}$ the following identity holds,

$$\sum_{m=0}^{n} \binom{m}{k} = \binom{n+1}{k+1}.$$

Hence,

$$\sum_{j=1}^{n-1} \binom{N-j}{N-\ell} = \sum_{j=0}^{N-1} \binom{j}{N-\ell} - \sum_{j=0}^{N-n} \binom{j}{N-\ell} = \binom{N}{N-\ell+1} - \binom{N-n+1}{N-\ell+1}. \qquad \square$$

Proof of Theorem 2.29. The infinitesimal generator of $\tilde{Y}$ acts on functions $f : [N]_0 \to \mathbb{R}$ and is given by $A_{\tilde{Y}} = A_Y^d + A_Y^s + A_{\tilde{Y}}^{\nu_0} + A_{\tilde{Y}}^{\nu_1}$, with A_Y^d and A_Y^s of (2.12) and (2.13), respectively, and

$$\begin{aligned} A_{\tilde{Y}}^{\nu_0} f(\ell) &:= \ell \, \frac{N-\ell}{N-\ell+1} \, u\nu_0 \, [f(\ell-1) - f(\ell)] + \frac{\ell}{N-\ell+1} \, u\nu_0 \, [f(N) - f(\ell)], \\ A_{\tilde{Y}}^{\nu_1} f(\ell) &:= (N-\ell) \, \frac{\ell}{\ell+1} \, u\nu_1 \, [f(\ell+1) - f(\ell)] + \frac{N-\ell}{\ell+1} \, u\nu_1 \, [f(0) - f(\ell)]. \end{aligned}$$

The infinitesimal generator of L acts on functions $\tilde{f} : [N] \to \mathbb{R}$ and is given by $A_L = A_R^d + A_R^s + A_L^{\nu_0} + A_L^{\nu_1}$, with A_R^d and A_R^s of (2.15) and (2.16), respectively, and

$$A_L^{\nu_0} \tilde{f}(n) := u\nu_0 \sum_{j=1}^{n-1} [\tilde{f}(j) - \tilde{f}(n)], \qquad A_L^{\nu_1} \tilde{f}(n) := u\nu_1 \, (n-1) \, [\tilde{f}(n-1) - \tilde{f}(n)].$$

We want to apply [JK14, Prop. 1.2]. Since the state space of $\tilde{Y}$ is finite, every function is in the domain of $A_{\tilde{Y}}$. In particular, $H(\cdot, n)(\ell)$ and $P_t^{\tilde{Y}} H(\cdot, n)(\ell)$ lie in the domain of $A_{\tilde{Y}}$, where $(P_t^{\tilde{Y}})_{t \geq 0}$ is the transition semigroup corresponding to $\tilde{Y}$. Similarly, $H(\ell, \cdot)(n)$ and $P_t^L H(\ell, \cdot)(n)$ lie in the domain of A_L, where $(P_t^L)_{t \geq 0}$ is the transition semigroup corresponding to L. In the proof of Theorem 2.14, we already showed that

$$A_Y^d H(\cdot, n)(\ell) = A_R^d H(\ell, \cdot)(n) \quad \text{and} \quad A_Y^s H(\cdot, n)(\ell) = A_R^s H(\ell, \cdot)(n).$$

Hence, it suffices to check that

$$A_{\tilde{Y}}^{\nu_0} H(\cdot, n)(\ell) = A_L^{\nu_0} H(\ell, \cdot)(n) \quad \text{and} \quad A_{\tilde{Y}}^{\nu_1} H(\cdot, n)(\ell) = A_L^{\nu_1} H(\ell, \cdot)(n),$$

which then implies

$$A_{\tilde{Y}} H(\cdot, n)(\ell) = A_L H(\ell, \cdot)(n), \qquad \forall \ell \in [N]_0, \, n \in [N].$$

Indeed, for the part corresponding to the type-1 mutation we obtain

$$\begin{aligned}
A_L^{\nu_1} H(\ell,\cdot)(n) &= (n-1)u\nu_1\big(N-n+1-(\ell-n+1)\big)\frac{\ell^{\underline{n-1}}}{N^{\underline{n}}} \\
&= u\nu_1(N-\ell)\,\frac{\ell-(\ell-n+1)}{\ell+1}\,\frac{(\ell+1)^{\underline{n}}}{N^{\underline{n}}} \\
&= u\nu_1(N-\ell)\,\frac{\ell}{\ell+1}\left[\frac{(\ell+1)^{\underline{n}}}{N^{\underline{n}}}-\frac{\ell^{\underline{n}}}{N^{\underline{n}}}\right]+u\nu_1\,\frac{N-\ell}{\ell+1}\left[-\frac{\ell^{\underline{n}}}{N^{\underline{n}}}\right] \\
&= A_{\tilde{Y}}^{\nu_1} H(\cdot,n)(\ell).
\end{aligned}$$

For the part associated to mutation to type 0, note that,

$$A_L^{\nu_0} H(\ell,\cdot)(n) = u\nu_0\sum_{j=1}^{n-1}\left[\frac{\ell^{\underline{j}}}{N^{\underline{j}}}-\frac{\ell^{\underline{n}}}{N^{\underline{n}}}\right] = u\nu_0\frac{1}{\binom{N}{\ell}}\sum_{j=1}^{n-1}\left[\binom{N-j}{N-\ell}-\binom{N-n}{N-\ell}\right].$$

Hence, we can use Lemma 2.30 and obtain

$$\begin{aligned}
A_L^{\nu_0} H(\ell,\cdot)(n) &= u\nu_0\frac{1}{\binom{N}{\ell}}\left[\binom{N}{N-\ell+1}-\binom{N-n+1}{N-\ell+1}-(n-1)\binom{N-n}{N-\ell}\right] \\
&= \frac{\ell u\nu_0}{N-\ell+1}-\ell u\nu_0\,\frac{(\ell-1)^{\underline{n-1}}}{N^{\underline{n}}}\left(\frac{N-n+1}{N-\ell+1}+(n-1)\right) \\
&= \frac{\ell u\nu_0}{N-\ell+1}-\ell\,\frac{1}{N-\ell+1}\,u\nu_0\,\frac{(\ell-1)^{\underline{n-1}}}{N^{\underline{n}}}\Big((N-\ell)n+\ell\Big) \\
&= \frac{\ell u\nu_0}{N-\ell+1}\left[1-\frac{\ell^{\underline{n}}}{N^{\underline{n}}}\right]+\ell\,\frac{N-\ell}{N-\ell+1}\,u\nu_0\left[\frac{(\ell-1)^{\underline{n}}}{N^{\underline{n}}}-\frac{\ell^{\underline{n}}}{N^{\underline{n}}}\right] \\
&= A_{\tilde{Y}}^{\nu_0} H(\cdot,n)(\ell).
\end{aligned}$$

□

We now obtain a characterisation of h_∞ that does not depend on L by exploiting the duality in Theorem 2.29. This is the analogue to [Tay07, Prop. 2.5] in the finite population setting.

Corollary 2.31 (Common ancestor type distribution as hitting probability). *For $\ell \in [N]_0$, we have*

$$h_\infty(\ell) = \mathbb{P}(\widetilde{T}_N < \infty \mid \tilde{Y}_0 = \ell), \tag{2.32}$$

In particular, h_∞ is the unique solution of the difference equation,

$$\begin{aligned}
&\left[\frac{2+s}{N}+\frac{u\nu_1}{\ell}+\frac{u\nu_0}{N-\ell}\right]h_\infty(\ell) \\
&= \left[\frac{1}{N}+\frac{u\nu_1}{\ell+1}\right]h_\infty(\ell+1)+\left[\frac{1+s}{N}+\frac{u\nu_0}{N-\ell+1}\right]h_\infty(\ell-1)+\frac{u\nu_0}{(N-\ell)(N-\ell+1)},
\end{aligned} \tag{2.33}$$

for $\ell \in [N-1]$; complemented by $h_\infty(0) = 0$ and $h_\infty(N) = 1$.

Proof. Taking $t \to \infty$ in (2.30), together with (2.22) leads to (2.32). The boundary conditions of h_∞ follow also by (2.22). A first-step decomposition of the absorption probability of $\tilde{Y}_t$ in N leads to

$$\begin{aligned}
&\left[\ell\,\frac{N-\ell}{N}(2+s)+(N-\ell)u\nu_1+\ell u\nu_0\right]h_\infty(\ell)\\
&=\left[\ell\,\frac{N-\ell}{N}+(N-\ell)\,\frac{\ell}{\ell+1}\,u\nu_1\right]h_\infty(\ell+1)\\
&+\left[\ell\,\frac{N-\ell}{N}\,(1+s)+\ell\,\frac{N-\ell}{N-(\ell-1)}\,u\nu_0\right]h_\infty(\ell-1)\\
&+\frac{\ell}{N-(\ell-1)}\,u\nu_0, \qquad\qquad (\ell\in[N-1]).
\end{aligned}$$

Dividing by $\ell(N-\ell)$ leads to (2.33). Note that the system of equations (2.33) can be written in matrix form. Let $D=(d_{ij})\in\mathbb{R}^{(N-1)\times(N-1)}$ be the corresponding matrix. D is a strictly diagonally dominant matrix, i.e. $|d_{ii}|>\sum_{j\neq i}|d_{ij}|$. It follows from the Levy-Desplanques theorem (e.g. [JH85, Cor. 5.6.17]) that D nonsingular. In particular, the solution of the recursion with the boundary conditions is unique. □

3

A probabilistic view on the deterministic mutation-selection equation

In this chapter we work out the backward point of view, so far reserved to stochastic models of population genetics, for the deterministic mutation-selection equation. We will work with the simplest model, namely, with haploid individuals, two types, selection, and mutation, and pursue two major aims. First, we obtain a representation of the solution of the deterministic mutation-selection equation and its equilibrium state(s) in terms of an appropriate killed ASG. This relation reveals the ancestral structure behind the bifurcation phenomena of the associated ODE (for an introduction to bifurcation theory see, e.g., [GH83, Ch. 3]). Second, we characterise in this framework the type distribution of the ancestors of today's individuals in the distant past by the pLD-ASG that arises as the large population limit of the pLD-ASG for finite populations (of Section 2.4). Even though the forward process is now deterministic, the ancestral processes remain stochastic and we can still apply the tools and techniques from the previous chapter. This enables us to recover the results from the previous chapter in the new setting by only minor modifications of the proofs.

This chapter is organised as follows. We set out in Section 3.1 to recall the Moran model with two types, selection, and mutation; but *without* pairwise interaction. It leads to the deterministic haploid mutation-selection equation via a law of large numbers. Next, the graphical constructions required to trace back ancestral lines in this setup are introduced; namely, the ASG (Section 3.2), the killed ASG (Section 3.3), and the pLD-ASG (Section 3.4), all in the smss-limit. In the special case of unidirectional mutation (only away from the beneficial type, without back mutation), the results shed new light on the bifurcations related to the so-called error threshold phenomenon. Finally (Section 3.5), we characterise the ancestral type distribution in two ways: first, by means of the pLD-ASG; and second, as the absorption probability of a piecewise-deterministic Markov process.

3.1 Deterministic limit of the Moran model with mutation and selection

Consider the two-type Moran model with mutation and selection. It coincides with the model in Chapter 2 if we set $\gamma = 0$. Let us recall the essentials. We have a haploid population of fixed size N. Each individual in this population has a type, which is either 0 or 1. Individuals of type 1 reproduce at rate 1, whereas individuals of type 0 reproduce at rate $1 + s$ with $s \geq 0$. We refer to type 0 as the fit or beneficial type, whereas type 1 is unfit or deleterious. When an individual reproduces, its single offspring inherits the parent's type and replaces a uniformly-chosen individual in the population, thereby keeping the population size constant. Each individual mutates at rate u; the type after the event is i with probability ν_i, $i \in \{0,1\}$. We assume throughout this chapter that u is positive and $\nu_0, \nu_1 \in [0,1]$ with $\nu_0 + \nu_1 = 1$. The graphical representation of this Moran model coincides with the one of the previous chapter (see Fig. 2.1) if we remove interactive and checking arrows.

Let $Y_t^{(N)}$ be the number of type-1 individuals at time t in a population of size N. It is explained in Section 2.1 that $Y^{(N)} = (Y_t^{(N)})_{t \geq 0}$ is a continuous-time Markov chain on $[N]_0$. In what follows, we study a deterministic limit of the Moran model. Cordero [Cor17b, Prop. 3.1] shows that, if $Y_0^{(N)}/N \longrightarrow y_0$ as $N \to \infty$, then $Y^{(N)}/N$ converges to the solution $y(t; y_0)$ of the initial value problem (IVP)

$$\begin{aligned} \frac{dy}{dt}(t) &= -sy(t)(1-y(t)) - u\nu_0 y(t) + u\nu_1(1-y(t)), \qquad t \geq 0, \\ y(0) &= y_0, \qquad \text{for } y_0 \in [0,1]. \end{aligned} \tag{3.1}$$

More precisely, we have for all $t \geq 0$ and $\varepsilon > 0$

$$\mathbb{P}\left(\sup_{\xi \leq t} \left| \frac{Y_\xi^{(N)}}{N} - y(\xi; y_0) \right| > \varepsilon \right) \xrightarrow{N \to \infty} 0, \tag{3.2}$$

i.e. the convergence is uniform on compact sets of time in probability and a special case of the dynamical law of large numbers of Kurtz [Kur70, Thm. 3.1] (alternatively [EK86, Thm. 11.2.1]). We prove (3.2) in the more general case with pairwise interaction in Proposition 4.6 of Section 4.2. Neither parameters nor time are rescaled. This corresponds to a smss-setting (strong mutation–strong selection). (Note that this in contrast to the usual diffusion limit, where parameters and time are rescaled with population size; this is suitable in a weak mutation–weak selection framework, see, e.g., [Dur08, Ch. 7.2].) We call the limit process the deterministic limit or smss-limit. If $\nu_0 \in (0,1)$, the convergence carries over to $t \to \infty$ in the sense that the stationary distribution of the Moran model converges in distribution to the point measure on $\bar{y}$ as $N \to \infty$ [Cor17b, Cor. 3.1], where $\bar{y}$ is the (unique) stable equilibrium in $[0,1]$ of the ODE. The ODE (3.1) is the classical haploid mutation-selection equation of population genetics [CK56; CK70]. It is a Riccati differential equation with constant coefficients and hence the solution is known explicitly [Cor17b, Lem. 3.1]. The equilibrium points of (3.1) are the solutions of the equation

$$sy^2 - (u+s)y + u\nu_1 = 0. \tag{3.3}$$

The stable equilibrium, known, for example, via [Cor17b, Lem. 3.1], is given by

$$\bar{y} = \begin{cases} \frac{1}{2}\left(1 + \frac{u}{s} - \sqrt{(1-\frac{u}{s})^2 + 4\nu_0 \frac{u}{s}}\right), & s > 0, \\ \nu_1, & s = 0. \end{cases} \tag{3.4}$$

If $s > 0$, there is an additional equilibrium of (3.1), which is unstable (see [Cor17b,

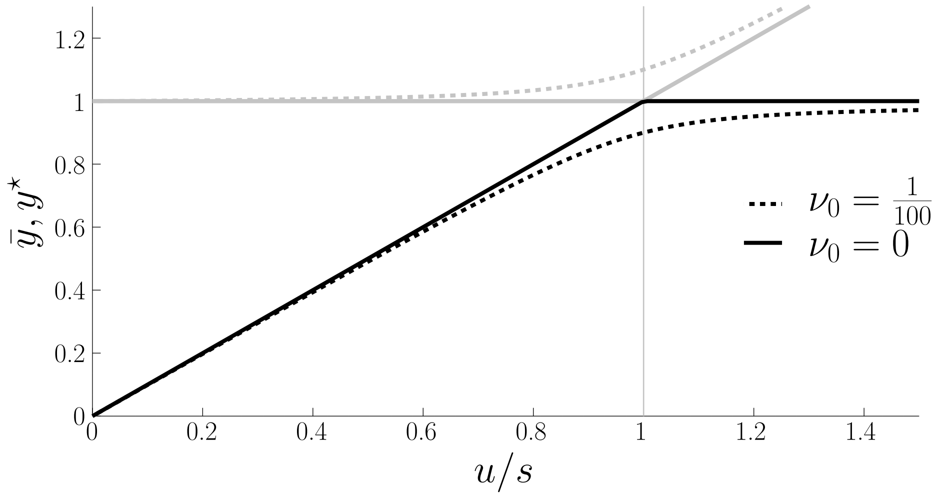


Figure 3.1. The equilibria of (3.1) as a function of u/s for $s > 0$. Black line: $\bar{y}$ (stable); grey line: $y^\star$ (unstable).

Lem. 3.1]), namely,

$$y^\star = \frac{1}{2}\left(1 + \frac{u}{s} + \sqrt{\left(1 - \frac{u}{s}\right)^2 + 4\nu_0\frac{u}{s}}\right).$$

If $\nu_0 > 0$, then $\bar{y} \in [0,1)$ and $y^\star > 1$; so $\bar{y}$ is the only relevant equilibrium. Furthermore, $\lim_{t\to\infty} y(t;y_0) = \bar{y}$ for all $y_0 \in [0,1]$. If $\nu_0 = 0$, the two equilibria reduce to $\bar{y} = \min\{u/s, 1\}$ and $y^\star = \max\{1, u/s\}$. In particular, if $\nu_0 = 0$ and $u \geq s$, then $\bar{y}$ is still the only relevant equilibrium and again $\lim_{t\to\infty} y(t;y_0) = \bar{y}$ for all $y_0 \in [0,1]$. But if $\nu_0 = 0$ and $u < s$, then there are two equilibria in the unit interval. In particular, then $\lim_{t\to\infty} y(t;y_0) = \bar{y}$ for $y_0 \in [0,1)$, while $y(t;1) \equiv 1$.

For $s > 0$, Fig. 3.1 shows how $\bar{y}$ and $y^\star$ depend on u/s and ν_0. The equilibrium frequency $\bar{y}$ is monotonically increasing in u/s; it converges to ν_1 for $u/s \to \infty$, which is also the equilibrium frequency when selection is absent. The case $s > 0$, $\nu_0 = 0$ deserves special attention because of its bifurcation structure. If $u/s < 1$, both $\bar{y} = u/s$ (stable) and $y^\star = 1$ (unstable) are in $[0,1]$; when u surpasses the critical value s, then $\bar{y} = 1$ is the only equilibrium in $[0,1]$, and is attracting for all $y_0 \in [0,1]$. This means that the population loses the fit type altogether; or put differently, the fit type can not persist if $u \geq s$. This phenomenon is known as the error threshold [Eig71; EMS89] and widely discussed in the biological literature. Extending (3.1) to $y_0 \in \mathbb{R}$ yields a *transcritical* (or *exchange of stability*) bifurcation of the equilibria at 1 and u/s: For $u < s$, the former is unstable and the latter is stable; and vice versa for $u > s$. See Baake and Wiehe [BW97] for more details on the equilibrium structure of the mutation-selection equation. (See [GH83, Ch. 3.4] for more details on transcritical bifurcations.) Let us only add here that the equilibrium at u/s in this classical mutation-selection equation with $\nu_0 = 0$ and $u < s$ is used to estimate fitness landscapes from molecular data via appropriate averaging [Zan+17, Sect. 2.2].

Remark 3.1. The term *bifurcation* seems to go back to Poincaré, who described the "splitting" of equilibrium solutions in a family of differential equations [GH83, Ch. 3.1]. Nowadays, the term is used not only for splittings. In general, if the variation of the parameters of a differential equation leads to a qualitative change in the asymptotic behaviour of its solutions, then this phenomenon is called bifurcation. The parameter values at which such a phenomenon occurs are called bifurcation points. ♢

Remark 3.2. Let us stress that we use the word *bifurcation* in two ways in this thesis. On the one hand, it describes the splitting of ancestral lines into two (or more) lines. On the other hand, we use it in the context of dynamical systems in the sense of Remark 3.1. ♢

3.2 ASG in the smss-limit

Setting $\gamma = 0$ in the construction of the ASG in the Moran model from Section 2.2 leads to the ASG in the Moran model without interaction; the only difference being that there are no interactive and checking arrows. In particular, in this setup all lines in the ASG (i. e. all potential influencers) correspond in fact to potential ancestors.

In the smss-limit, the ASG turns into the following construction. Branching, deleterious, and beneficial mutations occur at rate $s, u\nu_1$, and $u\nu_0$ per line, respectively. Since collisions and coalescences (see Section 2.2 for the definition of these notions) occur in the Moran model at rates of order $\mathcal{O}(1/N)$, these types of events vanish as $N \to \infty$. In particular, in the smss-limit, the graphs that contain the potential ancestors of the individuals in a sample evolve independently. It therefore suffices to consider a sample of size 1. The ASG in the smss-limit also admits a representation as a directed graph with labelled vertices, see Section 4.2 for details. The number of potential ancestors evolves like a pure birth process with linear birth rate, which is non-explosive (follows from, e.g., [Bre68, Ch. 15.7]). Hence, for every finite time horizon, the number of lines in the ASG remains finite. In the ASG in the smss-limit we type the potential ancestors at $r = t$ independently and identically according to the initial distribution $(1 - y_0, y_0)$. The notion of type propagation in the ASG for finite populations (see Definition 2.11) carries over to the ASG in the smss-limit if the types are assigned in this way. If we assign to leaf ℓ, site ℓ for $\ell \in \mathbb{N}$, then this allows us to propagate ancestral sites and leads to the notion of ancestry (see Definition 2.11) also in the smss-limit.

3.3 Killed ASG in the smss-limit

Our first aim now is to recover the solution of the deterministic mutation-selection equation (3.1) by genealogical means. We first derive a duality relation between the forward-in-time process and an appropriate ancestral structure. Later, we analyse the long-term behaviour of this ancestral structure.
Recall that the solution $y(t; y_0)$ gives the frequency of type 1 at time t; the smss-limit of the ASG is therefore the appropriate tool. Recall also that, due to the independence of the sampled individuals, it is sufficient to consider a single one.

We will work along the lines of the construction of the killed ASG for finite populations in Section 2.3. We repeat the general idea. Again, our starting point is the well-known

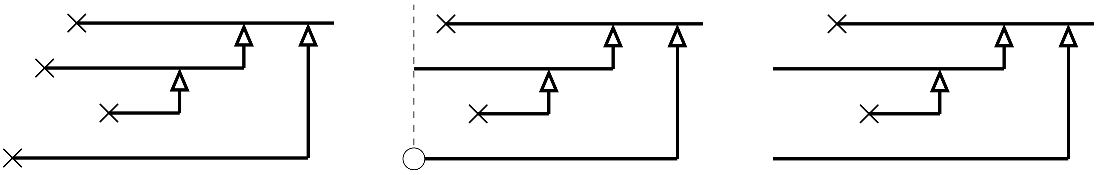


Figure 3.2. The killed ASG either absorbs in a state with 0 lines due to mutations to type 1 (left) or in a cemetery state Δ due to a mutation to type 0 (center); it may also grow to ∞ (not shown). The realisation on the right is still in a transient state.

observation (e.g. [KN97] and [Man09, Thm. 2.1]) that holds for the finite Moran model and carries over to the deterministic setting: In the absence of mutations, a single individual at time t is of type 1 if and only if all its potential ancestors at $t = 0$ are of type 1. This is easily verified via the pecking order (cf. Fig. 2.5). Namely, at every branching event, a type 0 on either the continuing or incoming line suffices for the descendant individual to be of type 0; iterating this over all branching events gives the statement. Mutations add further information about the types: they can determine the type of the sample even before we sample the initial types. More precisely, a mutation to type 1 determines the type of the line (to the right of the mutation) on which it occurs, so this line need not be traced back further into the past; it may be pruned. Next, the first mutation to type 0 (on any line that is still alive after the pruning) decides that the sampled individual has type 0, so that *no* potential ancestor must be considered any further and the process may be killed. This motivates the description of the following structure.

The *killed ASG in the smss-limit* starts with one line emerging from each of the n individuals in the sample. Every line branches at rate s (due to a selective arrow from outside the set of potential ancestors). Every line is pruned at rate $u\nu_1$ (due to a deleterious mutation). At rate $u\nu_0$ per line, the process is killed (due to a beneficial mutation), that is, it is sent to the cemetery state Δ. All the events occur independently on every line. Figure 3.2 depicts some realisations of the killed ASG. There, we adopt the convention that the incoming line is always placed immediately beneath the continuing line. The description gives also rise to a directed graph with labelled vertices similar to the one of Definition 2.12 (although the construction is now not based on the Moran model; see Section 4.2 for details). But as in the finite setting, for our purposes it suffices to count the number of lines in the killed ASG. We define the corresponding process below.

Remark 3.3. Note that, in contrast to the killed ASG for finite population from Section 2.2, collisions and coalescences are absent due to the scaling as described in Section 3.2. $\diamondsuit$

Definition 3.1 (Line-counting process of killed ASG)**.** The *line-counting process of the killed ASG* is the continuous-time Markov chain $R = (R_r)_{r\geq 0}$ with values in $\mathbb{N}_0^{\Delta} := \mathbb{N}_0 \cup \{\Delta\}$ and transition rates

$$q_R(n, n+1) = ns, \qquad q_R(n, n-1) = nu\nu_1, \qquad q_R(n, \Delta) = nu\nu_0 \tag{3.5}$$

for $n \in \mathbb{N}_0$.

The states 0 and Δ are absorbing; all other states are transient. The state 0 is reached if all lines are pruned due to deleterious mutations. The state Δ is reached upon the first beneficial mutation. Absorption in 0 (in Δ) implies that (not) all individuals in the sample are of type 1. If $\nu_0 = 0$, then R is a linear birth-death-process with birth rate s and death rate u. In particular, the process may then also grow to ∞ with positive probability if $u < s$.

We now establish a connection between the solution $y(\cdot\,; y_0)$ of the (deterministic) mutation-selection equation and the (stochastic) line-counting process $(R_r)_{r\geq 0}$, in terms of a duality relation, which formalises the ideas described above. Let $H : [0,1] \times \mathbb{N}_0^\Delta \to \mathbb{R}$ be defined as

$$H(y_0, n) = y_0^n, \quad \text{for } y_0 \in [0,1],\ n \in \mathbb{N}_0^\Delta, \tag{3.6}$$

where $y_0^\Delta := 0$ for all $y_0 \in [0,1]$. The function H returns the probability for n individuals to be of type 1 if each of them is independently of type 1 with probability y_0. Setting $y_0^\Delta = 0$ means that it is impossible to sample an unfit individual that has a beneficial mutation in its direct ancestry. The function H will serve as our duality function.

Theorem 3.2 (Moment duality). *The line-counting process $(R_r)_{r\geq 0}$ of the killed ASG and the solution $y(\cdot\,; y_0)$ of the deterministic mutation-selection equation* (3.1) *satisfy the duality relation*

$$y(t; y_0)^n = \mathbb{E}\Big[y_0^{R_t} \mid R_0 = n\Big] \quad \textit{for all } n \in \mathbb{N}_0^\Delta,\ y_0 \in [0,1],\ \textit{and } t \geq 0. \tag{3.7}$$

Remark 3.4. In the weak mutation–weak selection limit, the above duality appears in [SU86, Lem. 2.1] and [AS05, Thm. 1] if $\nu_0 = 0$. If $\nu_0 > 0$, [SU86, Lem. 2.1] establish a similar duality but with a Feynman-Kac-correction. To our knowledge, the moment duality for $\nu_0 > 0$ without a Feynman-Kac correction has so far not appeared in the literature. ♢

Proof of Theorem 3.2. We can consider $(y(t; y_0))_{t\geq 0}$ as a (deterministic) Markov process on $[0,1]$ with infinitesimal generator

$$A_y f(y) = A_y^s f(y) + A_y^u f(y)$$

for $f \in \mathcal{C}^1([0,1], \mathbb{R})$, where

$$A_y^s f(y) := -sy(1-y)\frac{\partial f}{\partial y} \quad \text{and} \quad A_y^u f(y) := [-u\nu_0 y + u\nu_1(1-y)]\frac{\partial f}{\partial y} \tag{3.8}$$

correspond to selection and mutation, respectively. Since the right-hand side of (3.1) as a function of $y(t)$ is continuously differentiable, it follows from a classic result of ODE theory [KP04, Thm. 8.43] that $y(t; \cdot) \in \mathcal{C}^1([0,1], \mathbb{R})$. In particular, $H(\cdot, n) \in \mathcal{C}^1([0,1])$ and $H(y(t; \cdot), n) \in \mathcal{C}^1([0,1])$. On the other hand, the infinitesimal generator of the line-counting process of the killed ASG reads

$$A_R \tilde{f}(n) = A_R^s \tilde{f}(n) + A_R^u \tilde{f}(n) \tag{3.9}$$

for $\tilde{f} : \mathbb{N}_0^\Delta \to \mathbb{R}$ bounded, where

$$\begin{aligned} A_R^s \tilde{f}(n) &:= ns[\tilde{f}(n+1) - \tilde{f}(n)], \\ A_R^u \tilde{f}(n) &:= nu\nu_1[\tilde{f}(n-1) - \tilde{f}(n)] + nu\nu_0[\tilde{f}(\Delta) - \tilde{f}(n)] \end{aligned}$$

again correspond to selection and mutation, respectively. Let $(P_t^R)_{t\geq 0}$ be the transition semigroup corresponding to R. Since $H(y,\cdot)$ and $P_t^R H(y,\cdot)$ are bounded, they are in the domain of A_R. We now show that

$$A_y H(\cdot,n)(y) = A_R H(y,\cdot)(n) \qquad \text{for } y \in [0,1] \text{ and } n \in \mathbb{N}_0^\Delta$$

and then apply [JK14, Prop. 1.2] to prove the duality. This matching of the generators is a straightforward calculation and can be done individually for the selection and mutation parts; for example,

$$A_y^s H(\cdot,n)(y) = -ns[y^n - y^{n+1}] = A_R^s H(y,\cdot)(n).$$

Similarly, $A_y^u H(\cdot,n)(y) = A_R^u H(y,\cdot)(n)$. □

Theorem 3.2 is the smss-limit equivalent to Theorem 2.14. It provides a stochastic representation of the solution of the deterministic mutation-selection equation. It tells us that the killed ASG is indeed the right process to determine the current type distribution. To see this, set $n=1$ and note that the right-hand side of (3.7) indeed equals the probability that a single individual at time t is of type 1: This is the case if either all lines have been pruned before time t; or if all lines still alive at time t are assigned type 1 when sampling from the initial distribution with weights $(1-y_0, y_0)$, see Figure 3.2.

Remark 3.5. Theorem 3.2 amounts to a weak duality between the forward and the backward process. We expect that this also holds strongly pathwise (see [JK14, Sect. 4] for the corresponding notions). A strong pathwise duality would require an embedding of both processes into the same probability space for all possible initial states. ♢

We are particularly interested in the equilibrium $\bar{y}$. Define $\mathcal{F}_r^R := \sigma(R_t : t \leq r)$, i.e. the natural filtration of $(R_r)_{r\geq 0}$. Let us note in passing.

Corollary 3.3.

$$\left(\bar{y}^{R_r}\right)_{r\geq 0} \textit{ is a } ((\mathcal{F}_r^R)_{r\geq 0})\textit{-martingale.}$$

Proof. Fix $r \geq 0$. Clearly, $\bar{y}^{R_r}$ is adapted to the natural filtration and integrable. Setting $y_0 = \bar{y}$ in (3.7) yields $\mathbb{E}[\bar{y}^{R_r} \mid R_0 = n] = \bar{y}^n$ for all $n \in \mathbb{N}_0^\Delta$. Hence, for $t \in [0,r]$, by the Markov property,

$$\mathbb{E}_n[\bar{y}^{R_r} \mid \mathcal{F}_t^R] = \mathbb{E}_{R_t}[\bar{y}^{R_{r-t}}] = \bar{y}^{R_t}.$$

The result follows. □

We now proceed to recover $\bar{y}$ via the probabilistic backward picture. To this end, we take the limit $t \to \infty$ in (3.7). This leads us to consider the long-term behaviour of R, which is stated in the following Lemma.

Lemma 3.4 (Long-term behaviour R). *(i) If $\nu_0 = 1$, R absorbs in Δ with probability 1.*

(ii) If $\nu_0 \in (0,1)$, R absorbs in $\{0,\Delta\}$ with probability 1.

(iii) If $\nu_0 = 0$ and $u < s$, R absorbs in 0 with probability < 1 and, conditional on non-absorption of R in 0, $R_r \to \infty$ with probability 1.

(iv) If $\nu_0 = 0$ and $u \geq s$, R absorbs in 0 with probability 1.

Proof. If $\nu_0 > 0$ and R is not absorbed in $\{0, \Delta\}$, there is always at least one line in the killed ASG. The time to the first beneficial mutation on any given line is exponentially distributed with parameter $u\nu_0$ and therefore finite almost surely. Hence, for any $n \in \mathbb{N}$,

$$\mathbb{P}(R_r \notin \{0, \Delta\} \mid R_0 = n) \leq \mathbb{P}(R_r \neq \Delta \mid R_r \neq 0, R_0 = n) \overset{r \to \infty}{\longrightarrow} 0.$$

This proves (ii). If $\nu_0 = 1$, we have $\nu_1 = 0$ and hence $\mathbb{P}(R_r = 0 \mid R_0 = n) = 0$ for all $n \in \mathbb{N}$ and $r \geq 0$. The same argument used for (ii) then leads to (i). For (iii), assume $\nu_0 = 0$. Then, R is a linear birth-death-process. Since $\sum_{n \geq 2} s(u/s)^n < \infty$, it follows by [KM57, Thm. 10] that the probability of absorption in 0 is < 1. It is a classic result of branching processes, that R either absorbs or $R_r \to \infty$ almost surely, see [AN72, Ch. III.4]. In particular, conditional on non-absorption, R is transient. The other cases follow by the classic absorption criterion [KM57, Sect. 5]. □

If we take the limit $t \to \infty$ in (3.7) for $n = 1$ and $y_0 \in [0, 1)$, we obtain a representation of the equilibrium frequency $\bar{y}$ in terms of the absorption probability of R in 0. Hence, we obtain the following corollary to the duality result in Theorem 3.2.

Corollary 3.5 (Equilibrium as absorption probability)**.**

$$\bar{y} = P(\lim_{r \to \infty} R_r = 0 \mid R_0 = 1). \tag{3.10}$$

Therefore, we can now recover (3.4) using only properties of R. To calculate the absorption probabilities, let $w_n := P(\lim_{r \to \infty} R_r = 0 \mid R_0 = n)$. A first-step decomposition yields

$$w_n = \frac{s}{u+s} w_{n+1} + \frac{u\nu_1}{u+s} w_{n-1}, \qquad n \geq 1, \tag{3.11}$$

together with $w_0 = 1$ and $w_\Delta = 0$. (3.11) is a homogeneous linear recurrence relation with constant coefficients. It can be solved by classic methods [KP91, Ch. 3.2 and 3.3]. We now take a shortcut to determine w_n and exploit that, due to the independence of the n lines, one has $w_n = w_1^n$, so that it suffices to show the following.

Proposition 3.6.

$$w_1 = \begin{cases} \frac{1}{2}\left(1 + \frac{u}{s} - \sqrt{\left(1 - \frac{u}{s}\right)^2 + 4\nu_0 \frac{u}{s}}\right), & s > 0, \\ \nu_1, & s = 0. \end{cases} \tag{3.12}$$

Remark 3.6. Note that, for $\nu_0 = 0$, (3.12) reduces to

$$w_1 = \begin{cases} \min\{\frac{u}{s}, 1\}, & \text{if } s > 0, \\ 1, & \text{if } s = 0. \end{cases}$$

♢

Proof of Proposition 3.6. The solution to (3.11) is uniquely determined by w_1. Using the product form of w_2, (3.11) evaluated for $n = 1$ leads to

$$sw_1^2 - (u+s)w_1 + u\nu_1 = 0, \tag{3.13}$$

i. e. w_1 satisfies Eq. (3.3). In particular, for $s = 0$, one has $w_1 = \nu_1$. For $s > 0$, we get $w_1 \in \{\bar{y}, y^\star\}$. In addition, if $\nu_0 > 0$ or $u > s$, we already know from Section 3.1 that $y^\star > 1$. Since w_1 is a probability, we therefore have $w_1 = \bar{y}$. If $\nu_0 = 0$ and $u < s$, then $y^\star = 1$ and $\bar{y} < 1$. But Lemma 3.4 implies $w_1 < 1$, so $w_1 = \bar{y}$. Finally, if $\nu_0 = 0$ and $s = u$, we have $w_1 = \bar{y} = y^\star$. □

Since $w_1 = \bar{y}$, Proposition 3.6 is in accordance with Corollary 3.5. We have thus found the desired genealogical interpretation of the solution of the deterministic mutation-selection equation (3.1) and, in particular, of its stable equilibrium $\bar{y}$. Let us explicitly describe what happens in the special case $\nu_0 = 0$, which brings about the bifurcation that corresponds to the error threshold. In this case, Δ cannot be accessed, R is a birth-death process with birth rate s and death rate u, and $w_1 = \bar{y}$ corresponds to its extinction probability. Namely, for $u \geq s$, the process dies out almost surely, whereas for $u < s$, it survives with positive probability $1 - u/s$ and then grows to infinite size almost surely. This is a classical result from the theory of branching processes [AN72, Ch. III.4]: Indeed, for $\nu_0 = 0$, (3.13) is the fixed point equation $w_1 = \varphi(w_1)$ for the generating function φ of the offspring distribution of a binary Galton-Watson process with probability $u/(u+s)$ for no offspring and $s/(u+s)$ for two offspring individuals. This connection sheds new light on the bifurcation observed in Section 3.1 and Fig. 3.1. Namely, let us consider the killed ASG starting from a single individual sampled from the equilibrium population (at some late time t, say), so $R_0 = 1$. If R converges to ∞, then $\lim_{r\to\infty} y_0^{R_r} = 0$ for all $y_0 \in [0,1)$, so the sampled individual is of type 0; whereas $y_0^{R_r} \equiv 1$ for $y_0 = 1$ and $r \geq 0$, which results in an individual of type 1. On the other hand, conditional on eventual absorption of R in 0, $\lim_{r\to\infty} y_0^{R_r} = 1$ for all $y_0 \in [0,1]$, which renders type 1 for the sampled individual.

3.4 pLD-ASG in the smss-limit

The genealogical perspective developed in the previous sections for the mutation-selection equation allows the notion of ancestry also in this setting. Let us now turn to the type of the *ancestor* of a single individual from the (equilibrium) population in the smss-limit. We have seen in Sections 2.4 for the Moran model that it is a more involved problem to determine the ancestral type distribution than to identify the (stationary) type distribution of the forward process, because we now must identify the parental branch (incoming or continuing, depending on the type) at every branching event in the genealogical structure, which requires nested case distinctions. Furthermore, some ancestral lines must be traced back beyond the first mutation. Nevertheless, mutations may still rule out certain potential ancestors. We will work along the lines of the construction of the pLD-ASG for finite populations in Section 2.4. We base our approach on Cordero's extension [Cor17a, Sect. 5] of the line-counting process of the pLD-ASG of Lenz et al. [Len+15] to the framework of the smss-limit. In this section, we recall the idea behind the underlying process and derive some of its properties.

Before we set out, let us specify what we mean by the ancestral type distribution in this setting. After the assignment of types to the leaves of an ASG, we can propagate types and ancestors also in the smss-limit as described in Definition 2.11.

Definition 3.7 (Ancestral type distribution (smss-limit))**.** The ancestral type distribution in the smss-limit, conditional on an initial type distribution $(1 - y_0, y_0)$, is defined as

$$g_r(y_0) := \mathbb{P}(J_r = 1 \mid y_0),$$

where J_r is the type of the ancestor of the root of the ASG in the smss-limit in $[0, r]$ started with a single lineage if we assign to each leaf of the ASG independently type 0 (resp. type 1) with probability $1 - y_0$ (resp. y_0) and propagate types and ancestral sites according to Definition 2.11.

The relevant potential ancestors in the smss-limit are the lines of the ASG that have a non-zero probability to be the ancestor of the single individual at the root of the ASG, in line with Definition 2.20. The pLD-ASG keeps track of these lines. We recall the construction.

The pLD-ASG starts from a single individual. The lines of the graph correspond to the potential ancestors (and hence to a subset of the ASG in the smss-limit) and are assigned consecutive levels, starting at level 1 (see Fig. 3.3). If a line is hit by a selective arrow, its level is increased by one and at the same time all lines above it are shifted up one level; thereby making space for the incoming line, which then occupies the former level of the line it hit. If the first event on a line that does not occupy the top level is a mutation to type 1, we can conclude that it will not be ancestral, since it will, at a later time, play the role of an unsuccessful incoming line, for its type is 1 due to the mutation. Hence we can cut away this line. The line occupying the top level is exempt from the pruning since, regardless of its type, this line will be ancestral if all lines below it are non-ancestral. If a line that is not the top line has a mutation to type 0, we can cut away all lines above it, because this line will, at some stage, be an incoming line and will, due to the mutation, succeed against lines above it. If the top line is hit by a mutation to type 0, this does not have an effect. This motivates the following description of the process.

The *pruned lookdown ASG in the smss-limit* starts at time $r = 0$ and proceeds in direction of increasing r. At every time r, the graph consists of a finite number L_r of lines. The lines are numbered by the integers $1, \dots, L_r$, to which we refer as *levels*. The process then evolves via the following transitions.

(1) Every line $i \leqslant L_r$ branches at rate s and a new line, namely the incoming branch, is inserted at level i and all lines at levels $k \geqslant i$ are pushed one level upward to $k + 1$; in particular, the continuing branch is shifted from level i to $i + 1$. L_r increases to $L_r + 1$.

(2) Every line $i \leqslant L_r$ experiences deleterious mutations at rate $u\nu_1$. If $i = L_r$, nothing happens. If $i < L_r$, the line at level i is pruned, and the lines above it slide down to 'fill the gap', rendering the transition from L_r to $L_r - 1$.

(3) Every line $i \leqslant L_r$ experiences beneficial mutations at rate $u\nu_0$. All the lines at levels $> i$ are pruned, resulting in a transition from L_r to i. Thus, no pruning happens if a beneficial mutation occurs on level L_r.

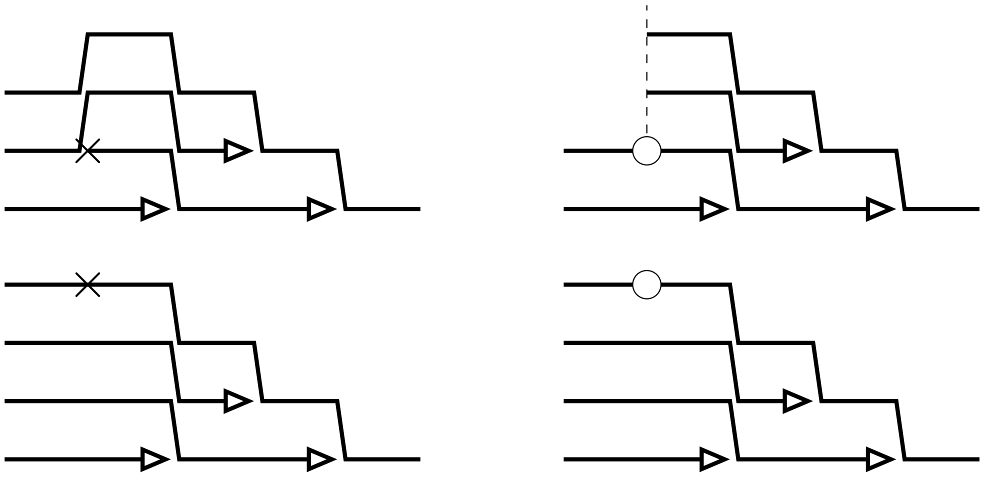


Figure 3.3. The pLD-ASG: Pruning due to a deleterious mutation on a line that is not at the top (top left); pruning of all lines above a beneficial mutation (top right); a deleterious and a beneficial mutation on the top line, which do not affect the number of potential ancestors (bottom).

All the events occur independently on every line. As in Chapter 2, for our purposes it will suffice to only keep track of the number of lines in the pLD-ASG.

Definition 3.8 (Line-counting process of pLD-ASG)**.** The *line-counting process of the pLD-ASG* $L = (L_r)_{r\geq 0}$ is a continuous-time Markov chain on $\mathbb{N}$ with transition rates

$$q_L(n, n+1) = ns, \quad q_L(n, n-1) = (n-1)u\nu_1 + \mathbb{1}_{\{n>1\}}u\nu_0, \quad q_L(n, n-\ell) = u\nu_0, \tag{3.14}$$

where $2 \leq \ell < n$.

Remark 3.7. Note that, in contrast to the pLD-ASG for finite population from Section 2.4, collisions, and coalescences are absent due to the scaling as described in Section 3.2. ◇

Remark 3.8. For later use, we do not insist on starting from a single individual; but one should keep in mind that if we start the process with $n > 1$ lines, then it does not correctly describe the relevant ancestry of n individuals. For example, assume that the first event is a beneficial mutation on line 1. This induces pruning of all other lines, which is incompatible with the relevant ancestry of n individuals. ◇

Remark 3.9. The pLD-ASG and its line-counting process can be constructed via three families of independent homogeneous Poisson processes with rates s, $u\nu_1$, and $u\nu_0$ on each level, indicating a selection, deleterious mutation, and beneficial mutation event, respectively. See, e.g., [BLW16, Sect. 3] for such a construction in the diffusion limit of the Moran model. ◇

Remark 3.10. Cordero [Cor17a, Prop 5.3] proves that the line-counting process of the pLD-ASG for finite populations of Definition 2.21 converges in distribution to L as $N \to \infty$. He works in a framework where L is positive recurrent. In fact, his proof does not rely on the positive recurrence of L and so the result also holds in the case in which L is not positive recurrent. ◇

For any given $r > 0$, a hierarchy is, by construction, imposed on the lines of the graph, such that if at least one line is assigned a 0, the lowest line occupied by a type-0 individual is the true ancestral line. In particular, the ancestor at time r is then of type 0. If all lines are occupied by individuals of type 1, the top line is the true ancestral line and the ancestor at time r is of type 1. (This hierarchy in the pLD-ASG allows us to determine the effect of mutations.) In the finite Moran model and in the diffusion limit, the line that is ancestral if all potential ancestors are of type 1 is called immune (the name originates from the immunity to pruning by deleterious mutations); in the smss-limit, the immune line is always the top line. This rationale gives rise to the following theorem that is a direct consequence of Cordero [Cor17a, Prop. 5.5, Cor. 5.6], which we state here without proof.

Theorem 3.9 (Representation ancestral type distribution). *Let* $y_0 \in [0,1]$ *and* $r \geq 0$. *We have,*

$$g_r(y_0) = \mathbb{E}_1\Big[y_0^{L_r}\Big].$$

The above results leads to a representation of the ancestor's type at any time r; but explicit results require the limit $r \to \infty$. We therefore now consider the long-term behaviour of L_r. Recall that we assume $u > 0$ throughout.

Proposition 3.10 (Long-term behaviour L). *(i) If* $s = 0$, L *absorbs in* 1 *almost surely.*

(ii) If $u < s$ *and* $\nu_0 = 0$, L *is transient, so* $L_r \to \infty$ *almost surely as* $r \to \infty$.

(iii) If $u = s$ *and* $\nu_0 = 0$, L *is null recurrent.*

(iv) If $u > s$ *or* $\nu_0 > 0$, L *is positive recurrent and the stationary distribution is geometric with parameter* $1 - p$, *where*

$$p = \begin{cases} \frac{s}{u\nu_1}\bar{y}, & \text{if } \nu_1 > 0, \\ \frac{s}{u+s}, & \text{if } \nu_1 = 0. \end{cases}$$

Remark 3.11. The parameter of the geometric distribution $p = p(u,s,\nu_1)$ is a function of u, s, and ν_1. Explicitly, it is given by

$$p(u,s,\nu_1) = \begin{cases} \frac{1}{2}\Big(\frac{u+s}{u\nu_1} - \sqrt{\big(\frac{u+s}{u\nu_1}\big)^2 - 4\frac{s}{u\nu_1}}\ \Big), & \text{if } \nu_1 > 0, \\ \frac{s}{u+s}, & \text{if } \nu_1 = 0. \end{cases} \tag{3.15}$$

It is continuous in ν_1, i. e. $\lim_{\nu_1 \to 0} p(u,s,\nu_1) = p(u,s,0)$. ♢

Remark 3.12. A proof of case (iv), for $\nu_1 > 0$, is given in [Cor17a, Lem. 5.3]. ♢

Proof of Proposition 3.10. Case (i) is trivial. Cases (ii) and (iii) are straightforward applications of [KM57, Thms. 1 and 2]. For case (iv), note that L is stochastically dominated by a Yule process with branching rate s. This Yule process is non-explosive. One easily checks that the claimed geometric distribution is invariant. Every process which is non-explosive and has an invariant distribution is positive recurrent, see [Nor98, Thm. 3.5.3]. The uniqueness of the stationary distribution follows from [Nor98, Thm. 3.5.2]. □

In what follows, the long-term behaviour of the tail probabilities of L_r is crucial. Let

$$a_n := \lim_{r\to\infty} \mathbb{P}_1(L_r > n)$$

if this limit exists (the subscript of $\mathbb{P}$ denotes the initial value). We first focus on the positive recurrent case where we know the limit exists.

Proposition 3.11 (Fearnhead's recursion). *If L is positive recurrent, the coefficients $(a_n)_{n\geq 0}$ satisfy*

$$a_n = \frac{s}{u+s} a_{n-1} + \frac{u\nu_1}{u+s} a_{n+1}, \qquad n \in \mathbb{N}, \tag{3.16}$$

with boundary condition $a_0 = 1$ and $\lim_{n\to\infty} a_n = 0$. Moreover, the solution to (3.16) with the boundary condition is unique.

Remark 3.13. Recursion (3.16) is the smss-limit analogue of Fearnhead's recursion for finite populations of Proposition 2.23 and for the diffusion limit [Fea02, Thm. 3]. ♢

Remark 3.14. If we interchange the roles of s and $u\nu_1$ in (3.16) and replace the boundary condition $\lim_{n\to\infty} a_n = 0$ by $a_\Delta = 0$, we obtain the recursion for w_n in (3.11) (note that $u = u\nu_0 + u\nu_1$ so that $u + s$ is invariant under the interchange of s and $u\nu_1$). ♢

Remark 3.15. To prove Fearnhead's recursion, we pursue a strategy that differs from the one in Proposition 2.23. Later, we also recover Proposition 3.11 via the strategy from the previous chapter, see part below Corollary 3.14. ♢

Proof of Proposition 3.11. Here, we give a direct proof via the graphical construction (see Fig. 3.4). The coefficients, as tail probabilities of a stationary distribution, satisfy the boundary conditions. Fix some $r > 0$ and $n \in \mathbb{N}$. Recall from Remark 3.9, that the pLD-ASG can be constructed via three families of independent Poisson processes on each level corresponding to selection, deleterious mutation, and beneficial mutation, respectively. We now look at the last events before r in backward time; which correspond to the first events after r in forward time. Let $T_s(r), T_{\nu_0}(r)$, and $T_{\nu_1}(r)$ be the (backward) times of the last selective, beneficial, and deleterious mutation event, respectively, that have occurred before time r on the first n levels. Set $T(r) := \max\{T_s(r), T_{\nu_0}(r), T_{\nu_1}(r)\}$. On $\{L_r > n\}$, we have that $T(r)$ is positive. Furthermore,

$$\begin{aligned}\mathbb{P}(L_r > n) = \mathbb{P}(L_r > n, T(r) = T_s(r)) + \mathbb{P}(L_r > n, T(r) = T_{\nu_1}(r))\\ + \mathbb{P}(L_r > n, T(r) = T_{\nu_0}(r)).\end{aligned}$$

Let $L_{T(r)-} := \lim_{\tilde{r}\nearrow T(r)} L_{\tilde{r}}$ be the state 'just before' the jump. Reading each transition in Fig. 3.4 from left to right, one concludes the following. If $T(r) = T_s(r)$, then $L_r > n$ if and only if $L_{T(r)-} > n - 1$. If $T(r) = T_{\nu_1}(r)$, then $L_r > n$ if and only if $L_{T(r)-} > n + 1$. The case $T(r) = T_{\nu_0}(r)$ contradicts $L_r > n$, so $\mathbb{P}(L_r > n,\ T(r) = T_{\nu_0}(r)) = 0$. On $\{L_r > n\}$, none of the first n lines is the immune line and therefore the probability that the last event is a selection event or a pruning due to a deleterious mutation is $s/(u+s)$ and $u\nu_1/(u+s)$,

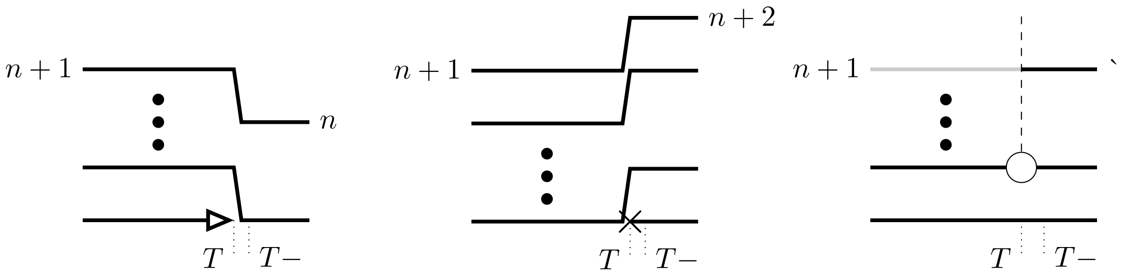


Figure 3.4. The first event on the first n (out of at least $n+1$) lines may be a branching (left) or a pruning due to a deleterious mutation (center); it cannot be a pruning due to a beneficial mutation (right).

respectively. Hence,

$$\begin{aligned}\mathbb{P}(L_r > n) = & \frac{s}{u+s}\mathbb{P}\big(L_{T(r)-} > n-1 \mid T(r) = T_s(r)\big) \\ & + \frac{u\nu_1}{u+s}\mathbb{P}\big(L_{T(r)-} > n+1 \mid T(r) = T_{\nu_1}(r)\big).\end{aligned}$$

But $L_{T(r)-}$ is independent of what happens at time $T(r)$, since this is in the future (in r-time). Taking $r \to \infty$ on both sides proves the assertion. (3.16) is a linear difference equation with constant coefficients and can be solved by classic methods [KP91, Ch. 3.3]. If $\nu_1 = 0$, then $a_n = (s/(u+s))^n$. If $\nu_1 > 0$, then the characteristic root(s) of the associated characteristic polynomial are

$$\lambda_{1/2} = \frac{u+s \mp \sqrt{(u+s)^2 - 4su\nu_1}}{2u\nu_1}.$$

Note that

$$(u+s)^2 - 4su\nu_1 \geq (u\nu_1 - s)^2 + 2su\nu_0 \geq 0,$$

so that the discriminant is positive and all solution are of the form $c_1\lambda_1^n + c_2\lambda_2^n$ [KP91, Thm. 3.6]. If $\nu_1 = 1$ and $s < u$, then $\lambda_1 = s/u$ and $\lambda_2 = 1$. The boundary conditions imply $c_1 = 1$ and $c_2 = 0$. If $\nu_1 \in (0,1)$, a straightforward calculation leads to $\lambda_1 \in [0,1]$ and $\lambda_2 > 1$. Again, by the boundary condition $c_1 = 1$ and $c_2 = 0$. In particular, the solution is unique. □

Remark 3.16. It is an essential ingredient of the present proof of Proposition 3.11 that the immune line is the line at the top. To use this strategy to prove the finite population analogue of Fearnhead's recursion (Proposition 2.23) is not straightforward, because the immune line line can then be on any occupied level. ♢

If L is positive recurrent, we denote by L_∞ a random variable on $\mathbb{N}$ distributed according to the stationary distribution of the line-counting process. Directly solving the recurrence relation (3.16) leads to the geometric distribution of L_∞. We now take another route to recover the distribution via a more probabilistic approach. We derive the memoryless property of L_∞ and conclude that the distribution is geometric, since this is the only discrete distribution without memory.

Proposition 3.12 (Lack of memory property). *If L is positive recurrent, then for all $k \in \mathbb{N}_0$,*

$$\mathbb{P}(L_\infty > n+k \mid L_\infty > n) = \mathbb{P}(L_\infty > k). \tag{3.17}$$

In particular,

$$L_\infty \sim \mathrm{Geom}(1-p), \quad \textit{with } p = \begin{cases} \frac{s}{u\nu_1}\bar{y}, & \textit{if } \nu_1 > 0, \\ \frac{s}{u+s}, & \textit{if } \nu_1 = 0. \end{cases} \tag{3.18}$$

Proof. Denote $b_k^{(n)} := \mathbb{P}(L_\infty > n+k \mid L_\infty > n)$. Clearly, $b_0^{(n)} = 1$ and $\lim_{k\to\infty} b_k^{(n)} = 0$ for all $n \in \mathbb{N}$. By Proposition 3.11,

$$b_k^{(n)} = \frac{a_{n+k}}{a_n} = \frac{1}{a_n}\Big(\frac{s}{u+s}a_{n+k-1} + \frac{u\nu_1}{u+s}a_{n+k+1}\Big) = \frac{s}{u+s}b_{k-1}^{(n)} + \frac{u\nu_1}{u+s}b_{k+1}^{(n)}.$$

Since the solution of (3.16) is unique with the boundary conditions, it follows that $b_k^{(n)} = a_k$ for all $k \in \mathbb{N}_0$. As a consequence, $\mathbb{P}(L_\infty > n) = a_1^n$. Now that we know (again) that L_∞ has indeed a geometric distribution, it remains to determine the parameter. By Proposition 3.11,

$$a_1 = \frac{s}{u+s} + \frac{u\nu_1}{u+s}a_1^2, \tag{3.19}$$

of which the solution is given by $a_1 = s\bar{y}/u\nu_1$ if $\nu_1 > 0$ and $a_1 = s/(u+s)$ if $\nu_1 = 0$. □

As in the finite population case, recursion (3.16) looks like a first-step decomposition for the absorption probabilities of some other process. And indeed, we establish a similar connection as in Proposition 2.25 in the smss-limit. Let $(D_t)_{t\geq 0}$ be the process on $\mathbb{N}^\Delta := \mathbb{N} \cup \{\Delta\}$ with transition rates given by

$$q_D(d, d-1) = (d-1)s, \qquad q_D(d, d+1) = (d-1)u\nu_1, \qquad q_D(d, \Delta) = (d-1)u\nu_0 \tag{3.20}$$

for $d \in \mathbb{N}$. We adopt the convention that $n < \Delta$ for all $n \in \mathbb{N}$.

Remark 3.17. The process D exhibits an interesting connection to the line-counting process of the killed ASG. Let D and L be as previously defined with given rates $u\nu_0$, $u\nu_1$, and s. Furthermore, let $\breve{D}$ and $\breve{L}$ be the same processes, but with rates s and $u\nu_1$ interchanged. Write $\breve{R}$ for the line-counting process of a killed ASG with beneficial and deleterious mutation rate $u\nu_0$ and s, respectively, and selection rate $u\nu_1$ (so $u\nu_1$ and s are interchanged). Note that the rate at which mutation events occur is then $\breve{u} = u\nu_0 + s$; similarly, given that a mutation occurs, the probabilities for beneficial and deleterious mutations are $\breve{\nu}_0 = u\nu_0/(u\nu_0 + s)$ and $\breve{\nu}_1 = s/(u\nu_0 + s)$, respectively. Comparing (3.5) and (3.20) immediately yields

$$\breve{R} \overset{d}{=} D - 1 \quad \text{if} \quad \breve{R}_0 = D_0 - 1. \tag{3.21}$$

In particular, the long-term behaviour of D follows by means of Lemma 3.4: If $\nu_0 > 0$, D absorbs in $\{1, \Delta\}$ with probability 1. If in addition $s = 0$, D absorbs in Δ with probability 1. If $\nu_0 = 0$ and $u \leq s$, D absorbs in 1 with probability 1. If $\nu_0 = 0$ and $u > s$, D absorbs in 1 with probability < 1 and, conditional on non-absorption of D in 1, $D_t \to \infty$ with probability 1. ♢

Proposition 3.13 (Siegmund duality). *L and D are Siegmund dual, i. e. for $t \geq 0$,*

$$\mathbb{P}(d \leq L_t \mid L_0 = n) = \mathbb{P}(D_t \leq n \mid D_0 = d), \qquad \forall n \in \mathbb{N},\ d \in \mathbb{N}^{\Delta}. \tag{3.22}$$

Remark 3.18. This is the smss-limit analogue to Proposition 2.25. ◊

Proof of Proposition 3.13. We denote the infinitesimal generators of L and D by A_L and A_D, respectively. They have the form

$$\begin{aligned} A_L f(n) = {} & ns[f(n+1) - f(n)] + \Big((n-1)u\nu_1 + \mathbb{1}_{\{n\geq 1\}}u\nu_0\Big)[f(n-1) - f(n)] \\ & + u\nu_0 \sum_{i=1}^{n-2} [f(i) - f(n)] \end{aligned}$$

for $f : \mathbb{N} \to \mathbb{R}$ bounded, and

$$\begin{aligned} A_D \tilde{f}(d) = {} & (d-1)s[\tilde{f}(d-1) - \tilde{f}(d)] + (d-1)u\nu_1[\tilde{f}(d+1) - \tilde{f}(d)] \\ & + (d-1)u\nu_0[\tilde{f}(\Delta) - \tilde{f}(d)] \end{aligned}$$

for $\tilde{f} : \mathbb{N}^{\Delta} \to \mathbb{R}$ bounded. In the case of a Siegmund duality, the duality function is

$$\bar{H}(n,d) = \mathbb{1}_{(d\leq n)}.$$

Denote by $(P_t^L)_{t\geq 0}$ and $(P_t^D)_{t\geq 0}$ the transition semigroup corresponding to L and D, respectively. $\bar{H}(\cdot,d)(n)$ and $P_t^L \bar{H}(\cdot,d)(n)$ are in the domain of A_L and $\bar{H}(n,\cdot)(d)$ and $P_t^D \bar{H}(n,\cdot)(d)$ are in the domain of A_D. We will show that $A_L\bar{H}(\cdot,d)(n) = A_D\bar{H}(n,\cdot)(d)$. The result follows then as an application of [JK14, Prop. 1.2]. Indeed,

$$A_L\bar{H}(\cdot,d)(n) = ns\mathbb{1}_{(d=n+1)} - (n-1)u\nu_1\mathbb{1}_{(d=n)} - u\nu_0 \sum_{i=1}^{n-1} \mathbb{1}_{(d\leq n<d+i)}.$$

Note that

$$\sum_{i=1}^{n-1} \mathbb{1}_{(d\leq n<d+i)} = \begin{cases} d-1, & \text{if } n \geq d > 1, \\ 0, & \text{otherwise.} \end{cases}$$

Thus, we can rewrite $A_L\bar{H}(\cdot,d)(n)$ as

$$(d-1)s\mathbb{1}_{(d-1=n)} - (d-1)u\nu_1\mathbb{1}_{(d=n)} - (d-1)u\nu_0\mathbb{1}_{(d\leq n)},$$

which equals $A_D\bar{H}(n,\cdot)(d)$, as required. □

Remark 3.19. The analogous result in the diffusion limit is proved in [BLW16, Lem. 4.5] via Clifford-Sudbury flights [CS85]. Their proof leads to a pathwise duality. We expect a similar argument to apply also in our setting. ◊

Corollary 3.14 (Tail probababilities as absorption probabilities).

$$\mathbb{P}(\lim_{t\to\infty} D_t = 1 \mid D_0 = n+1) = a_n, \qquad \forall n \in \mathbb{N}. \tag{3.23}$$

Proof. The proof is a direct consequence of Proposition 3.13. □

This result gives an alternative way to recover (3.16) via a first-step decomposition of the absorption probabilities of D. By Remark 3.17, the absorption probability of D in 1, given $D_0 = n+1$, equals the absorption probability of $\breve{R}$ in 0, given $\breve{R}_0 = n$. In particular, $a_n = \breve{\bar{y}}^n$, where

$$\breve{\bar{y}} := \begin{cases} \frac{1}{2}\left(1 + \frac{u\nu_0+s}{u\nu_1} - \sqrt{(1-\frac{u\nu_0+s}{u\nu_1})^2 + 4\frac{\nu_0}{\nu_1}}\right), & \nu_1 > 0, \\ \frac{s}{u+s}, & \nu_1 = 0. \end{cases} \tag{3.24}$$

This is consistent with Remark 3.14: The recursion for the absorption probability of $\breve{R}$ in 0 is obtained by interchanging the roles of $u\nu_1$ and s in Fearnhead's recursion. Hence, $\breve{\bar{y}}$ is as in (3.4), but with $u\nu_1$ and s interchanged; note that this implies replacement of $u = u\nu_0 + u\nu_1$ by $u\nu_0 + s$. On the other hand, as a consequence of Corollary 3.14, $\breve{\bar{y}} = p$ with p from Proposition 3.12. In a similar way, we can derive that

$$\bar{y} = \lim_{r\to\infty} \mathbb{P}_1(\breve{L}_r > 1).$$

We can now deal with the long-term behaviour when L is null recurrent (recall from Proposition 3.10 that this is the case for $\nu_0 = 0$ and $u = s$).

Corollary 3.15 (Large time behaviour for null recurrent L)**.** *If $u = s$ and $\nu_0 = 0$,*

$$\lim_{r\to\infty} \mathbb{P}_1(L_r > n) = 1.$$

Proof. The proof is an immediate consequence of Corollary 3.14 together with Remark 3.17. □

3.5 Ancestral type distribution in the smss-limit

In the Moran model and in the diffusion limit, all individuals at present originate from a single individual in the distant past, see [Cor17a, Sect. 3] and [KN97, Thm. 3.2], respectively. This individual is called the common ancestor, and the distribution of its type is the common ancestor type distribution. In the smss-limit, there are no coalescence and collision events (see Section 3.2, alternatively [Cor17a, Sect. 5][KN97]), so the notion of a common ancestor does not make sense. Instead Cordero [Cor17a, p. 617] introduces the representative ancestral type. This is the type of the ancestor of a generic individual in the population, denoted earlier [GB03] as the ancestral type of a typical individual. The general concept was developed by Jagers [Jag89; Jag92] in the context of branching processes. By a slight abuse of wording, we will call the representative ancestor just ancestor; and therefore the representative ancestral type will just be called the ancestral type. The ASG in the smss-limit describes the ancestry of a typical individual. We denote the ancestral type at backward time r of the individual at the root of the ASG by $J_r \in \{0,1\}$. The probability of an unfit ancestral type at backward time r if the leaves are sampled according to $(1-y_0, y_0)$ is given by $g_r(y_0) = \mathbb{P}(J_r = 1 \mid y_0)$, see Definition 3.7. By Theorem 3.9 (alternatively, see

[Cor17a, Prop. 5.5, Cor. 5.6]), we know

$$g_r(y_0) = \mathbb{E}_1\left[y_0^{L_r}\right] = 1 - (1 - y_0) \sum_{n\geq 0} \mathbb{P}_1(L_r > n) y_0^n. \tag{3.25}$$

This is consistent with the graphical picture: the ancestral type at backward time r is 1 if and only if all L_r lines are of type 1. The corresponding probability is given by $\mathbb{E}_1[y_0^{L_r}]$. Alternatively, we can partition the event of a beneficial ancestor according to the first level occupied by a type-0 individual. Namely,

$$\mathbb{P}_1(L_r > n)(1 - y_0)y_0^n$$

is the probability that at least $n + 1$ lines are present, the $(n + 1)$st line is of type 0, and the first n lines are of type 1. Summing this probability over n gives the probability of an ancestral type 0. The complementary probability leads to the right-hand side of (3.25).

Now let $g_\infty(y_0) := \lim_{r\to\infty} g_r(y_0)$ be the probability of an unfit ancestral type of an individual sampled at a very late time conditional on the initial frequency of unfit individuals being y_0 if the limit exists. If $\sum_{n\geq 0} a_n y_0^n < \infty$, equation (3.25) yields

$$g_\infty(y_0) = 1 - (1 - y_0) \sum_{n\geq 0} a_n y_0^n. \tag{3.26}$$

We can now exploit what we know about the a_n to obtain explicit expressions for g_∞. This is captured in the following theorem.

Theorem 3.16. *For $y_0 \in [0, 1]$, $g_\infty(y_0)$ exists.*

(i) If $s = 0$, $g_\infty(y_0) = y_0$ for all $y_0 \in [0, 1]$.

(ii) If $u \leq s$ and $\nu_0 = 0$, $g_\infty(y_0) = \begin{cases} 0, & \text{if } y_0 \in [0,1), \\ 1, & \text{if } y_0 = 1. \end{cases}$

(iii) If $s > 0$ and either $u > s$ or $\nu_0 > 0$, $g_\infty(y_0) = \frac{1-p}{1-py_0} y_0$.

Proof. In case (i), L absorbs in 1 and hence $a_0 = 1$ and $a_n = 0$ for all $n \geq 1$. In particular, $\sum_{n\geq 0} a_n y_0^n < \infty$ for all $y_0 \in [0, 1]$, so that together with (3.26) the result follows. For case (ii), we first treat the subcase $u < s$ and $\nu_0 = 0$. There, L is transient and hence $L_r \to \infty$ almost surely. Hence, $a_n = 1$ for all $n \geq 0$. For $y_0 \in [0, 1)$, again $\sum_{n\geq 0} a_n y_0^n < \infty$, and the result follows by (3.26). For $y_0 = 1$, we use that L is bounded for all $r > 0$. In particular, $\sum_{n\geq 0} \mathbb{P}(L_r > n \mid L_0 = 1) < \infty$. But then, $g_r(1) = 1$ for $r > 0$. Taking the limit $r \to \infty$ yields the result. The other subcase of (ii) is $u = s$ and $\nu_0 = 0$. There, Corollary 3.15 leads to $a_n \equiv 1$ $(n \geq 0)$. Case (iii) follows by summing the geometric series obtained from (3.26) via Proposition 3.12. □

Theorem 3.16 is consistent with the graphical picture. Case (i) corresponds to the neutral situation, in which each individual has exactly one potential ancestor at all times; the ancestor is then a single draw from the initial distribution. In particular, there is no bias towards one of the types. In case (iii), $L_\infty > 1$ with positive probability, so there is a bias towards the beneficial type. The reason is that a single beneficial potential ancestor suffices for the ancestral type to be of type 0, which manifests itself in the factor $(1-p)/(1-py_0) < 1$ for $y_0 < 1$. In case (ii), depending on whether $u = s$ or $u < s$, L is null recurrent or transient. In both cases, the number of potential ancestors in the limit $r \to \infty$ is infinite and the bias towards type 0 is taken to an extreme: Any positive proportion of beneficial types suffices to ensure that the ancestor has type 0. If there are no beneficial types in the population, the ancestor is of type 1 with probability 1.

Let us now study the dependence on ν_0 of the probability for an unfit ancestral type. To stress the dependence, we write $g_\infty(y_0, \nu_0)$. We recover Corollary 2.27 in the smss-limit. The result was first observed by [Len+15, Sect. 6.1] for the diffusion limit.

Proposition 3.17 (Monotonicity ancestral type distribution)**.** *Let* $y_0 \in (0,1)$*. If* $s > 0$ *and* $\nu_0, \bar{\nu}_0 \in [0,1]$ *with* $\nu_0 < \bar{\nu}_0$*, then*

$$g_\infty(y_0, \nu_0) < g_\infty(y_0, \bar{\nu}_0).$$

Proof. The same technique to prove Proposition 2.27 leads to the result. ☐

Consider now $(1-g_\infty(\bar{y}), g_\infty(\bar{y}))$, namely, the distribution of the ancestral type that lives in the stable equilibrium population $\bar{y}$. We call it the ancestral type distribution at equilibrium and characterise it in what follows. First note that both g_∞ and $\bar{y}$ are functions of ν_0. In order to stress this (double) dependence, we write $g_\infty(\bar{y}(\nu_0), \nu_0)$. We also write $L(\nu_0)$ instead of L. Recall from Section 3.1 that, for $\nu_0 = 0$, $\bar{y}(0) = \min\{u/s, 1\}$ and so, by Theorem 3.16,

$$g_\infty(\bar{y}(0), 0) = \begin{cases} 0, & \text{if } u < s, \\ 1, & \text{if } u \geq s. \end{cases}$$

This is the counterpart to the transcritical bifurcation (or error threshold) of the equilibrium

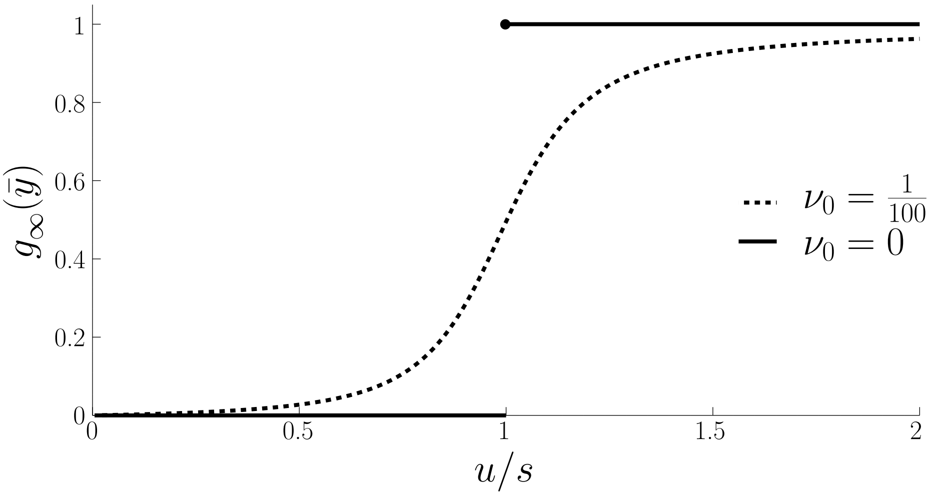


Figure 3.5. The probability of an unfit ancestral type at equilibrium.

frequency of the forward process. The probability of an unfit ancestral type at equilibrium exhibits an even more drastic behaviour: a jump from 0 to 1 if u surpasses the critical value s, see Fig. 3.5. If $\nu_0 \in (0,1)$, $L(\nu_0)$ is positive recurrent and $L_\infty(\nu_0)$ is almost surely finite. Moreover, in this case $\bar{y}(\nu_0) \in (0,1)$ for all $u > 0$. In particular, $g_\infty(\bar{y}(\nu_0), \nu_0) = \mathbb{E}_1[\bar{y}(\nu_0)^{L_\infty(\nu_0)}] \in (0,1)$, and hence

$$g_\infty(\bar{y}(\nu_0), \nu_0) \begin{cases} > g_\infty(\bar{y}(0), 0) = 0, & \text{if } u < s, \\ < g_\infty(\bar{y}(0), 0) = 1, & \text{if } u \geq s, \end{cases}$$

compare Figs. 3.1 and 3.5. It may seem surprising at first sight that, even though switching off beneficial mutations leads to an increase of $\bar{y}$ for all values of u, it decreases the probability of the deleterious type to be ancestral if $u < s$, but increases it for $u \geq s$. The reason for this is that:

- for $u < s$: in contrast to the case $\nu_0 \in (0,1)$, where $L_\infty(\nu_0)$ is finite, $L_\infty(0)$ is infinite, and hence beats $\bar{y}(0)$ regardless of its value.
- for $u \geq s$: in contrast to the case $\nu_0 \in (0,1)$, where $\bar{y}(\nu_0)$ is strictly positive, $\bar{y}(0) = 1$, and therefore, there is no chance to sample an ancestor of type 0, regardless of the value of $L_\infty(\nu_0)$.

In the end, we recover Theorem 2.29 in the smss-limit. Recall that Taylor [Tay07, Eq. (11)] shows that the common ancestor type distribution is a harmonic function for the generator of a jump-diffusion process (see also [Len+15, Sect. 7.1] for a more detailed discussion). The analogue of this jump-diffusion (or alternatively of the absorbing Markov chain in Section 2.4.4 of this thesis) in the smss-framework is the piecewise-deterministic Markov process $\tilde{Y} := (\tilde{Y}_t)_{t\geq 0}$ with generator

$$\begin{aligned} A_{\tilde{Y}} f(y) = {} & [-sy(1-y) - u\nu_0 y + u\nu_1(1-y)] \frac{\partial f}{\partial y} + \frac{u\nu_0 y}{1-y} [f(1) - f(y)] \\ & + \frac{u\nu_1(1-y)}{y} [f(0) - f(y)] \end{aligned} \tag{3.27}$$

for $f \in C_1([0,1], \mathbb{R})$ with $\lim_{y\to 1} A_{\tilde{Y}} f(y) = \lim_{y\to 0} A_{\tilde{Y}} f(y) = 0$. The latter means that $\tilde{Y}$ absorbs in 0 or 1. This process follows the dynamics of the mutation-selection equation up to a random jumping time. At this time the process jumps to one of the boundaries where it is absorbed, see also Fig. 3.6. Existence and uniqueness of a Markov process corresponding to $A_{\tilde{Y}}$ follow by proving that the jump rates, which diverge at the boundary, are in fact bounded along trajectories of the process over any finite time interval. If $\nu_0 \in (0,1)$ and $y_0 \in (0,1)$, then $y(t; y_0)$ never hits the boundary, since $\bar{y} \in (0,1)$. If $\nu_0 = 1$ or $\nu_0 = 0$ and $y_0 \in (0,1)$, then $y(t; y_0)$ hits 0 or 1, respectively. In both cases, the possibly diverging jump term is absent because either $\nu_1 = 0$ or $\nu_0 = 0$, respectively. We now strengthen Taylor's result and show that $\tilde{Y}$ is dual to L.

Theorem 3.18 (Duality). *The processes $\tilde{Y}$ and L are dual with respect to the duality function $H(y,n) = y^n$, that is, for $t \geq 0$,*

$$\mathbb{E}[\tilde{Y}_t^n \mid \tilde{Y}_0 = y_0] = \mathbb{E}[y_0^{L_t} \mid L_0 = n], \qquad \forall y_0 \in [0,1], \ n \in \mathbb{N}. \tag{3.28}$$

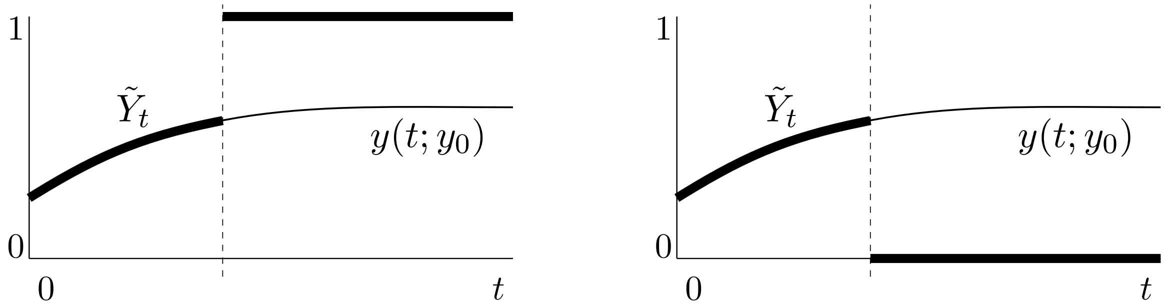


Figure 3.6. $\tilde{Y}$ follows $y(t;y_0)$ until a random time at which it jumps to either 1 (left) or 0 (right).

Remark 3.20. This is the smss-limit analogue to Theorem 2.29. ♢

Proof of Theorem 3.18. The infinitesimal generator of $\tilde{Y}$ is given by $A_{\tilde{Y}}\tilde{f}(y) = A_y^s\tilde{f}(y) + A_{\tilde{Y}}^u\tilde{f}(y)$, with A_y^s of (3.8) and

$$\begin{aligned}A_{\tilde{Y}}^u\tilde{f}(y) := {}& [-yu\nu_0 + u\nu_1(1-y)]\frac{\partial \tilde{f}}{\partial y} + \frac{y}{1-y}u\nu_0\left[\tilde{f}(1) - \tilde{f}(y)\right] \\ &+ \frac{1-y}{y}u\nu_1\left[\tilde{f}(0) - \tilde{f}(y)\right],\end{aligned}$$

for $\tilde{f} \in \mathcal{C}^1([0,1],\mathbb{R})$ with $\lim_{y\to 0} A_{\tilde{Y}}\tilde{f}(y) = \lim_{y\to 1} A_{\tilde{Y}}\tilde{f}(y) = 0$. ($A_{\tilde{Y}}^u$ should not be confused with A_Y^u of (3.8).) A straightforward calculation shows that H is in the domain of $A_{\tilde{Y}}$. The infinitesimal generator of L is given by $A_L f(n) = A_R^s f(n) + A_L^u f(n)$, with A_R^s of (3.9) and

$$A_L^u f(n) = (n-1)u\nu_1[f(n-1) - f(n)] + u\nu_0\sum_{k=1}^{n-1}[f(k) - f(n)],$$

for $f : \mathbb{N} \to \mathbb{R}$ bounded. We use [EK86, Cor. 4.4.13] to prove the result. Let us first verify the generator identity, i.e. $A_{\tilde{Y}}H(\cdot,n)(y_0) = A_L H(y_0,\cdot)(n)$. In the proof of Theorem 3.2, we already showed $A_y^s H(\cdot,n)(y) = A_R^s H(y,\cdot)(n)$. Hence, it suffices to check $A_{\tilde{Y}}^u H(\cdot,n)(y) = A_L^u H(y,\cdot)(n)$, which then implies

$$A_{\tilde{Y}}H(\cdot,n)(y) = A_L H(y,\cdot)(n), \qquad \forall y \in [0,1], n \in \mathbb{N}.$$

Indeed, we obtain

$$\begin{aligned}&A_{\tilde{Y}}^u H(\cdot,n)(y) \\ &= (n-1)u\nu_1\left[y^{n-1} - y^n\right] + u\nu_0\left[y\frac{1-y^n}{1-y} - ny^n\right] \\ &= (n-1)u\nu_1\left[H(y,n-1) - H(y,n)\right] + u\nu_0\sum_{j=1}^{n-1}\left[H(y,j) - H(y,n)\right] \\ &= A_L^u H(y,\cdot)(n).\end{aligned}$$

Next, we verify the integrability conditions. Fix $T > 0$, $y_0 \in [0,1]$, and $L_0 \in \mathbb{N}$. Clearly, $\sup_{r,t\in[0,T]} |H(\tilde{Y}_t, L_r)| \leq 1$. Furthermore,

$$\sup_{t,r\leq T} |A_L H(\tilde{Y}_t, L_r)| \leq 2(u+s) \sup_{r\leq T} L_r.$$

Note that L can be coupled to a pure birth process $\hat{L}$ with transition rate $q_{\hat{L}}(n, n+1) = ns$ such that $\hat{L}_0 = L_0$ and $\sup_{r\leq T} L_r \leq \hat{L}_T$. $\hat{L}_T$ is at any finite time finite almost surely (see, e.g., [Bre68, Ch. 15.7]); hence, so is the right hand side of the above equation. Finally, $A_{\tilde{Y}}$ and A_L are the generators of the progressive Markov processes $\tilde{Y}$ and L, respectively, and H is in the domain of their generators. Hence,

$$H(\tilde{Y}_t, n) - \int_0^t A_{\tilde{Y}} H(\tilde{Y}_s, n) ds \quad \text{and} \quad H(y_0, L_t) - \int_0^t A_L H(y_0, L_s) ds$$

are martingales for all $n \in \mathbb{N}$ and $y_0 \in [0,1]$ [EK86, Prop. 4.1.7]. The result follows. □

We now obtain a characterisation of $g_\infty(y_0)$ that does not depend on L by taking $t \to \infty$ in (3.25) and (3.28). The analogue result in the weak-mutation weak-selection limit is proved in [Tay07, Prop. 2.5].

Corollary 3.19 (Ancestral type distribution as hitting probability). *For $y_0 \in [0,1]$, we have*

$$g_\infty(y_0) = \mathbb{P}(\lim_{t\to\infty} \tilde{Y}_t = 1 \mid \tilde{Y}_0 = y_0).$$

Remark 3.21. The Kolmogorov backward equation for the absorption probability of $\tilde{Y}_t$ in 1 leads to the characterisation of g_∞ as the solution to the boundary value problem

$$A_{\tilde{Y}} g_\infty(y_0) = 0 \quad \text{for } y_0 \in (0,1), \tag{3.29}$$

complemented by $g_\infty(0) = 0$ and $g_\infty(1) = 1$. It is the smss-limit analogue of the boundary value problem in Taylor [Tay07, Eq. (11)] (diffusion limit) and of Corollary 2.31 (finite population). ♢

4

Ancestral lines in the mutation-selection equation with pairwise interaction

Our goal is to extend the results for the haploid mutation-selection equation of Chapter 3 to the case of pairwise interaction, where the reproduction rate of an individual depends on the type of a partner chosen uniformly from the population. Biologically, this is a special case of what is known as frequency-dependent selection, which also occurs in evolutionary game theory (see, e.g., [HS98, Ch. 22]). The resulting equation is also equivalent to the so-called diploid mutation-selection equation, which describes a population of individuals that carry two copies of the genetic information rather than one as in the haploid case.

The diploid mutation-selection equation is characterised by a cubic polynomial as opposed to a quadratic polynomial in the haploid case. This leads to a richer bifurcation structure. In particular, one observes bistability in certain parameter regions; this is absent in the haploid case. While this is well-known in the forward direction of time, the corresponding ancestral processes are largely unexplored territory. The ASG with pairwise interaction, which we described in Section 2.2, is our starting point. Even though the type-frequency process of the entire population evolves deterministically, the ancestry of each individual is still stochastic and the resulting construction is a specific random tree. To make this object tractable, we will *prune* the tree upon mutation, thus reducing it to its informative parts. The hierarchies inherent in the tree will play a crucial role and will be encoded systematically via ternary trees with weighted leaves; this will lead to the *stratified ASG*. The latter will serve as a dual to the forward process and provide a stochastic representation of the solution of the mutation-selection equation with pairwise interaction. It will also be our workhorse to reveal the genealogical structures inherent in the bifurcations of the equilibria of the ODE. Indeed, it will turn out that the random genealogical trees have very different properties in the various parameter regimes. Furthermore, we will establish constructions, again based on the stratified ASG, that allow us to trace back the ancestral lines into the distant past and obtain explicit results about the ancestral population in the biologically relevant case of unidirectional mutation to the deleterious type (i.e. $\nu_0 = 0$).

The chapter is organised as follows. In Section 4.1, we recapitulate the two-type mutation-selection differential equation with pairwise interaction, along with its equilibrium and bifurcation structure. We then connect the Moran model with pairwise interaction to the differential equation via a dynamical law of large numbers. In Section 4.2, applying the law of large numbers to the ASG leads to the ancestral process corresponding to the deterministic equation. The stratified ASG is introduced in Section 4.3, and some of its fundamental properties are proved. The core of the results are found in Section 4.4 and 4.5. Section 4.4 establishes the connection between the stratified ASG and the deterministic model in the form of a duality relation and provides a stochastic representation of the solution of the differential equation. In this way, the bifurcation structure in the case of unidirectional mutation is recovered by genealogical means. Finally, Section 4.5 establishes the constructions to trace back an ancestral line into the distant past; these are the stratified ASG with immune line and the forest of stratified ASGs. We use them to derive the ancestral type distribution under unidirectional mutation to the deleterious type (i.e. $\nu_0 = 0$), both for finite time horizons and at stationarity.

4.1 Mutation-selection equation with pairwise interaction

We consider a *deterministic model* for the evolution of a population subject to mutation, selection, and a special form of frequency-dependent selection that we call pairwise interaction. More precisely, the population is composed of two types, type 0 and type 1. Denote by $y(t; y_0)$ the proportion of type 1 at time t given that the proportion at time 0 was $y_0 \in [0,1]$. Then $y(t; y_0)$ evolves deterministically as the solution of the initial value problem (IVP)

$$\begin{aligned} \frac{dy}{dt}(t) &= -y(t)(1-y(t))\left[s+\gamma(1-y(t))\right] + u\nu_1(1-y(t)) - u\nu_0 y(t) =: F(y(t)), \\ y(0) &= y_0, \quad \text{for } y_0 \in [0,1]. \end{aligned} \tag{4.1}$$

Throughout, we assume $\nu_0, \nu_1 \geq 0$ and $\nu_0 + \nu_1 = 1$. We speak of the ODE as the mutation-selection equation with pairwise interaction. We will see (Proposition 4.6) that the model arises as the large population limit of the Moran model with mutation, selection, and pairwise interaction.

The form of the mutation terms is obvious; the reproduction term describes that type-0 individuals selectively reproduce at effective rate $(1-y(t))[s+\gamma(1-y(t))]$ and thus reduce the relative frequency of type-1 individuals according to their current proportion $y(t)$. Neutral reproduction (which we assumed to occur of rate 1) does not enter the equation since its net contribution is $-(1-y(t))y(t) + y(t)(1-y(t)) = 0$. Note for later use, that this generalises to any neutral reproduction rate $c \geq 0$. In particular, it is no restriction to set $c = 0$. In principle, the solution of (4.1) can be expressed explicitly in terms of the roots of F by standard methods, but we refrain from doing this here.

Remark 4.1. We have introduced the interaction in terms of frequency-dependent selection, but an alternative interpretation corresponds to a diploid population with two allelic types subject to mutation and selection; the ODE is then usually termed the diploid mutation-selection equation. The diploid genotypes are then given by the pairs $(i,j), i,j \in \{0,1\}$, and their reproduction rates w_{ij} are $w_{00} = 1+2s+\gamma$, $w_{01} = w_{10} = 1+s$, and $w_{11} = 1$. This choice of parameters corresponds to the case where allele 0 is (partially) recessive, that is, it needs another 0 to fully play out its selective advantage (see also [BW97]). ♢

Remark 4.2. The general form of the mutation-selection equation goes back to Wright [Wri49] and is intensively discussed by Crow and Kimura [CK56]. They understand it as a simplistic model for the evolution of a population under the forces of mutation, (frequency-dependent) selection, and migration; and note that it is suited for an approximation of

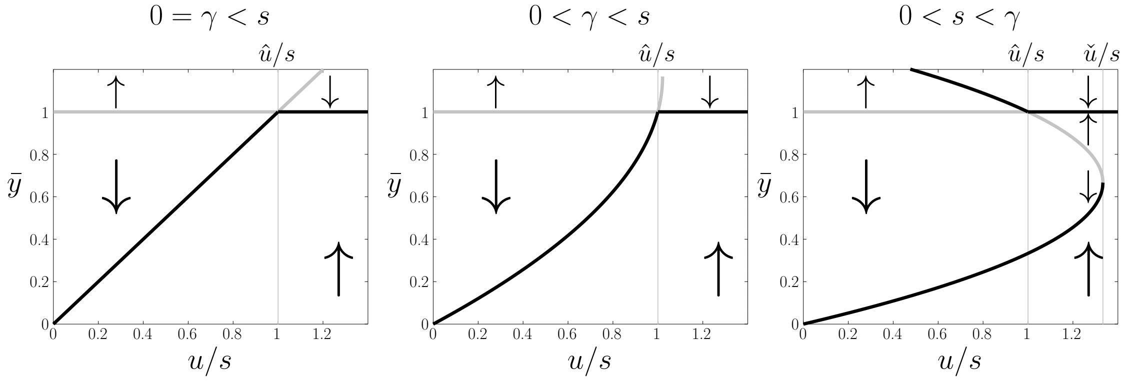


Figure 4.1. The equilibria $\bar{y}$ of (4.1) as a function of u/s for $\nu_0 = 0$ and $s = 1/30$. The left, middle, and right panels correspond to $\gamma = 0$, $\gamma = 1/40$, and $\gamma = 1/10$, respectively. Black lines: stable. Grey lines: unstable.

a large population with constant environmental factors [CK56, Sect. 1]. We consider a special case in which there is no migration, only two allelic types, and our special form of frequency-dependent selection. ♢

Remark 4.3. For a special choice of parameters, the ODE (4.1) corresponds to the mean-field limit (or law of large numbers) of the cooperative branching model on the complete graph as investigated by Mach et al. [MSS18b] (see also [Mac17]). More precisely, our notation translates to their case by interchanging the roles of type 0 and 1 and by setting $\gamma = \alpha$, $u = 1$, $\nu_0 = 0$, $\nu_1 = 1$, and $s = 0$. This leads to the mean-field equation [MSS18b, Eq. (1.36)]. ♢

Equilibria and bifurcation structure

To understand the large time behaviour of the solutions of the mutation-selection equation with pairwise interaction, we analyse the equilibria $\bar{y}$, namely the (real) roots of F in (4.1), along with their stability.

Since $F(0) \geq 0$, $F(1) \leq 0$, and F is continuous (it is a polynomial), the unit interval (which is the biologically relevant domain) is positive invariant under the flow and there exists at least one root of F in $[0,1]$. Let $\hat{y}_\infty$ and $\check{y}_\infty$ be the smallest and largest root of F in $[0,1]$, respectively. The conditions on the boundary, together with the continuity of F, imply that $\hat{y}_\infty$ and $\check{y}_\infty$ are attracting from the left and right, respectively, provided they are in $(0,1)$. We initially concentrate on the case $\nu_0 = 0$, in which expressions for the roots simplify significantly, and extend (4.1) to $\mathbb{R}$, considering *all* equilibria. Later, we specialise to those equilibria that lie in the unit interval. The mutation rate will play the role of the bifurcation parameter. It will be convenient to identify two specific threshold values for it.

Definition 4.1 (Critical mutation rates)**.** For $\gamma > 0$ and $s > 0$, let

$$\hat{u} := s \qquad \text{and} \qquad \check{u} := \frac{1}{\gamma}\left(\frac{s+\gamma}{2}\right)^2. \tag{4.2}$$

Note that $\check{u} - \hat{u} = (s-\gamma)^2/4\gamma \geq 0$, so that $\hat{u} \leq \check{u}$.

If $\nu_0 = 0$, the right-hand side of (4.1) reduces to

$$F(y) = (y-1)\,G(y) \tag{4.3}$$

with $G(y) = -\gamma y^2 + (s+\gamma)y - u$. Hence, the equilibria of (4.3) are

$$\bar{y}_1 = 1 \tag{4.4}$$

together with the real roots of G. Namely,

- for $\gamma = 0$ and $u \in [0, \infty)$,
$$\bar{y}_2 = \frac{u}{s} \tag{4.5}$$
(see Fig. 4.1, left);
- for $\gamma > 0$ and $u \in [0, \check{u}]$,
$$\bar{y}_2 = \frac{1}{2}\left(1 + \frac{s}{\gamma} - \sqrt{\sigma}\,\right) \quad \text{and } \bar{y}_3 = \frac{1}{2}\left(1 + \frac{s}{\gamma} + \sqrt{\sigma}\,\right), \tag{4.6}$$
where
$$\sigma := \left(1 + \frac{s}{\gamma}\right)^2 - 4\frac{u}{\gamma} \tag{4.7}$$
(see Fig. 4.1, middle and right).

Remark 4.4. Note that if $\gamma > 0$, then $u \leq \check{u}$ is equivalent to $\sigma \geq 0$. So, for $\gamma > 0$ and $u \in (\check{u}, \infty)$, the polynomial G has no real roots and $\bar{y}_1$ is the only equilibrium. ♢

In order to determine the stability (still for $\nu_0 = 0$), note that

$$F'(y) = G(y) + (y-1)\,G'(y),$$

where $G'(y) = -2\gamma y + s + \gamma$. Then,

- $F'(\bar{y}_1) = G(1) = s - u$, so $\bar{y}_1$ is stable (unstable) for $u < s$ $(u > s)$.
- $F'(\bar{y}_2) = (\bar{y}_2 - 1)G'(\bar{y}_2) = \gamma\sqrt{\sigma}(\bar{y}_2 - 1)$, so $\bar{y}_2$ is stable (unstable) if $\bar{y}_2 < 1$ $(\bar{y}_2 > 1)$.
- $F'(\bar{y}_3) = (\bar{y}_3 - 1)G'(\bar{y}_3) = -\gamma\sqrt{\sigma}(\bar{y}_3 - 1)$, so $\bar{y}_3$ is stable (unstable) if $\bar{y}_3 > 1$ $(\bar{y}_3 < 1)$.

We summarise this analysis in the following result.

Proposition 4.2 (Equilibria and stability). *Let* $\nu_0 = 0$ *and* $s, \gamma, u > 0$. *If* $\sigma \geq 0$, *the ODE* (4.1) *has equilibria* $\bar{y}_1$, $\bar{y}_2$, *and* $\bar{y}_3$ *given in* (4.4)–(4.6). *The corresponding stability is summarised in the following tables (where bold indicates equilibria that are stable in* $[0,1]$*).*

$s > \gamma$	$\sigma = 0$	$\sigma > 0$
$u < \hat{u}$	—	$0 < \bar{\mathbf{y}}_{\mathbf{2}} < \bar{y}_1 < \bar{y}_3$
$u = \hat{u}$	—	$\bar{\mathbf{y}}_{\mathbf{1}} = \bar{y}_2 < \bar{y}_3 = s/\gamma$
$u > \hat{u}$	$\bar{\mathbf{y}}_{\mathbf{1}} < \bar{y}_2 = \bar{y}_3$	$\bar{\mathbf{y}}_{\mathbf{1}} < \bar{y}_2 < \bar{y}_3$

$s = \gamma$	$\sigma = 0$	$\sigma > 0$
$u < \hat{u}$	—	$0 < \bar{\mathbf{y}}_{\mathbf{2}} < \bar{y}_1 < \bar{y}_3$
$u = \hat{u}$	$\bar{\mathbf{y}}_{\mathbf{1}} = \bar{\mathbf{y}}_{\mathbf{2}} = \bar{\mathbf{y}}_{\mathbf{3}}$	—
$u > \hat{u}$	—	—

$s < \gamma$	$\sigma = 0$	$\sigma > 0$
$u < \hat{u}$	—	$0 < \bar{\mathbf{y}}_{\mathbf{2}} < \bar{y}_1 < \bar{y}_3$
$u = \hat{u}$	—	$0 < \bar{\mathbf{y}}_{\mathbf{2}} = s/\gamma < \bar{y}_1 = \bar{y}_3$
$u > \hat{u}$	$\frac{1}{2} < \bar{y}_2 = \bar{y}_3 < \bar{\mathbf{y}}_{\mathbf{1}}$	$0 < \bar{\mathbf{y}}_{\mathbf{2}} < \bar{y}_3 < \bar{\mathbf{y}}_{\mathbf{1}}$

If $\sigma < 0$, $\bar{y}_1$ *is the only equilibrium and it is stable.*

Along the lines of [BW97, Sect. 3], we identify the following bifurcation phenomena with u as the bifurcation parameter. (Recall Remark 3.1 and the reference therein for more details on bifurcations.) For $0 < \gamma < s$ (resp. $0 < s < \gamma$), there is an exchange of stability, also known as transcritical bifurcation, at $u = \hat{u}$ of $\bar{y}_1$ and $\bar{y}_2$ (resp. of $\bar{y}_1$ and $\bar{y}_3$). When u surpasses the critical value $\hat{u}$, then $\bar{y}_1$ switches from unstable to stable; whereas $\bar{y}_2$ (resp. $\bar{y}_3$) switches from stable to unstable. For $s < \gamma$, there is an additional saddle-node bifurcation at $u = \check{u}$. There, the equilibria $\bar{y}_2$ and $\bar{y}_3$ (one stable, one unstable) collide and both vanish. If $\hat{u} = \check{u}$, we see a pitchfork bifurcation, where the unstable $\bar{y}_1$ passes through the collision point of $\bar{y}_2$ and $\bar{y}_3$ and becomes stable. (We refer the interested reader to [GH83, Ch. 3] for a general account of bifurcation theory for equilibria of ODEs.)

The equilibria directly lead to the long-term behaviour, which we now consider for $y_0 \in [0,1]$. Let $y(t;y_0)$ be the solution of (4.1) with $\nu_0 = 0$ and $s, \gamma, u > 0$. By the monotonicity of $y(\,\cdot\,;y_0)$, we infer that

$$y_\infty(y_0) := \lim_{t \to \infty} y(t;y_0) \tag{4.8}$$

exists and is always an equilibrium. The long-term type frequencies are characterised in the following corollary, which is a direct consequence of Remark 4.4 together with Proposition 4.2.

Corollary 4.3 (Asymptotic type frequencies). *Let* $y_0 \in [0,1]$, $\nu_0 = 0$, *and* $s, \gamma > 0$.

(i) If either $u < \hat{u}$ *or (*$s < \gamma$ *and* $u = \hat{u}$*),*

$$y_\infty(y_0) = \begin{cases} \bar{y}_2, & \text{if } y_0 \in [0,1), \\ \bar{y}_1, & \text{if } y_0 = \bar{y}_1. \end{cases}$$

(ii) If $s < \gamma$ *and* $u \in (\hat{u}, \check{u}]$,

$$y_\infty(y_0) = \begin{cases} \bar{y}_2, & \text{if } y_0 \in [0, \bar{y}_3), \\ \bar{y}_3, & \text{if } y_0 = \bar{y}_3, \\ \bar{y}_1, & \text{if } y_0 \in (\bar{y}_3, \bar{y}_1]. \end{cases}$$

(iii) If either $u > \check{u}$ *or (*$s \geq \gamma$ *and* $u \in [\hat{u}, \check{u}]$*), then for all* $y_0 \in [0,1]$,

$$y_\infty(y_0) = \bar{y}_1.$$

Remark 4.5. If $s < \gamma$ and $u = \check{u}$, we have $\bar{y}_2 = \bar{y}_3$ and

$$y_\infty(y_0) = \begin{cases} \bar{y}_2, & \text{if } y_0 \in [0, \bar{y}_2], \\ \bar{y}_1, & \text{if } y_0 \in (\bar{y}_2, \bar{y}_1]. \end{cases}$$

◇

Clearly, the equilibria $\bar{y}_i = \bar{y}_i(s, \gamma, u, \nu_0)$ $(i \in \{1,2,3\})$ are functions of s, γ, u and ν_0. The long-term type frequency $y_\infty(y_0) = y_\infty(y_0, s, \gamma, u, \nu_0)$ furthermore depends on y_0. By a straightforward application of L'Hôpital's rule, we see that $\bar{y}_2$ is continuous in γ at 0.

Proposition 4.4. *For $s, u > 0$ and $\nu_0 = 0$,*

$$\lim_{\gamma \to 0} \bar{y}_2(s, \gamma, u, 0) = \frac{u}{s}.$$

In particular,

$$\lim_{\gamma \to 0} y_\infty(y_0, s, \gamma, u, 0) = \begin{cases} \min\{\frac{u}{s}, 1\}, & \text{if } y_0 \in [0,1), \\ 1, & \text{if } y_0 = 1. \end{cases}$$

Let us recapitulate from [BW97, Sect. 3] the biological implications of Corollary 4.3. For $\gamma \leq s$, the fit type persists in the population for $u < \hat{u}$, but is lost for $u \geq \hat{u}$; this happens for any positive initial value and is an instance of the so-called error threshold [Eig71]. For $\gamma > s$, one has again persistence of the fit type for $u < \hat{u}$ and loss for $u > \check{u}$ for any positive initial value. But in the bistable regime $(\hat{u}, \check{u})$, the fit type will only persist if its initial frequency is at least $1 - \bar{y}_3$; otherwise it will be lost. In particular, a beneficial mutant arising in small frequency in a population that is otherwise unfit will not be able to establish itself.

Finally, let us briefly discuss the case $\nu_0 \in (0,1)$ (we do not explicitly consider the limiting case $\nu_0 = 1$; the biologically reasonable regime is $0 < \nu_0 \ll 1$, and $\nu_0 = 0$ is often a valid approximation). Since F is a cubic polynomial, explicit expressions for its roots are in principle available. In particular, a stability analysis as in the case $\nu_0 = 0$ is possible. But due to the length of expressions, the many case distinctions, and because we are primarily interested in the derivation of the bifurcation structure by genealogical means, we refrain from doing so here. Nevertheless, by Budan's theorem [Bud07], we can deduce a bound for the number of real roots in $[0,1]$.

Proposition 4.5. *Let $u, \gamma > 0$ and $\nu_0 \in (0,1)$. If $u < \hat{u}$ or $\gamma < s$, then F has exactly one root in* $[0,1]$.

Proof. Note first that if $u, \gamma > 0$ and $\nu_0, \nu_1 \in (0,1)$, then 0 and 1 are not roots of F. Recall that $F(y) = -\gamma y^3 + (2\gamma + s)y^2 - (\gamma + s + u)y + u\nu_1$. Define

$$F^+(y) := F(y+1) = -\gamma y^3 + (s - \gamma)y^2 + (s - u)y - u\nu_0.$$

Denote by o_F and o_{F^+} the number of sign changes of the coefficients of F and F^+, respect-

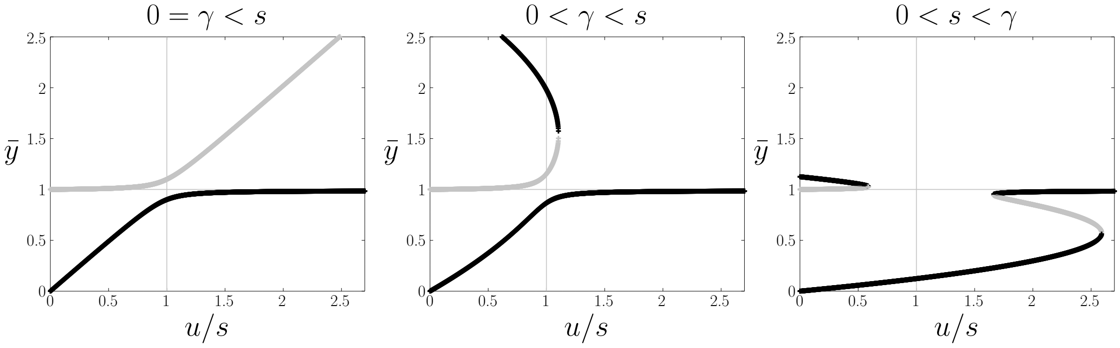


Figure 4.2. Equilibria of (4.1) evaluated numerically as functions of u/s for $\nu_0 = 1/100$ and $s = 1/30$. The left, middle, and right cases correspond to $\gamma = 0$, $\gamma = 1/60$, and $\gamma = 4/15$, respectively. Black lines: stable. Grey lines: unstable.

ively. Then, $o_F = 3$ and

$$o_{F+} = \begin{cases} 2, & \text{if } u < \hat{u} \text{ or } \gamma < s, \\ 0, & \text{if } s < \gamma \text{ and } u \geq \hat{u}. \end{cases}$$

Denote by r the number of roots of F in $(0,1)$. Now, Budan's theorem states that $r \leq o_F - o_{F+}$. On the other hand, we know from the discussion at the beginning of this section that there is at least one root in $[0,1]$. □

Remark 4.6. If (4.1) has three equilibria in $[0,1]$, then the equilibrium in the middle is unstable. To see this, note that F is positive before $\hat{y}$ and negative after $\check{y}$. Since there is an equilibrium in the middle, F has a positive derivative at this point. ♢

In contrast to $\nu_0 = 0$, if $\nu_0 \in (0,1)$, neither $y = 0$ nor $y = 1$ are equilibria and so both types coexist independently of y_0, see also Fig. 4.2. Yet, the long-term behaviour may again depend on y_0.

4.2 Mutation-selection equation as a limit of the Moran model: with pairwise interaction

Let us now relate the Moran model with pairwise interaction to the mutation-selection equation with pairwise interaction. The only restriction to the Moran model outlined in Section 2.1 is that we assume $u, s > 0$ throughout this chapter. To this end, we study the asymptotic behaviour of $Y^{(N)}/N$ as $N \to \infty$ without rescaling of parameters or time. This corresponds to the smsssi-limit (strong mutation–strong selection–strong interaction-limit). Subsequently, we consider the limit in the ASG of the Moran model to obtain the ASG in the smsssi-limit.
The following result provides the asymptotic behaviour of the stochastic process $Y^{(N)}/N$ and its connection to the deterministic model in the form of a dynamical law of large numbers.

Proposition 4.6 (Convergence Moran model). *Assume that* $\lim_{N\to\infty} Y_0^{(N)}/N = y_0 \in [0,1]$.

Then, for all $\varepsilon > 0$ *and* $t \geq 0$*, we have*

$$\lim_{N\to\infty} P\left(\sup_{\xi\leq t}\left|\frac{Y_\xi^{(N)}}{N} - y(\xi;y_0)\right| > \varepsilon\right) = 0,$$

where $y(\,\cdot\,;y_0)$ *is the solution of the IVP* (4.1)*, i. e.* $Y^{(N)}/N$ *converges to* $y(\,\cdot\,;y_0)$ *uniformly on compact time intervals in probability.*

Proof. The function F of (4.1) is Lipschitz continuous in $[0,1]$. We have $F(0) = u\nu_1 \geq 0$ and $F(1) = -u\nu_0 \leq 0$. Hence, the IVP (4.1) has a unique solution $y(\,\cdot\,;y_0)$ from $[0,\infty)$ to $[0,1]$. Note that we can rewrite the rates of $Y^{(N)}$ as $q_{Y^{(N)}}(k,k+\ell) = Nq\left(\frac{k}{N},\ell\right)$ for $\ell \in \mathbb{Z}\setminus\{0\}$, where $q : [0,1]\times\mathbb{Z}\setminus\{0\} \to \mathbb{R}$ is given by

$$q(y,1) = y(1-y) + (1-y)u\nu_0, \quad q(y,-1) = y(1-y)(1+s+\gamma(1-y)) + yu\nu_1,$$

together with $q(y,\ell) = 0$ for $|\ell| > 1$. Since q is continuous, $(Y^{(N)})_{N\geq 1}$ is a *density-dependent family* of Markov chains. Thus, in order to conclude, we only need to verify the following conditions of the dynamical law of large number for density-dependent families of Markov chains by Kurtz [Kur70, Thm. 3.1],

$$\sup_{y\in[0,1]}\sum_\ell |\ell| q(y,\ell) < \infty \quad \text{and} \quad \lim_{d\to\infty}\sup_{y\in[0,1]}\sum_{|\ell|>d}|\ell|q(y,\ell) = 0. \tag{4.9}$$

These conditions are clearly satisfied, which completes the proof. □

Remark 4.7. In the absence of interactions ($\gamma = 0$), the previous result coincides with [Cor17b, Prop. 3.1] (see also Eq. (3.2) in this thesis). ♢

The connection between the finite Moran models and the deterministic model provides a way to establish an ancestral picture for the mutation-selection equation with pairwise interaction. We start with the ancestral picture of the stochastic models and consider the same limit that connects the forward processes.

ASG in the smsssi-limit

The connection between the finite Moran models and the mutation-selection equation provided by Proposition 4.6 yields a way to construct the ASG in the deterministic model: We just let $N \to \infty$ in the ASG of the Moran model, which we described in Section 2.2. The resulting process will still be stochastic. We first describe the process in terms of the arrows in the spirit of the graphical representation in Section 2.1. Subsequently, we rigorously define the ASG in terms of a directed graph with labels.

Before we embark, note first that, since we assign types to the ASG in an exchangeable manner, when tracing back ancestries with the ASG in the Moran model, the relocation events do not affect the distribution of types, and so we can ignore them. For the same reason, we can assign a particular order to the lines in the ASG without changing the type distribution. In the following description of the process, it will be convenient to follow a lookdown-construction, which is the way we visualised the ASG in our figures. We do not

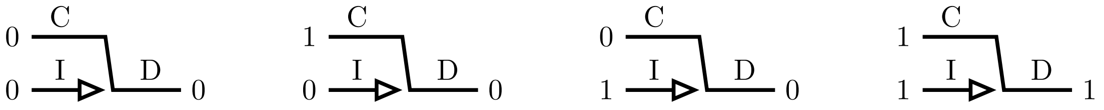


Figure 4.3. Bifurcation of a descendant line (D) into the continuing line (C) and the incoming line (I) along with the propagation of types according to the pecking order.

rely on this particular construction for the results that follow and, in fact, in the subsequent construction of the ASG as a directed graph with labelled vertices, we are *not* providing a lookdown construction. But for the moment, we proceed as in [Len+15] (see also [Cor17a] or Section 3.4 of this thesis) and construct the ASG in the smsssi-limit in a lookdown-like manner. This means that the lines are placed on consecutive levels, starting at level 1, according to a hierarchy reminiscent of the lookdown construction [DK99a; DK99b].
Since the rates of coalescence, simple collision, double-collision, and collision-bifurcation events vanish as N tends to infinity (they are $\mathcal{O}(1/N)$ per ordered pair of lines), these events will be absent in the asymptotic ASG, i. e. we will only see bifurcations, trifurcations, and mutations. (Recall that we use the word bifurcation twofold; in the context of dynamical systems and to describe the splitting of the ancestral lines, see Remark 3.2.) In particular, the ASG in the smsssi-limit of an initial sample of n individuals is distributed as n independent copies of an ASG started with a single individual. Thus, we can restrict ourselves to the evolution of an ASG starting with a single line. Let us first explain the ordering of the lines used in the construction of the ASG. At time 0, we place the single initial individual at level 1. We construct the ASG up to backward time r. A selective bifurcation at level i in the ASG in the smsssi-limit is represented by a horizontal open arrowhead at level i (see Fig. 4.3). The incoming branch emanates from the arrowhead and takes level i. All the lines at levels $\geq i$ are shifted one level upwards. This includes the descendant line, which then continues on level $i+1$ as the continuing line. A trifurcation event at level i in the ASG is represented by an arrowhead inscribed into an open square at level i (see Fig. 4.4). The incoming line emanates from the square at level i, the checking line emanates from the square and is placed at level $i+1$, and all the lines at levels $\geq i$ are shifted two levels upwards. Again, this includes the descendant line, which continues as the continuing line on level $i+2$.

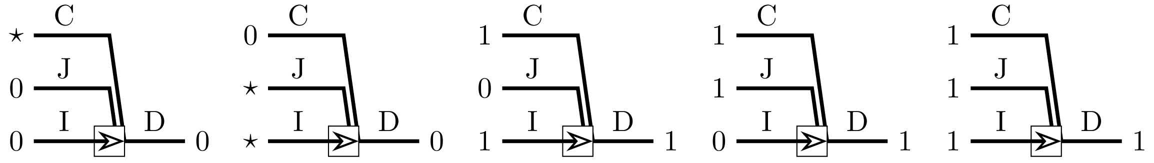


Figure 4.4. Trifurcation of the descendant line (D) into the continuing line (C), the checking line (J), and the incoming line (I) along with the associated type propagation rule ($\star$ stands for an arbitrary type.)

Each line in the ASG in the smsssi-limit independently bifurcates at rate s, thus increasing the number of lines by one. Each line trifurcates at rate γ independently of the others and independently of bifurcations. A trifurcation increases the number of lines by two. In addition, each line mutates to type 0 at rate $u\nu_0$ and to type 1 at rate $u\nu_1$ (see Fig. 4.5). Mutations occur independently on each line and independently of all the other events. For the remainder of this chapter, we refer to the ASG in the smsssi-limit just as ASG.

Figure 4.5. Propagating types across mutation events (note the parent independence).

To determine the true ancestry of the sample at present, we independently sample the type for each line at backward time r according to $(1-y_0, y_0)$ and propagate the types up to time 0 using the same rules as in the finite Moran model; see Definition 2.11. These rules are illustrated in Figs. 4.3–4.5. A visualisation of the ASG in $[0,r]$ is depicted in Fig. 4.9. In the following, we provide a construction of the AGS in the smsssi-limit as a directed graph with labelled vertices.

Construction of the ASG as a directed graph with labelled vertices

The following marked branching particle system is the basis of our construction of the ASG. Each particle occupies a site in $[0,1]$. Start at time $r=0$ with one particle on a site that is uniformly distributed in $[0,1]$. If there are n particles present at time r, then

- at rate sn, mark one of the existing particles chosen uniformly at random with $-\!\rhd$ and generate one new particle at a new site that is uniformly distributed in $[0,1]$ and independent from the other sites,
- at rate γn, mark one of the existing particles chosen uniformly at random with $-\!\boxed{\rhd}$ and generate two new particles each occupying a site uniformly distributed in $[0,1]$ and independent from the other sites,
- at rate $nu\nu_0$, mark one of the existing particles chosen uniformly at random with $\circ$,
- at rate $nu\nu_1$, mark one of the existing particles chosen uniformly at random with $\times$.

Note that, by construction, a particle at site w has an associated birth time b_w. Set the birth time of the particle present at time 0 to 0. Let $\mathcal{W}$ be the set of occupied sites between $r=0$ and $r=\infty$. The ASG is the following (infinite and uncountable) directed acyclic graph $G=(V,E)$ with a function l labelling the vertices, where

$$V \subseteq \mathbb{R}_{\geq 0} \times [0,1], \quad E \subseteq V \times V, \quad \text{and} \quad l : V \to \{-\!\rhd, -\!\boxed{\rhd}, \diamond\!-, \bullet\!-, \circ, \times, \emptyset\}.$$

As in the previous chapters, for $(r,w) \in V$, we refer to r as the time component, w the site, and $l(r,w)$ the label. The set of vertices is given by $V = \{(r,w) : r \in [b_w, \infty), w \in \mathcal{W}\}$, where $(r,w) \in \mathbb{R}_{\geq 0} \times [0,1]$ has label $-\!\rhd$, $-\!\boxed{\rhd}$, $\circ$, and $\times$ if the particle at site w is marked at time r with $-\!\rhd$, $-\!\boxed{\rhd}$, $\circ$, and $\times$, respectively. Let $\{S_i\}$ and $\{I_i\}$ be the set of times at which the number of particles increases by one and two, respectively. For I_i there are two newly occupied sites w_1 and w_2. Set $l(I_i, w_1) = \diamond\!-$ and $l(I_i, w_2) = \bullet\!-$. All remaining vertices are labelled with $\emptyset$. The set of edges E is given as follows.

- For S_i and $v, w \in \mathcal{W}$ such that $b_v = S_i$ and $l(S_i, w) = -\!\rhd$, we have $((S_i, w), (S_i, v)) \in E$.

- For I_i and $v_1, v_2, w \in \mathcal{W}$ with $b_{v_1} = b_{v_2} = I_i$, $v_1 \neq v_2$, $l(I_i, v_1) = \diamond\!\!-$, $l(I_i, v_2) = \bullet\!\!-$, and $l(I_i, w) = -\!\!\boxed{\triangleright}$, we have $((I_i, w), (I_i, v_1)) \in E$ and $((I_i, w), (I_i, v_2)) \in E$.
- For $w \in \mathcal{W}$ and $r, t \in [b_w, \infty)$ with $r < t$ such that $l(\tau, w) \notin \{-\!\!\triangleright, -\!\!\boxed{\triangleright}, \circ, \times\}$ for $\tau \in (r, t)$, we have $((r, w), (t, w)) \in E$.

The ASG in $[0, \infty)$ is then $\mathcal{A}_\infty := (G, l)$. Denote by $\mathcal{A}_r$ the restriction of $\mathcal{A}_\infty$ to the graph and label function that arises if only vertices with time component at most r are considered. $\mathcal{A}_r$ is called the ASG in $[0, r]$. We refer to the vertices in $\mathcal{A}_r$ with time component 0 as root(s) and to the vertices with time component r as leaves. The set of vertices sharing the same site is called a line.
A reversal of the direction of the edges and an assignment of types to the leaves allows us to propagate types and ancestral sites as described in Definition 2.11. In particular, the notion of ancestry translates into this setting. In all our figures, we use the description of the ASG from the previous subsection in terms of arrows, i.e. the edges that join different sites are already reverted.

The description of the ASG in terms of arrows from the previous section suffices for most of our purposes. In particular, when we prune and stratify the ASG in the next section, we primarily argue on the basis of the representation in terms of the arrows to ease the notation. But one should keep in mind that every operation defined in the next section has a counterpart as an operation acting on an ASG as a directed graph with labelled vertices.

Remark 4.8. We will often identify an ASG starting with n lines with the collection $\mathcal{A} = (\mathcal{A}(i))_{i \in [n]}$, where $\mathcal{A}(i)$ denotes the ASG associated with the ith line. $\diamondsuit$

Type distribution of the root in the ASG

Consider $\mathcal{A}_t$, i.e. an ASG in the time interval $[0, t]$. For $y_0 \in [0, 1]$, let $H(\mathcal{A}_t, y_0)$ be the probability that all lineages at the present are unfit if the types at the leaves are sampled according to $(1 - y_0, y_0)$. A natural way of computing $H(\mathcal{A}_t, y_0)$ is to determine first those assignments of types to the lines that lead to an unfit descendant and then to evaluate the probability of observing these assignments if we independently sample according to $(1 - y_0, y_0)$. The next lemma summarises some elementary properties of the function H.

Lemma 4.7. *Let $\mathcal{A}_t$ be an ASG in $[0, t]$ starting with n lines and, for $i \in \{1, ..., n\}$, let $\mathcal{A}_t(i)$ the ASG in $[0, t]$ associated with the ith line. Then, for all $y_0 \in [0, 1]$,*

$$H(\mathcal{A}_t, y_0) = \prod_{i=1}^{n} H(\mathcal{A}_t(i), y_0).$$

Moreover, for $n = 1$, if $T_\star$ is the time of the first event in $\mathcal{A}_t$, then

(1) if the event at time $T_\star$ is a deleterious mutation, then $H(\mathcal{A}_t, y_0) = 1$.

(2) if the event at time $T_\star$ is a beneficial mutation, then $H(\mathcal{A}_t, y_0) = 0$.

(3) if the event at time $T_\star$ is a selective branching, we denote by $\mathcal{A}^1_{T_\star,t}$ and $\mathcal{A}^2_{T_\star,t}$ the ASGs starting from the continuing and incoming line, respectively, at time $T_\star$ and ending at time t. Then

$$H(\mathcal{A}_t, y_0) = H(\mathcal{A}^1_{T_\star,t}, y_0) H(\mathcal{A}^2_{T_\star,t}, y_0).$$

(4) if the event at time $T_\star$ is a trifurcation, we denote by $\mathcal{A}^1_{T_\star,t}$, $\mathcal{A}^2_{T_\star,t}$, and $\mathcal{A}^3_{T_\star,t}$ the ASGs starting from the continuing, incoming, and checking line, respectively, at time $T_\star$ and ending at time t. Then,

$$H(\mathcal{A}_t, y_0) = H(\mathcal{A}^1_{T_\star,t}, y_0)\left[H(\mathcal{A}^2_{T_\star,t}, y_0) + H(\mathcal{A}^3_{T_\star,t}, y_0) - H(\mathcal{A}^2_{T_\star,t}, y_0) H(\mathcal{A}^3_{T_\star,t}, y_0)\right].$$

Remark 4.9. Let us at this point make a connection to [MSS18b]. In their context, deleterious mutations and trifurcations are captured by the local maps `dth` ('deaths') and `cob` ('cooperative branchings'), respectively. Hence, if $s = \nu_0 = 0$, $H(\mathcal{A}_t, y_0)$ corresponds to the concatenation of their higher-level maps $\widehat{\texttt{dth}}$ and $\widehat{\texttt{cob}}$, respectively. In particular, (1) and (4) of our Lemma 4.7 coincide with [MSS18b, Eq. (1.84)]. ◇

Proof of Lemma 4.7. The lines in the ASG do not interact, so the first result follows by independence. Now fix $n = 1$. For (1) and (2), note that if the first event is a mutation, the type of the line at time $r = 0$ is independent of the type assignment at time $r = t$ due to parent independence (cf. Fig. 4.5). If the first event is a selective event, the individual at time 0 is of type 1 if and only if the two lines involved in the event are of type 1. This leads to (3). If the first event is an interactive event, the individual at time $r = 0$ is of type 1 if and only if the continuing line is unfit and either the checking or the incoming line is unfit. This leads to (4). □

One of the main aims in the remainder of the chapter is to recover the bifurcation structure of the mutation-selection equation by means of the ancestral process and to analyse the ancestral type distribution in this framework. To connect the forward and the backward process, we want to derive a duality relation of the form

$$\mathbb{E}_{\mathcal{A}_0}[H(\mathcal{A}_t, y_0)] = H(\mathcal{A}_0, y(t; y_0)). \tag{4.10}$$

If we define $\tilde{y}(t; y_0) := H(\mathcal{A}_t, y_0)$, then we clearly have $H(\mathcal{A}_0, \tilde{y}(t; y_0)) = H(\mathcal{A}_t, y_0)$. But a priori, it is not clear how $\tilde{y}(t; y_0)$ relates to $y(t; y_0)$. One strategy to prove the duality relation is by noting that the ASG together with the rule of propagating the types can be formulated in the framework of recursive tree processes (see Remark 4.9). In particular, the above relation then follows as an application of [MSS18b, Thm. 6]. We take a different route in remainder of the chapter and retain more information of the underlying tree structure. This also allows us to analyse the ancestral type distribution, which we do in Section 4.5.

4.3 Stratified ASG

Quite generally there are two approaches to compute the function H. The natural way is to determine first the type assignments to the lines in the ASG that lead to an unfit descendant and then to evaluate the probability of observing these assignments if we independently sample according to $(1 - y_0, y_0)$. This is the approach pursed by Mach [Mac17], but the general idea is also present in the work of Dawson and Greven [DG14, Ch. 5.5]. In contrast, we aim at resolving all information contained in the tree on the spot. This leads to a reduction and pruning of the tree. A subsequent reorganization leads to a stratification of the ASG into distinct regions. This is the topic of the current section.

4.3.1 Motivation: The case without interaction

Let us recall the appropriate reduction of the ASG in the case of frequency-independent selection from Section 3.3.

In the absence of interaction and mutation, a sampled individual is of type 1 if and only if all its potential ancestors are of type 1; this follows from the pecking order and holds regardless of the tree structure. A mutation determines the type of the line on which it occurs, so this line need not be traced back further and hence can be pruned. Moreover, if a beneficial mutation occurs on a line that is not yet pruned, the type of the descendant will be fit so that we can stop reading the ASG, and we send the process to a cemetery point Δ. This reasoning gives rise to the killed ASG of Section 3.3. Recall, that its line-counting process $R = (R_r)_{r \geq 0}$ (Definition 3.1) is a Markov process on $\mathbb{N}_0 \cup \{\Delta\}$ with transition rates

$$q_R(k, k+1) = ks, \quad q_R(k, k-1) = ku\nu_1, \quad q_R(k, \Delta) = ku\nu_0, \qquad k \in \mathbb{N}_0.$$

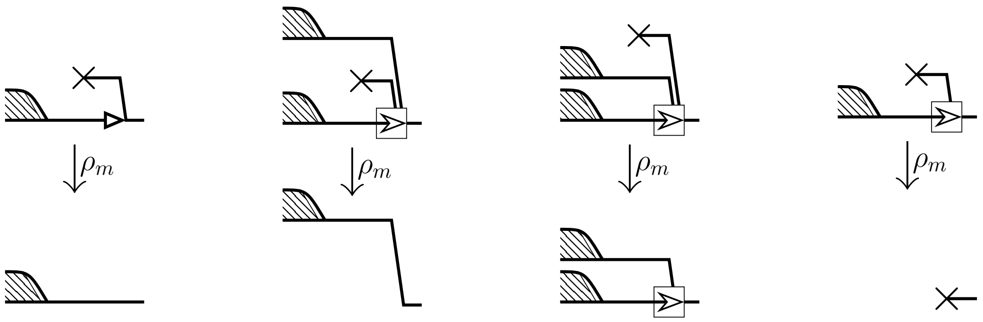
Absorption of R in 0 implies that all individuals in the sample are of type 1; whereas absorption of R in Δ implies that at least one individual in the sample is fit. The process R is in moment duality with the mutation-selection model without interaction (see Theorem 3.2), that is we have for $y_0 \in [0,1]$ and $n \in \mathbb{N}_0^\Delta$,

$$y(t; y_0)^n = \mathbb{E}_n\big[y_0^{R_t}\big]. \tag{4.11}$$

4.3.2 Reducing the interactive ASG

In the interactive case, a single fit line does not necessarily lead to a fit sampled individual (see Fig. 4.4), and hence we cannot use the same reasoning as in the non-interactive case. In particular, counting lines is not sufficient; rather, the tree structure plays an eminent role. A first step to circumvent this problem is to get rid of the mutation events present in the ASG. We now explain this in detail.

As in the non-interactive case, a mutation on a line in the ASG determines the effect of that line on the type of the sampled individual, and, therefore, we need not trace back its ancestry any further. In addition, the type assigned to the line by the mutation will propagate (forward in time), resolving on its way some of the selective and/or interactive events it encounters, by following the local rules presented in the previous section. A particularly interesting situation occurs when the first event after a trifurcation is a deleterious mutation in the corresponding continuing line. In this case, the type of the descendant line depends only on the type of the incoming and the checking line. For this reason, we consider a new type of event, which we call *interactive bifurcation* and which corresponds to an interactive event where the continuing line is pruned due to a deleterious mutation. Moreover, we denote as a *generalised ASG* an ASG consisting of selective and interactive bifurcations, trifurcations, and mutations. In particular, a generalised ASG is then again a directed graph with labelled vertices. In an interactive bifurcation a vertex with label ⍈ has never an incoming edge with label $\emptyset$. (But in a trifurcation, a vertex with label ⍈ still has an incoming edge with label $\emptyset$.) We extend the definition of the type propagation to generalised ASGs by adding the rule that in an interactive bifurcation the descendant line is fit if and only if both checking and incoming line are fit. In particular, this extends the definition of H to generalised ASGs. In what follows, we primarily use the description of the ASG



(1.a) (1.b) (1.c) (1.d)

Figure 4.6. 1-step pruning described by (1.a)-(1.d).

in terms of the arrows to ease the notation, which is also the representation we use in our figures. All operations acting on an ASG with arrows translate to an operation acting on an ASG as a directed graph with labelled vertices. We now define appropriate reduction- and pruning-operations that allow us to resolve the local effects of mutations in a generalised ASG. After their definition, the invariance of the type at the root under these operations for a given deterministic assignment of types to the leaves of the ASG is the content of Lemma 4.9.

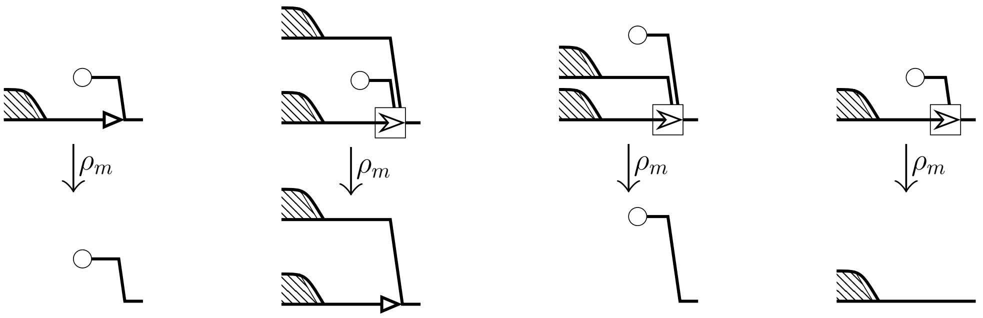
Definition 4.8 (Reduction and pruning)**.** Let $\mathcal{A}$ be a generalised ASG. The reduced ASG of $\mathcal{A}$, denoted by $\text{✂}(\mathcal{A})$, is obtained by removing from $\mathcal{A}$ all the sub-ASGs arising to the left of mutation events. In particular, $\text{✂}(\mathcal{A})$ has again a representation as a directed graph with labelled vertices. We refer to $\bar{\mathcal{A}}$ as a reduced ASG if there exists some generalised ASG $\mathcal{A}$ such that $\bar{\mathcal{A}} = \text{✂}(\mathcal{A})$. Assume that $\text{✂}(\mathcal{A})$ contains at least one branching event and at least one mutation event. If m denotes a mutation event in $\text{✂}(\mathcal{A})$, we call $\rho_m(\mathcal{A})$ the 1-step pruning of $\text{✂}(\mathcal{A})$ at m and define it as follows (cf. Figs. 4.6 and 4.7).

(1) If m is a deleterious mutation and

  (a) the event preceding m is a selective branching, we remove the arrow and the line segment between the two events, and we connect the other line involved in the branching to the descendant line.
  (b) the event preceding m is a trifurcation and the line involved in the mutation is incoming (resp. checking) to this event, we remove the line segment between the two events, the sub-ASG arising from the checking (resp. incoming) line, and the interactive arrow.
  (c) the event preceding m is a trifurcation and the line involved in the mutation is continuing to this event, we remove the line segment between the two events, thus transforming the trifurcation into an interactive bifurcation.
  (d) the event preceding m is an interactive bifurcation, we remove the sub-ASG arising at the interactive event and replace the interactive event by a deleterious mutation.

(2) If m is a beneficial mutation and

  (a) the event preceding m is a selective branching and the line involved in the mutation is continuing (resp. incoming), we remove the sub-ASG arising from the



(2.a) (2.b) (2.c) (2.d)

Figure 4.7. 1-step pruning described by (2.a)-(2.d)

incoming (resp. continuing) line and connect the remaining line to the descendant line.

(b) the event preceding m is a trifurcation and the line involved in the mutation is incoming (resp. checking) to this event, we remove the sub-ASG arising from the incoming (resp. checking) line and connect the checking (resp. incoming) line to the continuing line via a selective arrow.

(c) the event preceding m is a trifurcation and the line involved in the mutation is continuing to this event, we remove the sub-ASGs arising from the checking and incoming lines.

(d) the event preceding m is an interactive bifurcation and the line involved in the mutation is incoming (resp. checking) to this event, we remove the line segment between the two events and connect the checking (resp. incoming) line to the descendant line.

We call $\hat{\mathcal{A}}$ a pruning of $\mathcal{A}$ if $\hat{\mathcal{A}} = \text{✂}(\mathcal{A})$ or if it is obtained by successive 1-step prunings of $\text{✂}(\mathcal{A})$. Moreover, a pruning of $\mathcal{A}$ is called *total* if it is composed of a generalised ASG without mutations and/or a collection of lines ending in mutation events. We write $\mathcal{A} \sim \circ$ if $\mathcal{A}$ consists of a single line ending in a beneficial mutation. Similarly, we write $\mathcal{A} \sim \times$ if $\mathcal{A}$ consists of a single line ending in a deleterious mutation.

Remark 4.10. Note that a 1-step pruning reduces the number of events in a reduced ASG at least by one. Therefore, a total pruning is obtained after a finite number of 1-step prunings. ♢

Remark 4.11. Consider Fig. 4.6. If we want to determine the type of the descendent in case (1.a), both lines play a symmetric role. In the figure, we only present one of the two cases. Similarly, checking and incoming line have an equal effect on the type of the descendent in cases (1.b) and (1.d). The same reasoning applies in Fig. 4.7 for cases (2.a), (2.b), and (2.d). Note that, the reduced ASG in the current form is not an appropriate structure to determine the ancestor of the root, because the line-segments arising to the left of mutations might be relevant to determine the type of the ancestor. ♢

Remark 4.12. If $\mathcal{A}$ is a generalised ASG, then $\text{✂}(\mathcal{A})$ is, by construction, embedded into $\mathcal{A}$ (in the obvious way). Similarly, if $\bar{\mathcal{A}}$ is a reduced ASG and $\rho_m(\bar{\mathcal{A}})$ is a 1-step pruning of it, then the lines of $\rho_m(\bar{\mathcal{A}})$ can be embedded into $\bar{\mathcal{A}}$. However, the type of the connections

Figure 4.8. Resolving first the first (resp. second) mutation from the left leads to the total pruning on the left (resp. right). Note the difference in length of the line segment to the beneficial mutation after pruning.

between lines can differ between $\rho_m(\bar{\mathcal{A}})$ and $\bar{\mathcal{A}}$ (see Fig. 4.9). This identification of lines will be used implicitly all along in this section. $\diamondsuit$

Remark 4.13. Two total prunings are not necessarily identical. For an example see Fig. 4.8. $\diamondsuit$

Remark 4.14. An example for a total pruning is the following. Consider a generalised ASG $\mathcal{A}_t$ that starts with two lines. In one line the first event is a mutation. The ASG arising from the other line contains no mutations. If we remove the part beyond the first mutation on the first line, then this is a total pruning. $\diamondsuit$

Consider the reduced ASG that arises from $\mathcal{A}_t$. Assign types to the lines that are present at time backward time t (the lines ending before time t get their types from the corresponding mutation events). Next, propagate types as described in Section 4.2 (see also Definition 2.11). This extends the definition of H to reduced ASGs.

Remark 4.15. The statement of Lemma 4.7 remains true if $\mathcal{A}_t$ is a reduced ASG or a generalised ASG in $[0, t]$. Moreover, points (1), (2), (3) and (4) are complemented by

(5) if the event at time $T_\star$ is an interactive bifurcation, we denote by $\mathcal{A}^2_{T_\star,t}$ and $\mathcal{A}^3_{T_\star,t}$ the ASGs starting at time $T_\star$ from the incoming and checking line, respectively. Then,

$$H(\mathcal{A}_t, y_0) = H(\mathcal{A}^2_{T_\star,t}, y_0) + H(\mathcal{A}^3_{T_\star,t}, y_0) - H(\mathcal{A}^2_{T_\star,t}, y_0) H(\mathcal{A}^3_{T_\star,t}, y_0). \qquad \diamondsuit$$

The next lemma states that for any given type assignment to the leaves of an ASG, the type of the root is invariant under reduction and 1-step prunings.

Lemma 4.9 (Invariance of root type under reduction/pruning)**.** *Consider a generalised ASG $\mathcal{A}_t$ for some $t \geq 0$. For any given assignment of types to the leaves, the type of the root is invariant under reduction and 1-step prunings. In particular,*

$$H(\mathcal{A}_t, y_0) = H(\text{✂}(\mathcal{A}_t), y_0) = H(\rho_m(\mathcal{A}_t), y_0),$$

for any mutation m in $\text{✂}(\mathcal{A}_t)$, i.e. H is invariant under reduction and 1-step prunings.

Proof. By Lemma 4.7, we can without loss of generality assume that our ASG starts with one line. Let $\mathcal{A}_t$ be a generalised ASG in $[0, t]$ started with a single line and consisting of n lines at time t, denoted by $\ell_1, ..., \ell_n$. For $c := (c_i)_{i \in [n]} \in \{0, 1\}^n$, denote by $v_c(\mathcal{A}_t) \in \{0, 1\}$

the type of the single line present at time 0 in $\mathcal{A}_t$ if at time t, for each $i \in [n]$, line ℓ_i is assigned type c_i. From construction, if $y := (y_i)_{i\in[n]}$, $z := (z_i)_{i\in[n]} \in \{0,1\}^n$ are such that $y_i = z_i$ for all i with ℓ_i being present in $\text{✂}(\mathcal{A}_t)$, then $v_y(\mathcal{A}_t) = v_z(\mathcal{A}_t)$. Therefore, the type of the root is invariant under reductions.

We now prove that the type of the root is invariant under 1-step prunings. We only consider prunings induced by a deleterious mutation. The prunings induced by beneficial mutation can be treated analogously.

The first part of the proof implies that we can work on the basis of a reduced ASG. To ease the notation, we write for the remainder of this proof $\mathcal{A}_t$ instead of $\text{✂}(\mathcal{A}_t)$. First, assume that $\mathcal{A}_t$ is a reduced ASG in $[0,t]$ consisting of at least one branching and at least one mutation event. We have to show that for every mutation m on a line of $\mathcal{A}_t$, $v_c(\mathcal{A}_t) = v_c(\rho_m(\mathcal{A}_t))$. Denote by $T_\star$ the time of the last branching that involves the line with mutation m and by ℓ the line with this branching (i.e. the descendent line). Let $T_\star^\varepsilon := T_\star - \varepsilon$ for $\varepsilon > 0$ small enough such that there is no other event on line ℓ in $[T_\star^\varepsilon, T_\star]$, i.e. the time just before the branching. Let $\mathcal{A}^\ell_{T_\star^\varepsilon,t}$ be the ASG in $[T_\star^\varepsilon, t]$ arising from line ℓ in $\mathcal{A}_t$, i.e. the ASG arising from line ℓ directly before time $T_\star$. Then, by a slight abuse of notation, we write $v_c(\mathcal{A}^\ell_{T_\star^\varepsilon,t})$ for the type of its root if the types of the leaves are assigned according to c restricted to the lines in $\mathcal{A}^\ell_{T_\star^\varepsilon,t}$ at time t. In particular, this is the type on the descendent line at the branching event on line ℓ at time $T_\star$. Since $\rho_m(\mathcal{A}_t)$ only affects the lines contained in $\mathcal{A}^\ell_{T_\star^\varepsilon,t}$, it suffices to prove that $v_c(\mathcal{A}^\ell_{T_\star^\varepsilon,t}) = v_c(\rho_m(\mathcal{A}^\ell_{T_\star^\varepsilon,t}))$, where $v_c(\rho_m(\mathcal{A}^\ell_{T_\star^\varepsilon,t}))$ is the type of the root of $\rho_m(\mathcal{A}^\ell_{T_\star^\varepsilon,t})$ if the types are assigned according to c restricted to the lines in $\rho_m(\mathcal{A}^\ell_{T_\star^\varepsilon,t})$ at time t. It then follows that $v_c(\mathcal{A}_t) = v_c(\rho_m(\mathcal{A}_t))$.
Hence, we now prove $v_c(\mathcal{A}^\ell_{T_\star^\varepsilon,t}) = v_c(\rho_m(\mathcal{A}^\ell_{T_\star^\varepsilon,t}))$, which boils down to proving the invariance of the type at the root in the generalised ASGs under the 1-step prunings depicted in Fig. 4.6.

(1.a) If the branching is a selective bifurcation and m occurs in the continuing line, let $\mathcal{A}^{\ell,1}_{T_\star,t}$ and $\mathcal{A}^{\ell,2}_{T_\star,t}$ be the reduced ASGs in $[T_\star, t]$ arising at this branching event from the continuing and incoming line, respectively. We have $\mathcal{A}^{\ell,1}_{T_\star,t} \sim \times$, and therefore $v_c(\mathcal{A}^{\ell,1}_{T_\star,t}) = 1$. The type propagation rule at a selective bifurcation implies then $v_c(\mathcal{A}^\ell_{T_\star^\varepsilon,t}) = 0$ if and only if $v_c(\mathcal{A}^{\ell,2}_{T_\star,t}) = 0$. On the other hand, $v_c(\rho_m(\mathcal{A}^\ell_{T_\star^\varepsilon,t})) = 0$ if and only if

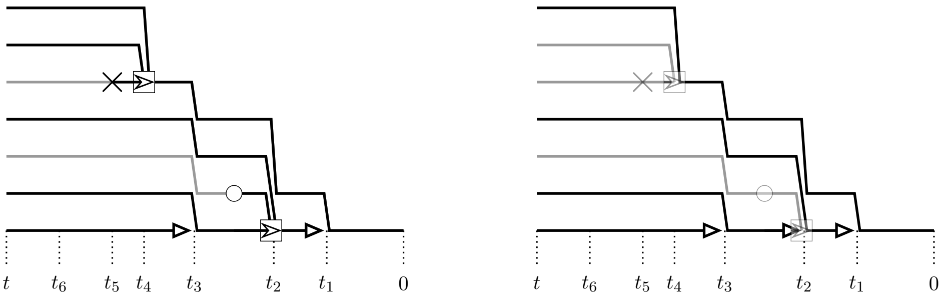


Figure 4.9. A realisation of the ASG $\mathcal{A}_t$ in the smsssi-limit in the time interval $[0,t]$ (grey and black, left and right), its reduced ASG $\text{✂}(\mathcal{A})$ (black, left), and its total pruning $\hat{\mathcal{A}}$ (black, right).

$v_c(\mathcal{A}^{\ell,2}_{T_\star,t}) = 0$. Hence, $v_c(\mathcal{A}^{\ell}_{T^\varepsilon_\star,t}) = v_c(\rho_m(\mathcal{A}^{\ell}_{T^\varepsilon_\star,t}))$.

(1.b) If the branching is a trifurcation and m occurs in the incoming line, denote by $\mathcal{A}^{\ell,1}_{T_\star,t}$, $\mathcal{A}^{\ell,2}_{T_\star,t}$, and $\mathcal{A}^{\ell,3}_{T_\star,t}$ the reduced ASGs in $[T_\star,t]$ arising from the continuing, checking, and incoming line, respectively. We have $\mathcal{A}^{\ell,3}_{T_\star,t} \sim \times$, and therefore $v_c(\mathcal{A}^{\ell,3}_{T_\star,t}) = 1$. The type propagation rule at an interactive trifurcation then implies $v_c(\mathcal{A}^{\ell}_{T^\varepsilon_\star,t}) = 1$ if and only if $v_c(\mathcal{A}^{\ell,1}_{T_\star,t}) = 1$. On the other hand, $v_c(\rho_m(\mathcal{A}^{\ell}_{T^\varepsilon_\star,t})) = 1$ if and only if $v_c(\mathcal{A}^{\ell,1}_{T_\star,t}) = 1$. Hence, $v_c(\mathcal{A}^{\ell}_{T^\varepsilon_\star,t}) = v_c(\mathcal{A}^{\ell,1}_{T_\star,t}) = v_c(\rho_m(\mathcal{A}^{\ell}_{T^\varepsilon_\star,t}))$. By symmetry, the same argument applies if the mutation is on the checking line.

(1.c) If the branching is a trifurcation and m occurs in the continuing line, we have $\mathcal{A}^{\ell,1}_{T_\star,t} \sim \times$, and therefore $v_c(\mathcal{A}^{\ell,1}_{T_\star,t}) = 1$. The type propagation rule at an interactive trifurcation then implies $v_c(\mathcal{A}^{\ell}_{T^\varepsilon_\star,t}) = 0$ if and only if $v_c(\mathcal{A}^{\ell,2}_{T_\star,t}) = 0$ and $v_c(\mathcal{A}^{\ell,3}_{T_\star,t}) = 0$. On the other hand, $v_c(\rho_m(\mathcal{A}^{\ell}_{T^\varepsilon_\star,t})) = 0$ if and only if $v_c(\mathcal{A}^{\ell,2}_{T_\star,t}) = 0$ and $v_c(\mathcal{A}^{\ell,3}_{T_\star,t}) = 0$. Hence, $v_c(\mathcal{A}^{\ell}_{T^\varepsilon_\star,t}) = v_c(\rho_m(\mathcal{A}^{\ell}_{T^\varepsilon_\star,t}))$.

(1.d) If the branching is an interactive bifurcation and m occurs on the incoming line, we have $\mathcal{A}^{\ell,2}_{T_\star,t} \sim \times$ and therefore $v_c(\mathcal{A}^{\ell,2}_{T_\star,t}) = 1$. The type propagation rule at an interactive bifurcation then implies $v_c(\mathcal{A}^{\ell}_{T^\varepsilon_\star,t}) = 1$. On the other hand, $v_c(\rho_m(\mathcal{A}^{\ell}_{T^\varepsilon_\star,t})) = 0$ if and only if $v_c(\mathcal{A}^{\ell,2}_{T_\star,t}) = 0$ and $v_c(\mathcal{A}^{\ell,3}_{T_\star,t}) = 0$. Hence, also $v_c(\rho_m(\mathcal{A}^{\ell}_{T^\varepsilon_\star,t})) = 1$. By symmetry, the same argument applies if the mutation is on the checking line.

To determine $H(\mathcal{A}_t, y_0)$, the type at every leaf is independently sampled according to $(1-y_0, y_0)$. Hence, from the invariance of the type of the root under reductions and 1-step prunings, the invariance of H under these operation follows. ☐

Lemma 4.10. *Let $\mathcal{A}_t$ be an ASG in $[0,t]$ starting with one line. Two total prunings of $\mathcal{A}_t$ are either identical or both consist of a single line ending in the same type of mutation.*

We provide the proof of Lemma 4.10 in Section 4.6. The previous result motivates the following definition.

Definition 4.11 (Pruned ASG)**.** Let $\mathcal{A}_t$ be an ASG in the time interval $[0,t]$ starting with one line. The *pruned ASG* associated with $\mathcal{A}_t$, denoted by $\rho(\mathcal{A}_t)$, is defined as follows. Let $\bar{\mathcal{A}}_t$ be a total pruning of $\mathcal{A}_t$. If $\bar{\mathcal{A}}_t \sim \times$, set $\rho(\mathcal{A}_t) := \emptyset$. If $\bar{\mathcal{A}}_t \sim \circ$, set $\rho(\mathcal{A}_t) := \Delta$, where Δ denotes a cemetery point. In the remaining case, we set $\rho(\mathcal{A}_t) := \bar{\mathcal{A}}_t$. Moreover, if $\mathcal{A}_t := (\mathcal{A}_t(i))_{i\in[n]}$ is an ASG in the time interval $[0,t]$ starting with n lines, then the pruned ASG associated with $\mathcal{A}_t$ is given by

$$\rho(\mathcal{A}_t) := \begin{cases} \Delta, & \text{if } \rho(\mathcal{A}_t(i)) = \Delta \text{ for some } i \in [n], \\ \emptyset, & \text{if } \rho(\mathcal{A}_t(i)) = \emptyset \text{ for all } i \in [n], \\ (\rho(\mathcal{A}_t(i)))_{i\in I_n}, & \text{otherwise,} \end{cases}$$

where $I_n := \{i \in [n] : \rho(\mathcal{A}_t(i)) \neq \emptyset\}$.

We set $H(\Delta, y_0) := 0$ and $H(\emptyset, y_0) := 1$, for $y_0 \in [0,1]$.

Lemma 4.12. *Let $\mathcal{A}$ be an ASG in a finite interval. For a given type assignment at the leaves of $\mathcal{A}$, the roots of $\mathcal{A}$ and $\rho(\mathcal{A})$ have the same type. In particular,*

$$H(\mathcal{A}, y_0) = H(\rho(\mathcal{A}), y_0), \qquad y_0 \in [0,1].$$

Proof. It follows by iterating Lemma 4.9. □

4.3.3 Stratifying the ASG

Pruning the ASG allows us to get rid of the mutation events. In the next step we partition or *stratify* the pruned ASG into regions within which types propagate as in the non-interactive case.
Let us explain the rationale. We say that two lines in a generalised ASG belong to the same *region* if they are connected only by means of selective arrows. A trifurcation gives rise to two new lines each of which is connected to the region of the continuing line via an incoming or checking arrow, respectively. In particular, each of the two new lines starts a new region. In this way, the line at the origin of a given region is assigned type 1 if and only if all the lines in that region are assigned type 1 and, in addition, at any trifurcation in this region (where the line with the trifurcation is continuing and hence part of the said region) either the incoming or the checking line is assigned type 1. The rationale at interactive bifurcations is analogous. Altogether, it is enough to keep track of the sizes of the regions and of the tree structure inherent to the connections between them. In order to encode this information, we use weighted ternary trees.

In what follows, a rooted tree will be an undirected, acyclic, finite, connected graph in which we identify a special vertex that is called the root. A ternary tree is a rooted tree in which each vertex that is not the root is labelled as left, middle, or right child. Rooted trees induce a natural order on the vertices by the distance to the root (this justifies the notion of children). Denote by Ξ the set of all ternary trees. For a ternary tree $\tau \in \Xi$, denote by V_τ the set of its vertices and by $L_\tau \subseteq V_\tau$ the set of its leaves. A weighted ternary tree is a pair $\mathcal{T} = (\tau, m_\tau)$, where $\tau \in \Xi$ and $m_\tau : L_\tau \to \mathbb{N}_0$. Denote by Υ the set of weighted ternary trees. Write $\textcircled{n}$ for the weighted ternary tree that consists only of a root of weight n.

For a given ASG, we will associate a weighted ternary tree such that: (1) each leaf of the weighted ternary tree is associated with a region in the corresponding pruned ASG, (2) the underlying tree structure provides the connections between the regions, and (3) the weight of a leaf corresponds to the number of lines in the associated region.

Definition 4.13 (Stratified ASG)**.** Let $\mathcal{A}_t$ be a generalised ASG in $[0,t]$. The stratified ASG associated with $\mathcal{A}_t$ is denoted by $S(\mathcal{A}_t) \in \Upsilon_\star := \Upsilon \cup \{\Delta\}$ and defined as follows. If $\rho(\mathcal{A}_t) = \emptyset$, set $S(\mathcal{A}_t) := \textcircled{0}$. If $\rho(\mathcal{A}_t) = \Delta$, set $S(\mathcal{A}_t) := \Delta$. Assume now that $\rho(\mathcal{A}_t) \notin \{\emptyset, \Delta\}$. If there are no interactive events in $\rho(\mathcal{A}_t)$, set $S(\mathcal{A}_t) := \textcircled{n}$, where n is the number of lines present at time t in $\rho(\mathcal{A}_t)$. In the remaining case, $S(\mathcal{A}_t)$ is defined recursively as follows. Denote by $\mathcal{A}^2_{T_\star,t}$ and $\mathcal{A}^3_{T_\star,t}$ the sub-ASGs arising at the first interactive event in $\rho(\mathcal{A}_t)$ from the checking and incoming lines, respectively. Moreover, denote by $\mathcal{A}^1_t$ the generalised ASG obtained by removing $\mathcal{A}^2_{T_\star,t}$ and $\mathcal{A}^3_{T_\star,t}$ from $\rho(\mathcal{A}_t)$. Then, define $S(\mathcal{A}_t)$ by joining the roots of $S(\mathcal{A}^1_t)$, $S(\mathcal{A}^2_{T_\star,t})$ and $S(\mathcal{A}^3_{T_\star,t})$ to a new node $\hat{\rho}$, such that they play the role of the left, middle, and right child of $\hat{\rho}$, respectively.

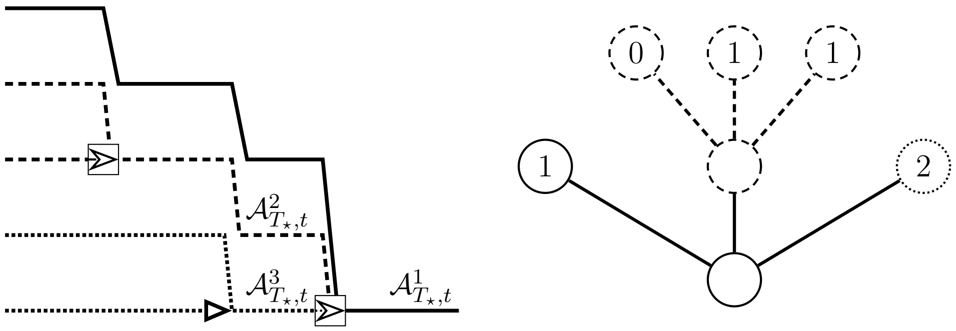


Figure 4.10. Left: A pruned ASG $\rho(\mathcal{A}_t)$. The sub-ASG $\mathcal{A}^2_{T_\star,t}$ (resp. $\mathcal{A}^3_{T_\star,t}$) arising at the first interactive event in $\rho(\mathcal{A}_t)$ from the checking (resp. incoming) line is plotted dashed (resp. dotted). The generalised ASG obtained by removing $\mathcal{A}^2_{T_\star,t}$ and $\mathcal{A}^3_{T_\star,t}$ from $\rho(\mathcal{A}_t)$ is $\mathcal{A}^1_t$ and corresponds to the solid part. In the last interactive event on $\mathcal{A}^2_{T_\star,t}$, the continuing line has disappeared due to pruning as in Fig. 4.7-(2.c). Right: The associated stratified ASG. The dashed (resp. dotted) subtree corresponds to $S(\mathcal{A}^2_{T_\star,t})$ (resp. $S(\mathcal{A}^3_{T_\star,t})$). The left subtree corresponds to $S(\mathcal{A}^1_t)$.

Figure 4.10 illustrates a pruned ASG and the recursive construction of the associated stratified ASG. Figure 4.11 illustrates how the building blocks of the pruned ASG translate to stratified ASGs.

Remark 4.16. Consider $\mathcal{A}_t$, i.e. a generalised ASG in $[0,t]$, with $\rho(\mathcal{A}_t) \notin \{\emptyset, \Delta\}$. There is a natural way to associate with any line present at time t in $\rho(\mathcal{A}_t)$ a leaf in $S(\mathcal{A}_t)$. If there are no interactive events in $\rho(\mathcal{A}_t)$, all the lines at time t in $\rho(\mathcal{A}_t)$ are associated with the root. Otherwise, using the notation in Definition 4.13, we can associate the lines of $\mathcal{A}^1_t$, $\mathcal{A}^2_{T_\star,t}$, and $\mathcal{A}^3_{T_\star,t}$ present at time t with the leaves of $S(\mathcal{A}^1_t)$, $S(\mathcal{A}^2_{T_\star,t})$, and $S(\mathcal{A}^3_{T_\star,t})$, respectively. In particular, the continuing, checking, and incoming region are associated to the subtree arising from the left, middle, and right child of the root, respectively. A recursive application of this procedure from the root to the leaves assigns each leaf in $\mathcal{A}_t$ a leaf in $S(\mathcal{A}_t)$. This construction will play an important role in Lemma 4.17. ♢

The previous remark motivates the notion of type propagation in a stratified ASGs. We say that the root of a stratified ASG $S(\mathcal{A}_t)$ is type 1 (resp. type 0) if the root of $\mathcal{A}_t$ is 1 (resp. 0). The recursive construction of $S(\mathcal{A}_t)$ implies the following type propagation rule

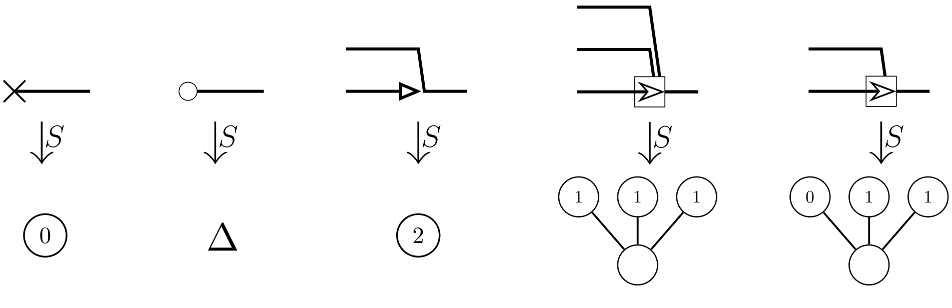


Figure 4.11. Building blocks of the pruned ASG and their associated stratified ASG.

in the stratified ASG. A leaf of the stratified ASG is of type 1 if and only if all lines in the associated region are of type 1. A vertex in the stratified ASG that is not a leaf is of type 1 if and only if the left child is of type 1 and either the middle or right child are of type 1. In the following definition, we propose a function that determines the probability of an unfit root in a stratified ASG. We subsequently prove that this function determines $H(\mathcal{A}, y_0)$ on the basis of $S(\mathcal{A})$.

Definition 4.14. Define $\mathcal{H} : \Upsilon_\star \times [0,1] \to [0,1]$ recursively. First, set for $y_0 \in [0,1]$,

$$\mathcal{H}(\textcircled{n}, y_0) := y_0^n, \quad n \in \mathbb{N}_0, \quad \text{and} \quad \mathcal{H}(\Delta, y_0) := 0$$

For $\mathcal{T} = (\tau, m_\tau) \in \Upsilon$ having at least three leaves, denote by κ^1, κ^2, and κ^3 the left, middle and right child of the root of τ, and recursively define

$$\mathcal{H}(\mathcal{T}, y_0) := \mathcal{H}(\mathcal{T}_{\kappa^1}, y_0)\left[\mathcal{H}(\mathcal{T}_{\kappa^2}, y_0) + \mathcal{H}(\mathcal{T}_{\kappa^3}, y_0) - \mathcal{H}(\mathcal{T}_{\kappa^2}, y_0)\mathcal{H}(\mathcal{T}_{\kappa^3}, y_0)\right], \tag{4.12}$$

where $\mathcal{T}_{\kappa^i}$ is the subtree of $\mathcal{T}$ that contains κ^i and all its descendants so that κ^i is the root ($i \in \{1,2,3\}$).

Remark 4.17. It follows from the recursive definition of $\mathcal{H}$ that $\mathcal{H}(\mathcal{T}, \cdot)$ is a polynomial for fixed $\mathcal{T} \in \Upsilon_\star$. ♢

The next theorem shows that the stratified ASG together with the function $\mathcal{H}$ is the right object to encode the probability that the initial lineages in an ASG are all unfit.

Theorem 4.15. *Consider $\mathcal{A}_t$, i.e. an ASG in $[0,t]$. We have*

$$H(\mathcal{A}_t, y_0) = \mathcal{H}(S(\mathcal{A}_t), y_0), \qquad \forall y_0 \in [0,1].$$

Proof. By Lemma 4.12, it is enough to prove that for $\mathcal{A}_t$, we have

$$H(\rho(\mathcal{A}_t), y_0) = \mathcal{H}(S(\mathcal{A}_t), y_0), \qquad \forall y_0 \in [0,1]. \tag{4.13}$$

If $\rho(\mathcal{A}_t) \in \{\emptyset, \Delta\}$, the result follows from the definition. Now we assume that $\rho(\mathcal{A}_t) \notin \{\emptyset, \Delta\}$, and we proceed by induction on the number m of interactive events present in $\rho(\mathcal{A}_t)$. For $m = 0$, iterating Lemma 4.7-(1), we obtain that $H(\rho(\mathcal{A}_t), y_0) = y_0^n$, where n is the number of lines present in $\rho(\mathcal{A}_t)$ at time t. In addition, by definition $S(\mathcal{A}_t) = \textcircled{n}$, and hence $\mathcal{H}(S(\mathcal{A}_t), y_0) = y_0^n$, and the result follows in this case. Now we assume that (4.13) holds for any ASG such that the corresponding pruned ASG consists of at most m interactive events. Assume that $\mathcal{A}_t$ is such that $\rho(\mathcal{A}_t)$ contains exactly $m+1$ interactive events. As in Definition 4.13, we denote by $\mathcal{A}^2_{T_\star,t}$ and $\mathcal{A}^3_{T_\star,t}$ the sub-ASGs arising at the first interactive event in $\rho(\mathcal{A}_t)$ from the checking and incoming line, respectively, and by $\mathcal{A}^1_t$ the generalised ASG obtained by removing $\mathcal{A}^2_{T_\star,t}$ and $\mathcal{A}^3_{T_\star,t}$ from $\rho(\mathcal{A}_t)$ (where $\mathcal{A}^1_t \sim \times$ in case of an interactive bifurcation). By construction, $\mathcal{A}^1_t$ and $\mathcal{A}^i_{T_\star,t}$ ($i \in \{2,3\}$) contain at most m interactive events, and therefore from the induction hypothesis $H(\rho(\mathcal{A}^1_t), y_0) = \mathcal{H}(S(\mathcal{A}^1_t), y_0)$ and $H(\rho(\mathcal{A}^i_{T_\star,t}), y_0) = \mathcal{H}(S(\mathcal{A}^i_{T_\star,t}), y_0)$ ($i \in \{2,3\}$). Hence, by Lemma 4.7 and the definition of $S(\mathcal{A}_t)$ and $\mathcal{H}$ (see Definition 4.13 and 4.14),

$$\begin{aligned}
&H(\mathcal{A}_t, y_0)\\
&= H(\mathcal{A}_t^1, y_0)\left[H(\mathcal{A}^2_{T_\star,t}, y_0) + H(\mathcal{A}^3_{T_\star,t}, y_0) - H(\mathcal{A}^2_{T_\star,t}, y_0)H(\mathcal{A}^3_{T_\star,t}, y_0)\right]\\
&= \mathcal{H}(S(\mathcal{A}_t^1), y_0)\left[\mathcal{H}(S(\mathcal{A}^2_{T_\star,t}), y_0) + \mathcal{H}(S(\mathcal{A}^3_{T_\star,t}), y_0) - \mathcal{H}(S(\mathcal{A}^2_{T_\star,t}), y_0)\mathcal{H}(S(\mathcal{A}^3_{T_\star,t}), y_0)\right]\\
&= \mathcal{H}(S(\mathcal{A}_t), y_0),
\end{aligned}$$

which proves the result. $\square$

4.3.4 Stratified ASG process

In this section we aim to describe the evolution of the process $(S(\mathcal{A}_r))_{r\geq 0}$. We first introduce some notation and operations on weighted ternary trees, which will serve as building blocks to explicitly describe the transitions of this process.

Fix $\mathcal{T} := (\tau, m_\tau) \in \Upsilon$. We denote by κ_τ and ℓ_τ^1 the root and the leftmost leaf of τ, respectively. For $v \in V_\tau \setminus \{\kappa_\tau\}$, a_v is the parent of v and $a_v^\star$ is either v if v is not the left child of its parent, or the youngest ancestor of v that is not the left child of its parent. In particular, $a^\star_{l^1_\tau} = \kappa_\tau$.

For $v \in V_\tau$, $\tau_v \in \Xi$ is the subtree of τ that contains v and all its descendants so that v is the root of τ_v. We write $\mathcal{T}_v = (\tau_v, m_{\tau_v}) \in \Upsilon$ for the ternary tree τ_v with weights given by the restriction of m_τ to L_{τ_v}, see also Fig. 4.12. Similarly, τ_v^C is the tree that arises from τ by removing all the descendants of v. The analogous weighted ternary tree is $\mathcal{T}_v^C = (\tau_v^C, m_{\tau_v^C})$, where

$$m_{\tau_v^C}(l) = \begin{cases} m_\tau(l), & \text{if } l \neq v,\\ 0, & \text{if } l = v. \end{cases}$$

For $\ell \in L_\tau$ and $\bar\tau \in \Xi$, we define $\tau \otimes_\ell \bar\tau \in \Xi$ as the tree that arises by concatenating $\bar\tau$ to τ at the leaf ℓ, see Fig. 4.13. In particular, the corresponding set of leaves is $L_{\tau\otimes_\ell\bar\tau} = (L_\tau \setminus \{\ell\}) \cup L_{\bar\tau}$. For $\bar{\mathcal{T}} = (\bar\tau, m_{\bar\tau}) \in \Upsilon$, the concatenation of $\bar{\mathcal{T}}$ at a leaf ℓ of $\mathcal{T}$ is defined

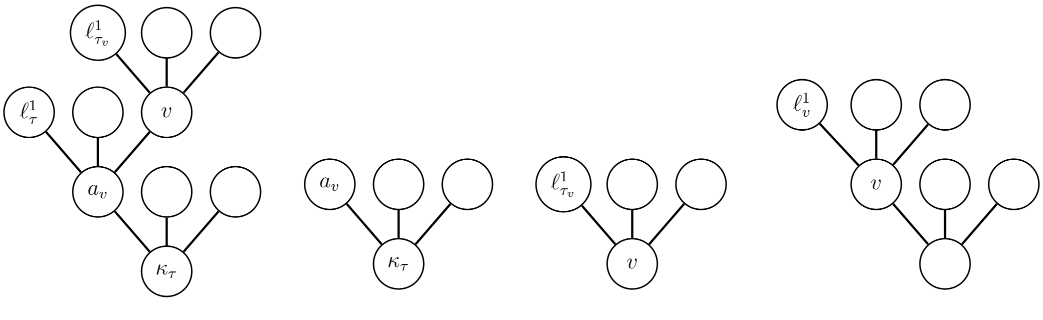


Figure 4.12. From left to right: ternary tree τ with leftmost leaf ℓ^1_τ and root κ_τ; the tree $\tau^C_{a_v}$ that arises if we remove from τ all the descendants of a_v; the restriction τ_v of τ to the subtree induced by v; the concatenation $\tau^C_{a_v} \otimes_{a_v} \tau_v$ of $\tau^C_{a_v}$ with τ_v at leaf a_v. Note that the labels do not indicate the weights here.

as $\mathcal{T} \otimes_\ell \bar{\mathcal{T}} = (\tau \otimes_\ell \bar{\tau}, m_{\mathcal{T}} \otimes_\ell m_{\bar{\mathcal{T}}})$, where

$$m_{\mathcal{T}} \otimes_\ell m_{\bar{\mathcal{T}}} = \begin{cases} m_{\mathcal{T}}(l), & \text{if } l \in L_{\mathcal{T}} \setminus \{\ell\}, \\ m_{\bar{\mathcal{T}}}(l), & \text{if } l \in L_{\bar{\mathcal{T}}} \setminus \{\ell^1_{\bar{\mathcal{T}}}\}, \\ m_{\mathcal{T}}(\ell) + m_{\bar{\mathcal{T}}}(l), & \text{if } l = \ell^1_{\bar{\mathcal{T}}}, \end{cases}$$

i.e. the weight of the left leaf of the grafted subtree is the sum of the weight of ℓ and $\ell^1_{\bar{\mathcal{T}}}$. All other leaves of the concatenated tree keep their original weight. See also Fig. 4.13 for an example. The reason for the definition is the following. The leftmost leaf of a stratified ASG corresponds to the oldest continuing region in the corresponding pruned ASG. If we attach a pruned ASG to a specific region in another pruned ASG, as we do in some pruning operations (e.g. Definition 4.8-(2.c)), the lines of the oldest region of the attached pruned ASG should be indistinguishable from the lines of the region at which we attach the pruned ASG.

The total weight of $\mathcal{T}$ is

$$M(\mathcal{T}) := \sum_{v \in L_{\mathcal{T}}} m_{\mathcal{T}}(v). \tag{4.14}$$

A generalised ASG in which the only event is an interactive bifurcation corresponds to a weighted ternary tree with three leaves, where the left leaf has weight 0 and middle and right leaf have weight 1. We denote the ternary tree with three leaves by $\Psi \in \Xi$. Moreover, set $\Psi^\star = (\Psi, m_\Psi) \in \Upsilon$ as the ternary tree Ψ with left leaf of weight 0 and middle and right leaf of weight 1, corresponding to the just-described generalised ASG.

Let us sketch, how an event in an ASG affects its stratified ASG. This leads to the appropriate transition operations in Definition 4.16.

(1) A selection event increases the number of lines in its region by one. In the stratified ASG, the weight of the corresponding leaf increases by one.

(2) In an interactive event, two new regions arise: the incoming and the checking region. In the stratified ASG, we add three new vertices with label left, middle, and right and connect them to the leaf associated with the region of the interactive event. The

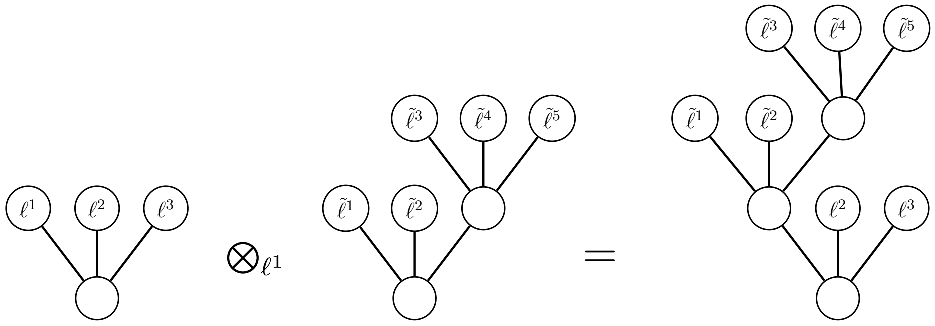


Figure 4.13. Concatenation of two trees at a given leaf. Note that the labels do not indicate weights here. The weight of leaf $\tilde{\ell}^1$ in the right tree is the weight of ℓ^1 plus the weight of $\tilde{\ell}^1$ in the middle tree.

former leaf is now an inner vertex and the three newly added vertices are leaves. The left leaf takes the weight of the parent (which is the former leaf), and middle and right leaf obtain weight 1, because their associated regions initially contain only one line.

(3) For a deleterious mutation on a line associated to leaf ℓ of the stratified ASG, we need to distinguish the following cases.
   - If the mutation occurs in a region with more than one line, then the removal of the line gives the corresponding pruned ASG (Fig. 4.6-(1.a)). In the stratified ASG this leads to a weight decrease (by one) of the associated leaf. Similarly, if the mutation occurs in a continuing region containing only a single line, then this leads to the removal of that line. The trifurcation transforms into an interactive bifurcation (Fig. 4.6-(1.c)). In the corresponding stratified ASG, the associated leaf is a left leaf with weight 1. The mutation decreases its weight to 0.
   - If the mutation occurs in a checking (resp. incoming) region with a single line, then this leads to the removal of the associated incoming (resp. checking) region (Fig. 4.6-(1.b)). In the stratified ASG, this corresponds to the removal of the associated middle and right leaf, and the replacement of the subtree arising from the parent by the subtree arising from the left child. In case the interactive event was a bifurcation (Fig. 4.6-(1.d)), it is possible that after this transformation another region with just a single incoming (or checking) line ends in a deleterious mutation. Then, we have to resolve another deleterious mutation, which can lead to the pruning of more regions. We have to resolve the mutation for the last time if it is in a region that is either a continuing region or at which the continuing region at the last trifurcation has more than one line. Denote by b_ℓ the youngest ancestor of a leaf ℓ in the stratified ASG that is either the left child of its parent, or that has left child b_ℓ^1 such that $M(\tau_{b_\ell^1}) > 0$, or which is the root. Then b_ℓ is associated to the region at which the effect of a deleterious mutation is only local and we need to replace the subtree arising from b_ℓ by its left child.

(4) For a beneficial mutation on a line associated to leaf ℓ of the stratified ASG, we need to distinguish the following cases.
   - If the mutation occurs in a checking (resp. incoming) region, the corresponding incoming (resp. checking) region is connected to the continuing region to obtain the pruned ASG (Fig. 4.7-(1.b)). In the stratified ASG, such a region corresponds to a leaf with label middle (resp. right). We assign the combined weight of the left and right leaf (resp. left and middle) to the parent, and then remove the three leaves (turning the parent into a leaf). If the mutations occurs in a continuing region (Fig. 4.7-(1.c)), we need to identify first the region to which the line is incoming or checking and apply the above procedure there. This regions is associated with the subtree arising from $a_{a_\ell^\star}$ in the stratified ASG.
   - If the mutation occurs in a region that is continuing to all interactive events, the pruned ASG is a single line ending in a beneficial mutation (Fig. 4.7-(1.d)). Such a region is associated to the leftmost leaf in the stratified ASG. A mutation in this region leads to the cemetery state. This idea also applies if there are only selective events in the pruned ASG (Fig. 4.7-(1.a)).

This leads to the following transformations that play the role of transitions of the process

$(S(\mathcal{A}_r))_{r\geq 0}$ (see Figs. 4.14 and 4.15). After their definition, we prove that these are indeed the appropriate operations (see Lemma 4.17).

Definition 4.16 (Transformations of weighted ternary trees)**.** For $\mathcal{T} = (\tau, m_\tau) \in \Upsilon$ and $\ell \in L_\tau$, define $\mathcal{T}^\ell_{\mathsf{Y}}, \mathcal{T}^\ell_{\Psi}, \mathcal{T}^\ell_{\times}, \mathcal{T}^\ell_{\circ} \in \Upsilon_\star$ as follows.

(1) $\mathcal{T}^\ell_{\mathsf{Y}} := (\tau^\ell_{\mathsf{Y}}, m_{\tau^\ell_{\mathsf{Y}}})$ with $\tau^\ell_{\mathsf{Y}} := \tau$, $m_{\tau^\ell_{\mathsf{Y}}}(\ell) := m_\tau(\ell) + 1$ and $m_{\tau^\ell_{\mathsf{Y}}}(l) := m_\tau(l)$ for $l \neq \ell$.

(2) $\mathcal{T}^\ell_{\Psi} := (\tau^\ell_{\Psi}, m_{\tau^\ell_{\Psi}})$ with $\tau^\ell_{\Psi} := \tau \otimes_\ell \Psi$ and $m_{\tau^\ell_{\Psi}} := m_\tau \otimes_\ell m_\Psi$ (i.e. $\mathcal{T}^\ell_{\Psi} = \mathcal{T} \otimes_\ell \Psi^\star$).

(3) $\mathcal{T}^\ell_{\times} := (\tau^\ell_{\times}, m_{\tau^\ell_{\times}})$ where

- if $m_\tau(\ell) > 1$ or if $m_\tau(\ell) = 1$ and ℓ is the left child of its parent or if ℓ is the root, then $\tau^\ell_{\times} := \tau$, $m_{\tau^\ell_{\times}}(\ell) := m_\tau(\ell) - 1$ and $m_{\tau^\ell_{\times}}(l) := m_\tau(l)$ for $l \neq \ell$.
- if $m_\tau(\ell) = 1$ and ℓ is not the left child of its parent, $\mathcal{T}^\ell_{\times} := \mathcal{T}^C_{b_\ell} \otimes_{b_\ell} \mathcal{T}_{b^1_\ell}$.

(4) $\mathcal{T}^\ell_{\circ} := (\tau^\ell_{\circ}, m_{\tau^\ell_{\circ}})$ where

- if $\ell \neq \ell^1_\tau$, denote by v and w the two children of $a_{a^\star_\ell}$ other than $a^\star_\ell$ ordered from left to right. Then set, $\mathcal{T}^\ell_{\circ} := (\mathcal{T}^C_{a_{a^\star_\ell}} \otimes_{a_{a^\star_\ell}} \mathcal{T}_v) \otimes_{\ell^1_{(\tau_v)}} \mathcal{T}_w$.
- if $\ell = \ell^1_\tau$, set $\mathcal{T}^\ell_{\circ} := \Delta$.

Lemma 4.17. *Let $\mathcal{A}_t$ be an ASG in $[0,t]$. Assume that $\rho(\mathcal{A}_{r-}) \notin \{\emptyset, \Delta\}$, $r \in (0,t]$, and that at time r a line present in $\rho(\mathcal{A}_{r-})$ is affected by an event in $\mathcal{A}_r$. If the affected line is associated to leaf $\ell \in S(\mathcal{A}_{r-})$ and the event corresponds to a*

(1) branching, then $S(\mathcal{A}_r) = (S(\mathcal{A}_{r-}))^\ell_{\mathsf{Y}}$.

(2) trifurcation, then $S(\mathcal{A}_r) = (S(\mathcal{A}_{r-}))^\ell_{\Psi}$.

(3) deleterious mutation, then $S(\mathcal{A}_r) = (S(\mathcal{A}_{r-}))^\ell_{\times}$.

(4) beneficial mutation, then $S(\mathcal{A}_r) = (S(\mathcal{A}_{r-}))^\ell_{\circ}$.

Proof. We prove the lemma for case (3). For the other cases the proof works analogously. We proceed by induction on the number k of interactive events present in $\rho(\mathcal{A}_{r-})$. For $k = 0$, we have that $S(\mathcal{A}_{r-}) =$ ⓝ, where n denotes the number of lines present at time $r-$

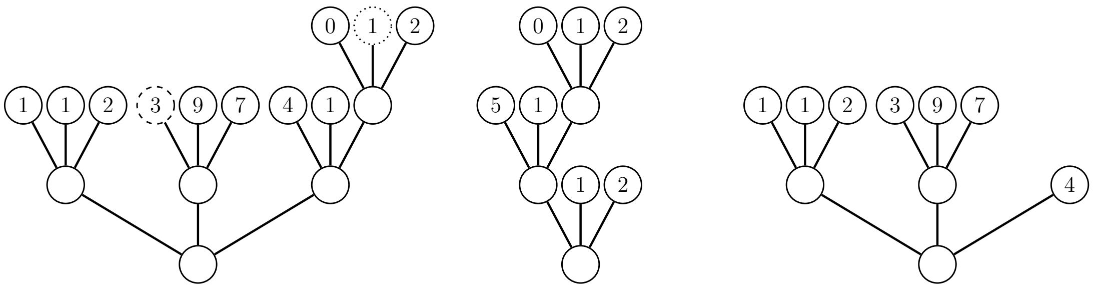


Figure 4.14. The weighted ternary tree in the middle arises if we apply the '∘'-operation at the dashed leaf in the left tree. The weighted ternary tree of the right arises if we apply the '×'-operation at the dotted leaf in the left tree. The labels indicate the weights of the leaves.

in $\rho(\mathcal{A}_{r-})$. In particular, all the lines present at time $r-$ in $\rho(\mathcal{A}_{r-})$ are associated with the root κ. Hence, $(S(\mathcal{A}_{r-}))^{\kappa}_{\times} = \textcircled{j}$, where $j := n-1$. In addition, $\rho(\mathcal{A}_r)$ does not contain interactive events and consists of j lines at time r. Therefore, $S(\mathcal{A}_r) = \textcircled{j}$, which proves the result for $k=0$. Now, we assume the result is true whenever $\rho(\mathcal{A}_{r-})$ contains at most k interactive events. Denote by $T_\star$ the time of the first interactive event in $\rho(\mathcal{A}_{r-})$. Denote by $\mathcal{A}^2_v$ and $\mathcal{A}^3_v$, for $v=r$ or $v=r-$, the ASG in $[T_\star, v]$ arising in $\mathcal{A}_v$ from the incoming and checking line at this interactive event, respectively. Moreover, denote by $\mathcal{A}^1_v$ the generalised ASG obtained by removing $\mathcal{A}^2_v$ and $\mathcal{A}^3_v$ from $\mathcal{A}_v$.

Let us assume that the line affected by the event is in $\mathcal{A}^2_{r-}$ (the other cases follow in a similar way). If $\rho(\mathcal{A}^2_r) = \emptyset$, then $\rho(\mathcal{A}_r) = \rho(\mathcal{A}^1_r)$. In particular, $S(\mathcal{A}_r) = S(\mathcal{A}^1_r) = S(\mathcal{A}^1_{r-})$. In addition, by the induction hypothesis $S(\mathcal{A}^2_{r-})^{\ell}_{\times} = S(\mathcal{A}^2_r) = \textcircled{0}$. This implies that the mass of ℓ at time $r-$ is 1. In addition, in $S(\mathcal{A}_{r-})$ we have that b_ℓ is the root. Hence, $S(\mathcal{A}_{r-})^{\ell}_{\times} = S(\mathcal{A}^1_{r-}) = S(\mathcal{A}_r)$. It remains to prove the result in the case where $\rho(\mathcal{A}^2_r) \neq \emptyset$. In this case, $S(\mathcal{A}_v)$, for $v=r$ or $v=r-$, is the weighted ternary tree obtained by joining from left to right $S(\mathcal{A}^1_v)$, $S(\mathcal{A}^2_v)$, and $S(\mathcal{A}^3_v)$ to a new root κ. By construction $S(\mathcal{A}^i_r) = S(\mathcal{A}^i_{r-})$ for $i \in \{1,3\}$, and by the induction hypothesis, $S(\mathcal{A}^2_r) = S(\mathcal{A}^2_{r-})^{\ell}_{\times}$. The result follows by noting that $S(\mathcal{A}_{r-})^{\ell}_{\times}$ is obtained by replacing $S(\mathcal{A}^2_{r-})$ by $S(\mathcal{A}^2_{r-})^{\ell}_{\times}$.

□

The effect of an event in an ASG to the associated stratified ASG is captured in the previous Lemma. Before defining the corresponding stochastic process, we discuss the appropriate topology in the following remark.

Remark 4.18 (Topology on weighted ternary trees)**.** We consider the discrete metric on $\Upsilon_\star$, the space of weighted ternary trees, which makes the topology induced by this metric discrete. In particular, $\Upsilon_\star$ equipped with this topology is then a Polish space, which makes it a suitable state space for stochastic processes. ♢

Definition 4.18 (Stratified ASG process)**.** The *stratified ASG process* $\mathscr{T} = (\mathscr{T}(r))_{r\geq 0}$ is the continuous-time Markov chain with values in $\Upsilon_\star$ and transition rates

$$\begin{aligned} q_{\mathscr{T}}(\mathcal{T}, \mathcal{T}^{\ell}_{\Upsilon}) &:= s\, m_\tau(\ell), & q_{\mathscr{T}}(\mathcal{T}, \mathcal{T}^{\ell}_{\Psi}) &:= \gamma\, m_\tau(\ell), \\ q_{\mathscr{T}}(\mathcal{T}, \mathcal{T}^{\ell}_{\times}) &:= u\nu_1\, m_\tau(\ell), & q_{\mathscr{T}}(\mathcal{T}, \mathcal{T}^{\ell}_{\circ}) &:= u\nu_0\, m_\tau(\ell), \end{aligned}$$

for $\mathcal{T} = (\tau, m_\tau) \in \Upsilon$ and $\ell \in L_\tau$. The states $\textcircled{$0$}$ and Δ are absorbing.

The infinitesimal generator of $\mathscr{T}$ is then given by

$$\mathcal{G}_{\mathscr{T}} f(\mathcal{T}) = \mathcal{G}_{\Upsilon} f(\mathcal{T}) + \mathcal{G}_{\Psi} f(\mathcal{T}) + \mathcal{G}_{\times} f(\mathcal{T}) + \mathcal{G}_{\circ} f(\mathcal{T}), \tag{4.15}$$

where $f : \Upsilon_\star \to \mathbb{R}$ bounded and for $\mathcal{T} = (\tau, m_\tau)$

$$\begin{aligned} \mathcal{G}_{\Upsilon} f(\mathcal{T}) &:= \sum_{\ell \in L_\tau} s\, m_\tau(\ell) [f(\mathcal{T}^{\ell}_{\Upsilon}) - f(\mathcal{T})], & \mathcal{G}_{\Psi} f(\mathcal{T}) &:= \sum_{\ell \in L_\tau} \gamma\, m_\tau(\ell) [f(\mathcal{T}^{\ell}_{\Psi}) - f(\mathcal{T})], \\ \mathcal{G}_{\times} f(\mathcal{T}) &:= \sum_{\ell \in L_\tau} u\nu_1\, m_\tau(\ell) [f(\mathcal{T}^{\ell}_{\times}) - f(\mathcal{T})], & \mathcal{G}_{\circ} f(\mathcal{T}) &:= \sum_{\ell \in L_\tau} u\nu_0\, m_\tau(\ell) [f(\mathcal{T}^{\ell}_{\circ}) - f(\mathcal{T})]. \end{aligned} \tag{4.16}$$

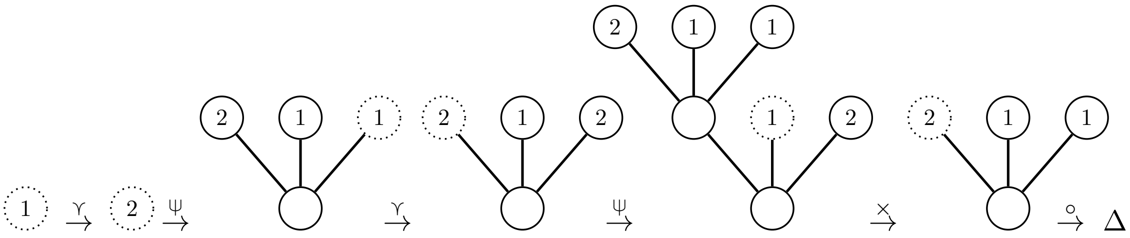


Figure 4.15. Stratified ASG process. The dotted leaf is the leaf affected by the operation associated with the subsequent transition arrow. Here, the labels in the leaves indicate their respective weight.

The following result implies that the stratified ASG process is, in distribution, equal to the stratified ASG associated with an ASG.

Theorem 4.19. *If $S(\mathcal{A}_0) = \mathscr{T}(0)$, then we have*

$$(S(\mathcal{A}_r))_{r\geq 0} \overset{(d)}{=} (\mathscr{T}(r))_{r\geq 0}.$$

Proof. Note that by construction $S(\mathcal{A}_r)$ is only affected by the events happening to the lines in $\rho(\mathcal{A}_r)$ at time r. Since the number of lines in $\rho(\mathcal{A}_r)$ that are associated with a given leaf $\ell \in S(\mathcal{A}_r) =: (\tau_r, m_r)$ is $m_r(\ell)$, the result follows as a direct application of Lemma 4.17. □

4.4 Type distribution via stratified ASG

In this section we aim to connect the solution of the deterministic mutation-selection equation with pairwise interaction (4.1) with the stratified ASG, both for a finite time horizon and at stationarity. The formal relation will be given as a duality with respect to the function $\mathcal{H}$ of Definition 4.14. After establishing the duality, we derive the long-term behaviour of the stratified ASG. In the end, we recover the bifurcation structure from (4.1) by genealogical means.

In a first step, we study the effect of the generator of $\mathscr{T}$ on $\mathcal{H}$ for a fixed initial frequency of unfit types $y_0 \in [0,1]$.

Lemma 4.20. *For every $\mathcal{T} \in \Upsilon_\star$, we have $\mathcal{H}(\mathcal{T},\cdot) \in \mathcal{C}^1([0,1],\mathbb{R})$. Moreover, for any $y_0 \in [0,1]$,*

$$\begin{aligned} \mathcal{G}_\Upsilon \mathcal{H}(\cdot,y_0)(\mathcal{T}) &= -sy_0(1-y_0)\frac{\partial \mathcal{H}(\mathcal{T},y)}{\partial y}(y_0), \quad \mathcal{G}_\Psi \mathcal{H}(\cdot,y_0)(\mathcal{T}) = -\gamma y_0(1-y_0)^2\frac{\partial \mathcal{H}(\mathcal{T},y)}{\partial y}(y_0),\\ \mathcal{G}_\times \mathcal{H}(\cdot,y_0)(\mathcal{T}) &= (1-y_0)u\nu_1\frac{\partial \mathcal{H}(\mathcal{T},y)}{\partial y}(y_0), \qquad \mathcal{G}_\circ \mathcal{H}(\cdot,y_0)(\mathcal{T}) = y_0u\nu_0\frac{\partial \mathcal{H}(\mathcal{T},y)}{\partial y}(y_0). \end{aligned} \tag{4.17}$$

We provide the proof of Lemma 4.20 in Section 4.6.

The next result establishes the aforementioned duality between the solution of the mutation-selection equation with pairwise interaction and the stratified ASG process.

Theorem 4.21 (Duality). *The stratified ASG process* $(\mathscr{T}(t))_{t\geq 0}$ *and the solution* $(y(t;y_0))_{t\geq 0}$ *of the IVP* (4.1) *satisfy the duality relation for* $t \geq 0$

$$\mathcal{H}(\mathcal{T}, y(t;y_0)) = \mathbb{E}_{\mathcal{T}}[\mathcal{H}(\mathscr{T}(t), y_0)] \quad \textit{for} \quad y_0 \in [0,1],\ \mathcal{T} \in \Upsilon_\star. \tag{4.18}$$

In particular, for $t \geq 0$ *and* $y_0 \in [0,1]$,

$$y(t;y_0) = \mathbb{E}_{①}[\mathcal{H}(\mathscr{T}(t), y_0)]. \tag{4.19}$$

Remark 4.19. If $s = \nu_0 = 0$ and $u = 1$, our theorem resembles [Mac17, Prop. I.2.1.4]. But, as mentioned in the beginning of Section 4.3, the two processes dual to the ODE are conceptually different. ♢

Proof of Theorem 4.21. We consider $y := (y(t;y_0))_{t\geq 0}$ as a (deterministic) Markov process on $[0,1]$ with generator given by

$$\mathcal{G}_y g(\tilde{y}) = F(\tilde{y}) \frac{dg}{dy}(\tilde{y}) \tag{4.20}$$

for $g \in \mathcal{C}^1([0,1],\mathbb{R})$. Fix $\mathcal{T} \in \Upsilon_\star$ and $t \geq 0$. By Lemma 4.20, $\mathcal{H}(\mathcal{T},\cdot) \in \mathcal{C}^1([0,1],\mathbb{R})$. Since F is continuously differentiable, it follows from a classic result of ODE theory [KP04, Thm. 8.43] that $y(t;\cdot) \in \mathcal{C}^1([0,1],\mathbb{R})$. Hence, also $P_t^y \mathcal{H}(\mathcal{T},\cdot) = \mathcal{H}(\mathcal{T}, y(t;\cdot)) \in \mathcal{C}^1([0,1],\mathbb{R})$, where $(P_t^y)_{t\geq 0}$ is the transition semigroup corresponding to y. $\Upsilon_\star$ is countable and equipped with the discrete topology. The number of possible transitions of the stratified ASG process at any given state is finite and each transition occurs at a finite rate. Therefore, the domain of its generator contains any bounded function from $\Upsilon_\star$ to $\mathbb{R}$. In particular, for $y_0 \in [0,1]$, $\mathcal{H}(\cdot, y_0)$ and $P_t^{\mathscr{T}} \mathcal{H}(\cdot, y_0)$ lie in the domain of its generator, where $(P_t^{\mathscr{T}})_{t\geq 0}$ is the transition semigroup corresponding to $\mathscr{T}$. Using Lemma 4.20, we deduce that

$$\mathcal{G}_{\mathscr{T}} \mathcal{H}(\cdot, \tilde{y})(\mathcal{T}) = \mathcal{G}_F \mathcal{H}(\mathcal{T}, \cdot)(\tilde{y}) \quad \text{for } \mathcal{T} \in \Upsilon_\star \text{ and } \tilde{y} \in [0,1].$$

Since $\mathcal{H}$ is bounded and continuous, the result follows from [JK14, Prop. 1.2]. □

The duality between the stratified ASG process and the solution of the mutation-selection equation implies the duality sketched in (4.10) for the ASG.

Corollary 4.22. *Let* $\mathcal{A}_t$ *be an ASG in* $[0,t]$ *starting with a single line and* $y_0 \in [0,1]$. *Then,*

$$y(t;y_0) = \mathbb{E}[H(\mathcal{A}_t, y_0)].$$

Proof. The result follows from the duality in Theorem 4.21 and then applying Theorem 4.19 and Theorem 4.15. □

We now deduce the main properties of the process $(\mathcal{H}(\mathscr{T}(r), y_0))_{r\geq 0}$. Denote by $\mathcal{F}_r^{\mathscr{T}} := \sigma(\mathscr{T}(t) : t \leq r)$, i.e. the natural filtration of $(\mathscr{T}(r))_{r\geq 0}$.

Theorem 4.23. *Let* $y_0 \in [0,1]$.

- *If* $F(y_0) > 0$, *then* $(\mathcal{H}(\mathscr{T}(r), y_0))_{r\geq 0}$ *is a bounded* $((\mathcal{F}_r^{\mathscr{T}})_{r\geq 0})$*-submartingale.*
- *If* $F(y_0) = 0$, *then* $(\mathcal{H}(\mathscr{T}(r), y_0))_{r\geq 0}$ *is a bounded* $((\mathcal{F}_r^{\mathscr{T}})_{r\geq 0})$*-martingale.*

- *If* $F(y_0) < 0$*, then* $(\mathcal{H}(\mathscr{T}(r), y_0))_{r\geq 0}$ *is a bounded* $((\mathcal{F}_r^{\mathscr{T}})_{r\geq 0})$*-supermartingale.*

In particular, $\mathcal{H}_\infty(y_0) := \lim_{r\to\infty} \mathcal{H}(\mathscr{T}(r), y_0) \in [0,1]$ *exists almost surely.*

Remark 4.20. This generalises Corollary 3.3 to the smsssi-limit. ♢

Proof of Theorem 4.23. Fix $r \geq 0$ and $y_0 \in [0,1]$. Since $\mathcal{H}(\cdot, y_0) \in [0,1]$, it is clearly integrable. Furthermore, $\mathcal{H}(\mathscr{T}(r), y_0)$ is measurable with respect to the σ-algebra generated by $\mathscr{T}(r)$ and therefore also adapted with respect to $\mathcal{F}_r^{\mathscr{T}}$. Fix $t \in [0, r]$ and $\mathcal{T} \in \Upsilon_\star$. By the Markov property and the duality, we have

$$\mathbb{E}_{\mathcal{T}}[\mathcal{H}(\mathscr{T}(r), y_0) \mid \mathcal{F}_t^{\mathscr{T}}] = \mathbb{E}_{\mathscr{T}(t)}[\mathcal{H}(\mathscr{T}(r-t), y_0)] = \mathcal{H}(\mathscr{T}(t), y(r-t; y_0)).$$

Note that if $F(y_0) > 0$ (resp. $F(y_0) < 0$), then $y(t; y_0)$ is non-decreasing (resp. non-increasing) for all $t > 0$. If $F(y_0) = 0$, then $y(t; y_0) \equiv y_0$ for all $t > 0$. Since $\mathcal{H}$ is monotone in the second argument,

$$\mathcal{H}(\mathscr{T}(r), y(t-r; y_0)) \begin{cases} \geq \mathcal{H}(\mathscr{T}(r), y_0), & \text{if } F(y_0) > 0, \\ = \mathcal{H}(\mathscr{T}(r), y_0), & \text{if } F(y_0) = 0, \\ \leq \mathcal{H}(\mathscr{T}(r), y_0), & \text{if } F(y_0) < 0. \end{cases}$$

In particular, $(\mathcal{H}(\mathscr{T}(r), y_0))_{r\geq 0}$ is a non-negative bounded sub/super-martingale. Hence, a straightforward application of Doob's martingale convergence theorem yields the last result. □

Eventually, we want to recover the long-term behaviour of the forward process in terms of the backward process by taking $t \to \infty$ in (4.19). To describe the long-term behaviour, we need some more notation. Denote the time of absorption of the stratified ASG process in ⓪ and Δ by

$$T_{⓪} := \inf\{r > 0 : \mathscr{T}(r) = ⓪\} \quad \text{and} \quad T_\Delta := \inf\{r > 0 : \mathscr{T}(r) = \Delta\},$$

respectively. Denote by

$$T_{\text{abs}} := \min\{T_{⓪}, T_\Delta\}.$$

Note that $\mathcal{H}_\infty(y_0) = 1$ on $\{T_{⓪} < \infty\}$ for any $y_0 \in [0,1]$. Similarly, $\mathcal{H}_\infty(y_0) = 0$ on $\{T_\Delta < \infty\}$ for any $y_0 \in [0,1]$. In particular, conditional on $\{T_{\text{abs}} < \infty\}$, $\mathcal{H}_\infty(y_0)$ is a Bernoulli random variable. If $\mathscr{T}$ is not absorbed, the analysis of $\mathcal{H}_\infty(y_0)$ is more involved.

Proposition 4.24. *If* y_0 *is not an unstable equilibrium of* (4.1) *and* $\mathscr{T}(0) = ①$*, then* $\mathcal{H}_\infty(y_0)$ *is a Bernoulli random variable with parameter* $y_\infty(y_0)$*.*

Remark 4.21. Let us make a link to [MSS18b] and endogeny. A very simplified description of a recursive tree process is as follows. Consider a tree and assign types to its leaves. The types propagate through the tree as follows. At each branching in the tree (this includes degenerate branchings with one branch), there is a random function attached to it. Given the types of the descendents, the random function determines the type of the ancestor. If the state at the root of this tree is measurable with respect to the σ-algebra generated by the random functions attached to all branching points, then the recursive tree process is said to be endogenous (see [AB05; MSS18a] for more details). It follows from [MSS18b, Prop. 15] together with [MSS18a, Thm. 5] (alternatively, [MSS18b, Prop. 16]) that if y_0 is an equilibrium of (4.1), then $\mathcal{H}_\infty(y_0)$ is Bernoulli if and only if the recursive tree process corresponding to y_0 is endogenous. In particular, Proposition 4.24 implies that in our setup, the recursive tree processes corresponding to stable equilibria are always endogenous. An alternative way to recover Proposition 4.24 is via [MSS18b, Prop. 19, see also Sect. 2.1] (alternatively, [AB05, Lem. 15]).

For $s = \nu_0 = 0$ and $u = 1$, Proposition 4.24 can be also recovered from [MSS18b, Thm. 17]. Moreover, if y_0 is an unstable equilibrium, Mach et al. [MSS18b, Lem. 18] determine the first and second moments of $\mathcal{H}_\infty(y_0)$. This allows them to infer that $\mathcal{H}_\infty(y_0)$ is then not Bernoulli. Furthermore, they complement the result by numerical evaluations of the distribution function [MSS18b, Fig. 2]. $\diamondsuit$

Proof of Proposition 4.24. Fix $y_0 \in [0,1]$ such that it is not an unstable equilibrium of (4.1). In particular, $y_\infty(y_0)$ is then attracting (from at least one side) and, therefore, $F'(y_\infty(y_0)) \leq 0$. Consider $(\mathscr{T}(r))_{r\geq 0}$ with $\mathscr{T}(0) =$ ①. Denote by $T_\Upsilon, T_\Psi, T_\circ$, and $T_\times$ the time of the first selective, interactive, beneficial mutation, and deleterious mutation event, respectively. Let $T = \min\{T_\Upsilon, T_\Psi, T_\circ, T_\times\}$. For sufficiently large $r > 0$,

$$\begin{aligned}\mathcal{H}(\mathscr{T}(r), y_0) = \mathbb{1}_{\{T=T_\times\}} &+ \mathbb{1}_{\{T=T_\Upsilon\}}\mathcal{H}(\mathscr{T}^1(r-T), y_0)\mathcal{H}(\mathscr{T}^2(r-T), y_0)\\ &+ \mathbb{1}_{\{T=T_\Psi\}}\mathcal{H}(\mathscr{T}^3(r-T), y_0)\big[\mathcal{H}(\mathscr{T}^4(r-T), y_0) + \mathcal{H}(\mathscr{T}^5(r-T), y_0)\\ &- \mathcal{H}(\mathscr{T}^4(r-T), y_0)\mathcal{H}(\mathscr{T}^5(r-T), y_0)\big],\end{aligned}$$

where $\mathscr{T}^1, \mathscr{T}^2, \mathscr{T}^3, \mathscr{T}^4$, and $\mathscr{T}^5$ are independent stratified ASGs all of which are started in ①. Taking the limit $r \to \infty$ yields,

$$\begin{aligned}\mathcal{H}_\infty(y_0) = \mathbb{1}_{\{T=T_\times\}} &+ \mathbb{1}_{\{T=T_\Upsilon\}}\mathcal{H}^1_\infty(y_0)\mathcal{H}^2_\infty(y_0)\\ &+ \mathbb{1}_{\{T=T_\Psi\}}\mathcal{H}^3_\infty(y_0)[\mathcal{H}^4_\infty(y_0) + \mathcal{H}^5_\infty(y_0) - \mathcal{H}^4_\infty(y_0)\mathcal{H}^5_\infty(y_0)].\end{aligned}$$

In particular, for every $G \in \mathcal{C}([0,1],\mathbb{R})$, we have

$$\begin{aligned}\mathbb{E}[G(\mathcal{H}_\infty(y_0))] = \mathbb{P}(T = T_\times)\mathbb{E}\big[G(1)\big] &+ \mathbb{P}(T = T_\circ)\mathbb{E}\big[G(0)\big]\\ &+ \mathbb{P}(T = T_\Upsilon)\mathbb{E}\Big[G\big(\mathcal{H}^1_\infty(y_0)\mathcal{H}^2_\infty(y_0)\big)\Big]\\ &+ \mathbb{P}(T = T_\Psi)\mathbb{E}\Big[G\big(\mathcal{H}^3_\infty(y_0)\big[\mathcal{H}^4_\infty(y_0) + \mathcal{H}^5_\infty(y_0) - \mathcal{H}^4_\infty(y_0)\mathcal{H}^5_\infty(y_0)\big]\big)\Big].\end{aligned} \tag{4.21}$$

Set $E(y_0) := \mathbb{E}[\mathcal{H}_\infty(y_0)]$ and $V(y_0) := \mathbb{E}[\mathcal{H}_\infty(y_0)^2]$. Note that if X is a random variable in $[0,1]$, then X is Bernoulli if and only if $\mathbb{E}[X(1-X)] = 0$. In particular, $\mathcal{H}_\infty(y_0)$ is Bernoulli if and only if $E(y_0) = V(y_0)$. Choosing $G(x) = x(1-x)$ in (4.21) and using independence, we obtain

$$\begin{aligned}E(y_0) - V(y_0) = &\frac{s}{u+s+\gamma}(E(y_0) - V(y_0))(E(y_0) + V(y_0)) \\ &+ \frac{\gamma}{u+s+\gamma}(E(y_0) - V(y_0)) \\ &\cdot (V(y_0)^2 + V(y_0)(2 - 3E(y_0)) + E(y_0)(2 - E(y_0))),\end{aligned} \tag{4.22}$$

By the duality,

$$E(y_0) = y_\infty(y_0) \tag{4.23}$$

and hence $E(y_0)$ is an equilibrium of (4.1). Furthermore, $\mathcal{H}_\infty(y_0) \in [0,1]$ and hence

$$0 \leq V(y_0) = \mathbb{E}[\mathcal{H}_\infty(y_0)^2] \leq \mathbb{E}[\mathcal{H}_\infty(y_0)] = E(y_0).$$

We want to further narrow down the value of $V(y_0)$. To do so, we consider equation (4.22) with unknown $x = V(y_0)$, i.e. we rewrite (4.22) as

$$(E(y_0) - x)\,\hat{p}(E(y_0), x) = 0,$$

where

$$\hat{p}(E(y_0), x) := s(E(y_0) + x) + \gamma(x^2 + x(2 - 3E(y_0)) + E(y_0)(2 - E(y_0))) - (u + s + \gamma).$$

Note that

$$\hat{p}(E(y_0), 0) = s(E(y_0) - 1) + \gamma(E(y_0)(2 - E(y_0)) - 1) - u \leq -u$$

and $\hat{p}(E(y_0), E(y_0)) = F'(E(y_0)) \leq 0$, where the inequality follows from (4.23) and the observation at the beginning of the proof. In particular, since $\hat{p}(E(y_0), x)$ is a quadratic polynomial with positive quadratic term in x, $\hat{p}(E(y_0), x) \neq 0$ for all $x \in [0, E(y_0))$. Altogether, this implies $V(y_0) = E(y_0)$. □

Denote the probability of absorption in ⓪ of the stratified ASG started in ① by

$$w_1 := \mathbb{P}(T_{⓪} < \infty \mid \mathscr{T}(0) = ①).$$

Similarly, denote the probability of *not* getting absorbed in Δ of the stratified ASG process started in ① by

$$d_1 := \mathbb{P}(T_\Delta = \infty \mid \mathscr{T}(0) = ①).$$

By definition, we have $w_1 \leq d_1$. Moreover, define for $y_0 \in [0,1]$,

$$p(y_0) := \mathbb{E}[\mathbb{1}_{\{T_{⓪} = \infty\}} \mathcal{H}_\infty(y_0) \mid \mathscr{T}(0) = ①].$$

If y_0 is not an unstable equilibrium, it follows from Proposition 4.24 that

$$p(y_0) = \mathbb{P}(\mathcal{H}_\infty(y_0) = 1, T_{⓪} = \infty \mid \mathscr{T}(0) = ①).$$

If the stratified ASG process absorbs in ⓪, then the descendant is of type 1. On the other hand, if the stratified ASG process absorbs in Δ, then the descendant is of type 0. In both cases this is independent of the sampling probability y_0 at the leaves. There are parameter regions where $w_1 = d_1$, so these are then the only possibilities. If the stratified ASG does not absorb, the type depends on y_0. We will examine this in detail now. Recall from Section 4.1 that $\hat{y}_\infty$ and $\check{y}_\infty$ denote, respectively, the smallest and largest equilibrium of (4.1) in $[0, 1]$.

Theorem 4.25. *For any* $y_0 \in [0, 1]$, *we have*

$$y_\infty(y_0) = w_1 + p(y_0). \tag{4.24}$$

In particular, $\hat{y}_\infty = w_1$ *and* $\check{y}_\infty = d_1$. *Furthermore, the two following statements are equivalent:*

(i) $\mathbb{P}(T_{\text{abs}} < \infty \mid \mathscr{T}(0) = \textcircled{1}) = 1$,

(ii) $\hat{y}_\infty = w_1 = d_1 = \check{y}_\infty$ *is the unique equilibrium in* $[0, 1]$.

Moreover, both (i) and (ii) imply that $w_1 = d_1$ *is stable.*

Remark 4.22. Theorem 4.25 in combination with the facts collected at the beginning of Section 4.1 implies that w_1 is never unstable. Furthermore, if $\nu_0 > 0$, then d_1 is not unstable. $\diamondsuit$

Proof of Theorem 4.25. We decompose $\mathcal{H}(\mathscr{T}(t), y)$ according to $\{T_{\textcircled{0}} < \infty\}$ and $\{T_{\textcircled{0}} = \infty\}$. More precisely, starting from the duality,

$$y(t; y_0) = \mathbb{E}_{\textcircled{1}}[\mathcal{H}(\mathscr{T}(t), y_0)\mathbb{1}_{\{T_{\textcircled{0}} < \infty\}}] + \mathbb{E}_{\textcircled{1}}[\mathcal{H}(\mathscr{T}(t), y_0)\mathbb{1}_{\{T_{\textcircled{0}} = \infty\}}]. \tag{4.25}$$

Since $\mathcal{H}(\textcircled{0}, y_0) = 1$, the first term in (4.25) converges to w_1 as $t \to \infty$. By Theorem 4.23, the second term of (4.25) converges to $p(y_0)$ as $t \to \infty$, thus proving (4.24). Since $\mathcal{H}_\infty(0) = 0$ on $\{T_{\textcircled{0}} = \infty\}$, $\mathcal{H}_\infty(1) = 1$ on $\{T_\Delta = \infty\}$, and $\mathcal{H}_\infty(1) = 0$ on $\{T_\Delta < \infty\}$,

$$p(0) = 0 \quad \text{and} \quad p(1) = \mathbb{P}(T_{\textcircled{0}} = \infty, T_\Delta = \infty) = d_1 - w_1.$$

Since p is increasing, we deduce that $\hat{y}_\infty = w_1$ and $\check{y}_\infty = d_1$. Note that under (i), $p(y_0) = 0$ for all $y_0 \in [0, 1]$, and hence the stability of w_1 follows by taking the limit when $t \to \infty$ in (4.25). Finally, the equivalence between (i) and (ii) follows using the identity

$$\mathbb{P}(T_{\textcircled{0}} = \infty, T_\Delta = \infty) = d_1 - w_1. \qquad \square$$

We denote the domain of attraction of an equilibrium y_∞ by

$$\text{Attr}(y_\infty) := \{y_0 \in [0, 1] : \lim_{t \to \infty} y(t; y_0) = y_\infty\}$$

and refine the statement of Proposition 4.24. The following result is a corollary to Theorem 4.25.

Corollary 4.26. *Assume* $\mathscr{T}(0) = ①$. *If* $y_0 \in \mathit{Attr}(w_1)$, *then a.s.*

$$\mathcal{H}_\infty(y_0) = \begin{cases} 0, & \text{if } T_{⓪} = \infty, \\ 1, & \text{if } T_{⓪} < \infty. \end{cases}$$

If $y_0 \in \mathit{Attr}(d_1)$, *then a.s.*

$$\mathcal{H}_\infty(y_0) = \begin{cases} 0, & \text{if } T_\Delta < \infty, \\ 1, & \text{if } T_\Delta = \infty. \end{cases}$$

In particular, on $\{T_{\text{abs}} = \infty\}$,

$$\mathcal{H}_\infty(y_0) = \begin{cases} 0, & \text{if } y_0 \in \mathit{Attr}(w_1), \\ 1, & \text{if } y_0 \in \mathit{Attr}(d_1). \end{cases}$$

Remark 4.23. For $y_0 \in \{w_1, d_1\}$, Corollary 4.26 makes the endogeny of the underlying recursive tree process corresponding to y_0 explicit. ♢

Remark 4.24. $\mathrm{Attr}(w_1) = \mathrm{Attr}(d_1)$ if and only if $w_1 = d_1$. By Theorem 4.25, $w_1 = d_1$ if and only if $\mathbb{P}(T_{⓪} = \infty, T_\Delta = \infty) = 0$. ♢

Proof of Corollary 4.26. If $y_0 \in \mathrm{Attr}(w_1)$, then by definition $w_1 = y_\infty(y_0)$ and together with Theorem 4.25,

$$w_1 = y_\infty(y_0) = w_1 + p(y_0),$$

which implies that $p(y_0) = 0$. It follows that $\mathcal{H}_\infty(y_0) = 0$ if $T_{⓪} = \infty$. Moreover, we clearly have $\mathcal{H}_\infty(y_0) = 1$ on $\{T_{⓪} < \infty\}$, since $\mathcal{H}(⓪, y_0) = 1$. This proves the first part. For the second part, assume that for $y_0 \in \mathrm{Attr}(d_1)$

$$\mathbb{P}(\mathcal{H}_\infty(y_0) = 1 \mid T_\Delta = \infty) < 1.$$

In this case also $\mathbb{E}[\mathcal{H}_\infty(y_0) \mid T_\Delta = \infty] < 1$. On the other hand, by Theorem 4.25

$$d_1 = y_\infty(y_0) = w_1 + p(y_0) = \mathbb{E}[\mathbb{1}_{\{T_\Delta = \infty\}} \mathcal{H}_\infty(y_0)] = d_1 \mathbb{E}[\mathcal{H}_\infty(y_0) \mid T_\Delta = \infty] < d_1,$$

which is a contradiction. Moreover, we clearly have $\mathcal{H}_\infty(y_0) = 0$ on $\{T_\Delta < \infty\}$, since $\mathcal{H}(\Delta, y_0) = 0$. The last statement is a direct consequence of the two previous results. ☐

The following proposition implies that the set of values for y_0 such that $\mathcal{H}_\infty(y_0) \notin \{0, 1\}$ on non-absorption is at most a singleton.

Proposition 4.27. *If* $\mathbb{P}(T_{\text{abs}} = \infty \mid \mathscr{T}(0) = ①) > 0$, *we have that*

$$\begin{aligned} y_c &:= \inf\Big\{y_0 \in [0,1] : \mathbb{P}(\mathcal{H}_\infty(y_0) = 1 \mid T_{\text{abs}} = \infty,\ \mathscr{T}(0) = ①) = 1\Big\} \\ &= \sup\Big\{y_0 \in [0,1] : \mathbb{P}(\mathcal{H}_\infty(y_0) = 0 \mid T_{\text{abs}} = \infty,\ \mathscr{T}(0) = ①) = 1\Big\}. \end{aligned}$$

Furthermore, $y_c \in [w_1, d_1]$.

Proof. Define $y^{(1)} := \inf\{y_0 \in [0,1] : \mathbb{P}(\mathcal{H}_\infty(y_0) = 1 \mid T_{\text{abs}} = \infty, \mathscr{T}(0) = \textcircled{1}) = 1\}$ and $y^{(0)} := \sup\{y_0 \in [0,1] : \mathbb{P}(\mathcal{H}_\infty(y_0) = 0 \mid T_{\text{abs}} = \infty, \mathscr{T}(0) = \textcircled{1}) = 1\}$. By the properties of sup and inf, $y^{(0)} \leq y^{(1)}$. By Theorem 4.25, $w_1 \leq d_1$ are equilibria and $y_\infty(y^{(0)}), y_\infty(y^{(1)}) \in [w_1, d_1]$. By definition of sup and inf and $p(y_0)$, we have that for all $\varepsilon > 0$ the inequalities $w_1 < y_\infty(y^{(0)} + \varepsilon)$ and $d_1 > y_\infty(y^{(1)} - \varepsilon)$ hold. Assume $y^{(0)} < y^{(1)}$. In particular, for $y_0 \in (y^{(0)}, y^{(1)})$, we have that $w_1 < y_\infty(y_0) < d_1$. Since (4.1) has at most three equilibria, for all $y_0 \in (y^{(0)}, y^{(1)})$ we have $y_\infty(y_0) \equiv c$ for some c in (w_1, d_1). But then c, which is enclosed by w_1 and d_1, is stable. This contradicts the findings from Remark 4.6. By Corollary 4.26, it follows that $y_c \in [w_1, d_1]$. $\square$

Let us study in more detail the long-term behaviour of the stratified ASG process that is conditioned to not absorb. Note that in the case $\gamma = 0$, the total weight of the stratified ASG behaves like a linear birth-death process with killing (which occurs at rate $u\nu_0$ per line). This process either absorbs or grows to ∞; see also Lemma 3.4. The next proposition sheds light on the case $\gamma > 0$. Recall from (4.14) that $M(\mathcal{T})$ is the total weight of $\mathcal{T}$.

Proposition 4.28. *On* $\{T_{\text{abs}} = \infty\}$, *we have that* $\lim_{r\to\infty} M(\mathscr{T}(r)) = \infty$.

Proof. Let $T_1, T_2, \dots$ be the jump times of $\mathscr{T}(t)$. Assume there is $A \subseteq \{\min\{T_{\textcircled{0}}, T_\Delta\} = \infty\}$ with $\mathbb{P}(A) > 0$ such that for all $\omega \in A$ there exists $R(\omega) < \infty$ and $J \subseteq \mathbb{N}$ with $|J| = \infty$ such that $M(\mathscr{T}(T_j)(\omega)) < R(\omega)$ for all $j \in J$; i.e. there exists a infinite sequence of jumping times such that the mass of the tree is below a certain threshold. Define the set of weighted ternary trees with (positive) mass smaller than n as $\Upsilon_n = \{\mathscr{T} \in \Upsilon \setminus \{\textcircled{0}\} \mid M(\mathscr{T}) < n\}$. This set is finite. Define

$$\mathscr{T}^n(t) = \begin{cases} \mathscr{T}(t), & \text{if } \mathscr{T}(t) \in \Upsilon_n, \\ \Theta, & \text{if } \mathscr{T}(t) \in \Upsilon \setminus \Upsilon_n, \end{cases}$$

where Θ is an arbitrary auxiliary state not in Υ_n. Note that, for all $y_0 \in (0,1)$,

$$\min_{\mathscr{T} \in \Upsilon_n} \mathcal{H}(\mathscr{T}, y_0) = i_n > 0 \quad \text{and} \quad \max_{\mathscr{T} \in \Upsilon_n} \mathcal{H}(\mathscr{T}, y_0) = s_n < 1.$$

Hence, for $\omega \in A, y_0 \in (0,1)$, and $j \in J$,

$$0 < i_{R(\omega)} \leq \mathcal{H}(\mathscr{T}(T_j), y_0) = \mathcal{H}(\mathscr{T}^{R(\omega)}(T_j), y_0) \leq s_{R(\omega)} < 1.$$

Taking the limit $j \to \infty$ implies that for all $\omega \in A$, $0 < i_{R(\omega)} \leq \mathcal{H}_\infty(y_0)(\omega) \leq s_{R(\omega)} < 1$. Recall from Proposition 4.24 that if y_0 is not an unstable equilibrium $\mathcal{H}_\infty(y_0) \in \{0,1\}$ almost surely so that $\mathbb{P}(A) = 0$ which contradicts the assumption and the result follows. $\square$

The following corollary is a collection of our previous results and connects the long-term behaviour of the stratified ASG process with the bifurcation structure of the mutation-selection equation.

Corollary 4.29. *(1) If* $\mathbb{P}(T_{\text{abs}} = \infty \mid \mathscr{T}(0) = \textcircled{1}) = 0$, *then* $w_1 = d_1$ *is the unique equilibrium of* (4.1) *in* $[0,1]$ *and it is stable.*

(2) If $\mathbb{P}(T_{\text{abs}} = \infty \mid \mathscr{T}(0) = ①) > 0$ *and* $y_c = d_1$*, then* w_1 *and* d_1 *are the only equilibria of* (4.1) *in* $[0,1]$ *with* $Attr(w_1) = [0, d_1)$ *and* $Attr(d_1) = [d_1, 1]$*.*

(3) If $\mathbb{P}(T_{\text{abs}} = \infty \mid \mathscr{T}(0) = ①) > 0$ *and* $y_c = w_1$*, then* w_1 *and* d_1 *are the only equilibria of* (4.1) *in* $[0,1]$ *with* $Attr(w_1) = [0, w_1]$ *and* $Attr(d_1) = (w_1, 1]$*.*

(4) If $\mathbb{P}(T_{\text{abs}} = \infty \mid \mathscr{T}(0) = ①) > 0$ *and* $w_1 < y_c < d_1$*, then* $w_1 < y_\infty(y_c) < d_1$ *are the only equilibria of* (4.1) *in* $[0,1]$ *with* w_1 *and* d_1 *being stable and* $y_\infty(y_c)$ *being unstable.*

Proof. The first claim is already part of Theorem 4.25. The second and third claims follow from Proposition 4.27 and Theorem 4.25. The last claim follows by Proposition 4.27, Theorem 4.25, and Remark 4.6. ☐

Let us discuss in detail how Corollary 4.29 relates to the forward picture. In (1) the stratified ASG absorbs almost surely so that the type of the descendant is independent of the sampling probability y_0, which results in a unique equilibrium in $[0,1]$. Its (global) stability reflects the independence of the sampling step. In (2) to (4), the stratified ASG may also grow to ∞ in which case the sampling probability y_0 becomes relevant. In (2) and (3), $y_c \in \{w_1, d_1\}$ so that $[0,1] = \text{Attr}(w_1)\dot\cup\text{Attr}(d_1)$, whereas in (4), $y_c \in (w_1, d_1)$ so that $[0,1] = \text{Attr}(w_1)\dot\cup\{y_c\}\dot\cup\text{Attr}(d_1)$. Corollary 4.26 describes the type of the descendant when the leaves in an infinite stratified ASG are sampled according to $y_0 \in \text{Attr}(w_1) \cup \text{Attr}(d_1)$: for $y_0 \in \text{Attr}(w_1)$ ($y_0 \in \text{Attr}(d_1)$) the descendant is of type 0 (of type 1) almost surely (on the set of non-absorption). For $y_0 = y_c$, the probability of the stratified ASG to not absorbs and having an unfit descendant is $p(y_c)$.

In the remainder of this section, we concentrate on the case $\nu_0 = 0$, where we can make more explicit statements.

Proposition 4.30. *If* $\nu_0 = 0$*, then* $w_1 = \min\{\bar{y}_2, 1\}$ *and* $d_1 = 1$*. In particular,* $w_1 = d_1$ *if and only if we are in one of the following parameter regimes*

$$(1)\quad \sigma < 0, \qquad (2)\quad \sigma = 0,\ u = s = \gamma, \qquad (3)\quad \sigma \geq 0, u \geq \hat{u}, \textit{and } s > \gamma,$$

where σ *is defined in* (4.7)*.*

Proof. The proposition is a direct consequence of Theorem 4.25 together with Proposition 4.2. ☐

Let us now connect the genealogical backward picture in the case $\nu_0 = 0$ with the bifurcation structure described in Section 4.1.

(i) By the analysis of the forward picture, if either $u < \hat{u}$ or ($s < \gamma$ and $u = \hat{u}$), $y_\infty(y_0) \in [0,1)$ unless $y_0 = 1$. The genealogical picture in this case is as follows. Either the stratified ASG absorbs in a state in which all leaves vanished due to deleterious mutations (probability $w_1 < 1$); or, by Proposition 4.28, the total weight of the stratified ASG tends to ∞ and, as a consequence of Corollary 4.26, the probability of an unfit root is 0.

(ii) If either $u > \check{u}$ or ($\gamma \leq s$ and $u \in [\hat{u}, \check{u}]$), then $y_\infty(y_0) = 1$ for all $y_0 \in [0,1]$. By Theorem 4.25, this corresponds to the almost sure absorption of the stratified ASG in ⓪ and the sampled individual is unfit regardless of y_0.

(iii) If $s < \gamma$ and $u = \check{u}$, both w_1 and 1 are attracting from the left, i. e. $y_\infty(y_0) = w_1$ for $y_0 \in [0, w_1]$ and $y_\infty(y_0) = 1$ for $y_0 \in (w_1, 1]$. By the same arguments as in case (i), we have the following backward picture. If the stratified ASG does not absorb, which occurs with probability $w_1 < 1$, its total weight tends to ∞. Whether or not the root of this infinite tree is unfit depends on the initial type frequency. If $y_0 \leq \bar{y}_2$, the root is fit almost surely (on non-absorption); if $y_0 > \bar{y}_2$, it is unfit almost surely (on non-absorption).

(iv) If $s < \gamma$ and $u \in (\hat{u}, \check{u})$, there exist three equilibria in $[0,1]$. By arguments similar to case (i), the backward picture is as follows. If the process does not absorb, where absorption occurs with probability $w_1 < 1$, the number of lines tends to ∞. Whether or not the root of this infinite tree is unfit depends on the initial type frequency. If $y_0 < \bar{y}_3$, the root is fit almost surely (on non-absorption); if $y_0 > \bar{y}_3$ the root is unfit almost surely (on non-absorption).

In the cases (i), (ii), and (iii), where the outcome depends on y_0, the question arises whether a more detailed analysis of the stratified ASG explains the dependence on y_0 in terms of the tree structure. This can be done in the case $u < \hat{u}$, where w_1 is the smallest solution of

$$-\gamma y^2 + (s+\gamma)y - u = 0, \tag{4.26}$$

i.e $w_1 = (1 + s/\gamma - \sqrt{\sigma})/2$; see (4.3) and (4.6). Let us rederive this expression in terms of the stratified ASG. To do so, we analyse $1 - w_1$, which is the survival probability of the stratified ASG started in ①. Define $\tilde{\gamma} := \gamma(1 - w_1)$ as well as

$$a := 1 - \frac{u}{s} \quad \text{and} \quad b = \frac{u}{s}\frac{\gamma}{s + \tilde{\gamma} - u}.$$

Note that, by a straightforward calculation, we can rewrite (4.26) (with $y = w_1$) as a quadratic equation in $1 - w_1$, namely

$$b(1-w_1)^2 - (1-w_1) + a = 0. \tag{4.27}$$

Using the solution formula and the series expansion of the square root at 1 leads to

$$1 - w_1 = \frac{1 - \sqrt{1 - 4ab}}{2b} = \frac{1}{2b}\sum_{n=1}^{\infty} \frac{(2n)!}{(2n-1)(n!)^2} a^n b^n.$$

A straightforward calculation then leads to the following result for which we subsequently provide an additional probabilistic proof. The result and its probabilistic proof shed light on the various ways, a stratified ASG may escape to ∞.

Proposition 4.31. *Let $\nu_0 = 0$. For $u < \hat{u}$, we have*

$$1 - w_1 = \sum_{n=0}^{\infty} C_n a^{n+1} b^n, \tag{4.28}$$

where C_n is the nth Catalan number.

Probabilistic proof. Note that the number of lines inside a continuing region (i. e. only the lines that are connected to the initial line of the region by selective arrows) behaves like a simple birth-death process with birth rate s and death rate u. Recall that this process dies out with probability u/s and grows to ∞ with probability $1-u/s$ (note that we consider here the parameter regime $u<\hat{u}=s$, so $u/s\in[0,1)$). Hence, a is the probability that the leftmost leaf in the stratified ASG does not reach weight 0. On the other hand, (4.26) is equivalent to

$$y=\frac{u}{s}-\frac{u}{s}\frac{\gamma(1-y)}{s+\gamma(1-y)}.$$

Using the fact that w_1 solves the above equation, a straightforward calculation yields the survival probability of the entire tree as

$$1-w_1=\frac{s+\tilde{\gamma}-u}{s+\tilde{\gamma}}. \tag{4.29}$$

In what follows, we work on the basis of an ASG in the time interval $[0,\infty)$ that is started with one line. Here, we refer to a region either as the entire ASG or as the ASG arising from the incoming or checking line in a trifurcation. We call a trifurcation in an ASG *successful* if both the associated checking and incoming lines give rise to ASGs that have non-absorbing stratified ASGs (note that, if one of them absorbs, both lines will be pruned in the long run). We say that a region survives without the help of trifurcations if the leftmost leaf of its associated stratified ASG does not reach weight 0; otherwise we say that the region goes extinct without trifurcations. On the other hand, we say that a region survives due to trifurcations if its stratified ASG does not absorb but its leftmost leaf reaches weight 0. Denote by π the probability that, conditional on the ASG going extinct without the help of trifurcations, the ASG goes extinct. We claim that $\pi=s/(s+\tilde{\gamma})$. Indeed, conditional on the event that the ASG goes extinct without trifurcations, the weight of the leftmost leaf evolves as a birth-death process with birth rate u and death rate s; see Lemma 4.45. Hence, a first-step decomposition of π leads to

$$(\gamma(1-w_1)^2+s+u)\pi=s+u\pi^2.$$

In the parameter regime we consider here, the solution is unique in $[0,1]$ and is given by $s/(s+\tilde{\gamma})$, which proves the claim. In particular, the probability that the ASG survives due to trifurcations is

$$\frac{u}{s}(1-\pi)=\frac{u}{s}\frac{\gamma(1-w_1)^2}{s+\tilde{\gamma}-u}=b(1-w_1)^2,$$

where we used (4.29). The probability that the incoming and checking lines of a successful trifurcation both survive without the help of trifurcations is $a^2/(1-w_1)^2$. Hence, a^2b is the probability that the ASG survives due to trifurcations and the incoming and the checking line of the first successful trifurcation survive without the help of trifurcations.

We now partition the event of non-absorption of the stratified ASG according to the possible ways the stratified ASG may survive. We do this by associating to every such stratified

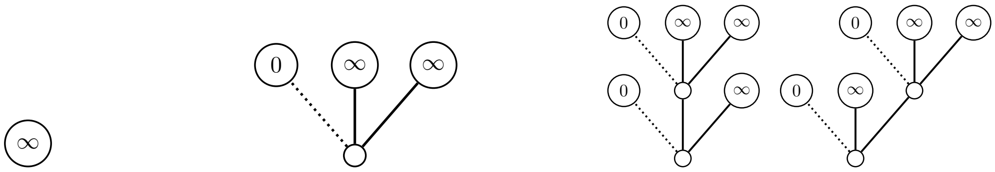


Figure 4.16. Ways of the stratified ASG process to survive (and then grow to infinite size) if $\nu_0 = 0$. The label ∞ in a leaf means that the region survives without the help of trifurcations. For an ASG that survives without the help of trifurcations, there is $C_0 = 1$ binary tree (left). For $n = 1$ there is $C_1 = 1$ binary tree (middle). For $n = 2$ there are $C_2 = 2$ binary trees (right). The solid lines correspond to the binary trees.

ASG a binary tree $\mathcal{B}$. We start with the binary tree with one leaf β. If the ASG survives without the help of trifurcations, we set $\mathcal{B} = \beta$. If not, we add two leaves β_1 and β_2 that are associated with the incoming and checking region in the first successful trifurcation, respectively. If one of these regions survives without the help of trifurcations, the corresponding leaf stops branching. Otherwise, we attach to it two new leaves and continue this procedure in a recursive way. By construction, $\mathcal{B}$ is a binary Galton-Watson tree with offspring distribution $p_2 = a/(1 - w_1)$ and $p_0 = 1 - p_2$. A straightforward calculation, which uses (4.27), shows that $p_0 > p_2$ and hence $\mathcal{B}$ is almost surely finite (on the event of non-absorption). Let N be the random number of bifurcations. In the previous construction a branching corresponds to a successful trifurcation in a region of the reduced ASG that goes extinct without trifurcations and a leaf corresponds to a region that survives without the help of trifurcations; see Fig. 4.16. By a simple induction argument on the number of bifurcations n in the binary tree, one can show that the probability that the ASG corresponds to a given binary tree with $n+1$ leaves is $a^{n+1}b^n$. Indeed, we have already proved that for $n = 0$ this probability is a and for $n = 1$ it is $a^2 b$. By construction, computing the probability of a tree with $n + 1$ bifurcations amounts to replacing a leaf by a tree with one bifurcation, which entails replacing a factor a by a factor $a^2 b$. Since there are C_n binary trees with $n+1$ leaves,

$$\mathbb{P}(T_{\textcircled{0}} = \infty, N = n) = C_n a^{n+1} b^n.$$

Summing over the possible values of N leads to (4.28). □

Assume that the binary tree associated with a realisation of the stratified ASG has n leaves. Each leaf corresponds to a region that survives without trifurcations and therefore, the leftmost leaf in the corresponding stratified ASG has infinite weight. Hence, if $y_0 < 1$, the descendant is of type 0 almost surely. If $y_0 = 1$, the descendant is of type 1. We thus also have a purely genealogical proof of the first case in Corollary 4.26.

If $u \geq \check{u}$, by a straightforward comparison with a simple birth-death process, each leaf in the stratified ASG eventually reaches weight 0 almost surely. However, for $u \in [\hat{u}, \check{u}]$ the stratified ASG can survive by escaping via trifurcations the effects of absorbing leaves. The analysis of the underlying tree structures is more difficult and we leave the details to future work.

4.5 Ancestral type distribution via stratified ASG

So far we have been concerned with a randomly chosen individual at present and determined its type via the stratified ASG. Let us now change perspective and consider the type of the *ancestor*, at time r before the present, of our current individual. It will turn out that this can also be tackled by means of a construction that builds on the stratified ASG. Indeed the specific structure of our reduced and pruned trees will be essential.

In line with Sections 3.4 and 3.5, we denote by J_r the type of the ancestor at backward time r of the root of an ASG in $[0,r]$ started with a single line if we propagate types and ancestors as described in Definition 2.11. We abbreviate the probability of J_r to be unfit, conditional on the sampling distribution at the leaves at backward time r being $(1-y_0, y_0)$, by $g_r(y_0) := \mathbb{P}_{y_0}(J_r = 1)$.

4.5.1 Ancestral type distribution without interaction

In the absence of interaction, the pLD-ASG described in Section 3.4 is a tool to determine the ancestral type distribution. We first recall the construction (without interaction) and then introduce a new perspective on the problem.
The pLD-ASG starts from a single individual. The lines of the graph correspond to the relevant potential ancestors (see Section 3.4 for the notion of relevant potential ancestor in the smss-limit) and are assigned consecutive levels, starting at level 1. If a line is hit by a selective arrow, its level is increased by one and at the same time all lines above it are shifted up one level; thereby making space for the incoming line, which then occupies the former level of the line it hit. If the first event on a line that does not occupy the top level is a deleterious mutation, we can conclude that it will not be ancestral, since it will, at a later time, play the role of an unsuccessful incoming line, for its type is 1 due to the mutation. We therefore prune this line at the time of the mutation event. The line occupying the top level is exempt from the pruning since, regardless of its type, this line will be ancestral if all lines below it are non-ancestral. We call this line the immune line. If a line that is not the top line has a mutation to type 0, we can cut away all lines above it, because this line will, at some stage, be incoming to all lines above it. It will, due to the beneficial mutation, be successful in all these selection events. If the top line is hit by a mutation to type 0, this does not have an effect, see Fig. 4.17. This reasoning gives rise to the line-counting process of the pLD-ASG $L = (L_r)_{r\geq 0}$ as a continuous-time Markov chain on $\mathbb{N}$ with transition rates

$$q_L(n,n+1) = ns, \quad q_L(n,n-1) = (n-1)u\nu_1 + \mathbb{1}_{\{n>1\}}u\nu_0, \quad q_L(n,n-j) = u\nu_0, \qquad n \in \mathbb{N},$$

where $j \in \{2,\dots,n-1\}$. The convenient feature of the above construction is that the original individual has an unfit ancestor if and only if all potential ancestors in the pLD-ASG are unfit. The above reasoning leads to (3.25), i.e.

$$g_r(y_0) = \sum_{n=1}^{\infty} \mathbb{P}(L_r = n \mid L_0 = 1)\, y_0^n. \tag{4.30}$$

As a warm-up for the case with interaction, we introduce a different perspective on the pLD-ASG without interaction and without beneficial mutation ($\nu_0 = 0$). In this case, the pLD-

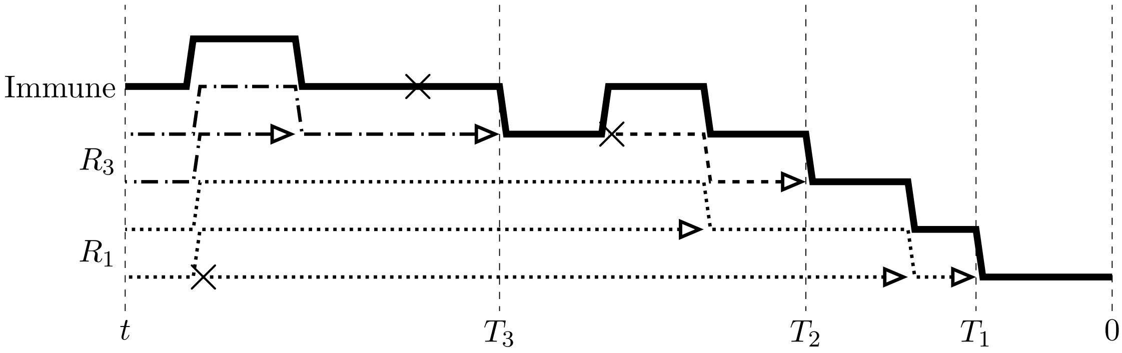


Figure 4.17. A new perspective on the pLD-ASG: from the immune line emerge three independent killed ASGs. They are R_1 (····), R_2 (- -), and R_3 (–·) at times T_1, T_2, and T_3. Here, R_2 has absorbed before backward time t, but R_1 and R_3 have not.

ASG is just a killed ASG (we have recalled its definition in the beginning of Section 4.3.1, see also Section 3.3) for which we ignore the mutations on the immune line. Hence, we can separate the pLD-ASG into two parts. The first part is the immune line, which gives rise to new lines at rate s. Since it is not pruned upon deleterious mutations, it persists indefinitely. The lines that emerge from the immune line evolve independently and each of them is the origin of a new killed ASG . The collection of these mutually independent killed ASGs forms the second part.

Under this new perspective, the ancestor of the root of the pLD-ASG is unfit if and only if 1) at the time of sampling, an unfit type is associated with the immune line, and 2) all killed ASGs emerging from it are unfit at the time of their origin. We will see that the independence of the killed ASGs, the duality (4.11), and standard properties of Poisson processes allow the derivation of the following generalisation of Theorem 3.16 to finite time horizons.

Proposition 4.32 (Ancestral type distribution)**.** *Let* $\nu_0 = 0$ *and* $\gamma = 0$*. Then*

$$g_r(y_0) = y_0 \exp\Big(- s \int_0^r (1 - y(\xi; y_0)) d\xi \Big), \qquad y_0 \in [0,1]. \tag{4.31}$$

In particular,

$$g_r(y_0) = \begin{cases} y_0 \, \frac{u - s\, y(r; y_0)}{u - s y_0}, & \textit{if } y_0 \in [0,1) \setminus \{\frac{u}{s}\}, \\ y_0 \, \exp(-rs(1 - y_0)), & \textit{if } y_0 \in \{\frac{u}{s}, 1\}. \end{cases} \tag{4.32}$$

Furthermore, $g_\infty(y_0) = \lim_{r \to \infty} g_r(y_0)$ *exists and is given as follows.*

- *(i) If* $s = 0$*,* $g_\infty(y_0) = y_0$ *for all* $y_0 \in [0,1]$*.*
- *(ii) If* $u \leq \hat{u}$*,* $g_\infty(y_0) = \begin{cases} 0, & \textit{if } y_0 \in [0,1), \\ 1, & \textit{if } y_0 = 1. \end{cases}$
- *(iii) If* $u > \hat{u}$*,* $g_\infty(y_0) = y_0 \, \frac{u - s}{u - s y_0}$*.*

Proof. The proof of (4.31) is given in the next section in a more general setting (see proof of Theorem 4.40). Given (4.31), (4.32) follows by standard integration techniques. To see this, consider $y_0 < u/s$. Then $y(r; y_0)$ increases and hence

$$-s\int_0^r (1-y(\xi,y_0))d\xi = \int_{y_0}^{y(r;y_0)} \frac{-s}{u-s\eta}d\eta = \ln\left(\frac{u-sy(r;y_0)}{u-sy_0}\right),$$

where we substituted $y(\xi; y_0) = \eta$. Together with (4.31) this leads to (4.32). We can proceed similarly for $y_0 \in (u/s, 1)$. For $y_0 \in \{u/s, 1\}$, one has $y(r; y_0) \equiv y_0$ and the result follows. (i)–(iii) are a consequence of (4.32) together with the form of $y_\infty(y_0)$ from Corollary 4.3 if $\gamma = 0$. □

Let us use Proposition 4.32 to once more make the connection with the deterministic dynamics. Consider the following transformation,

$$z_0(t) := (1-y(t;y_0))\, f(t) \quad \text{and} \quad z_1(t) := y(t;y_0)\, f(t), \tag{4.33}$$

where

$$f(t) = f_0 \exp\left(\int_0^t s(1-y(\xi;y_0))d\xi\right),$$

for some $f_0 > 0$ and $y_0 \in [0, 1]$. The transformation (4.33) is frequently used in deterministic population genetics because it leads to the transformation of the nonlinear equation (4.1) into a linear system (if $\gamma = 0$):

$$\frac{dz_0}{dt}(t) = \Big(sy(t)(1-y(t)) - (1-y(t))u\nu_1\Big)f(t) + s(1-y(t))(1-y(t))\, f(t) = z_0(t)(s-u\nu_1),$$

and, similarly, $(dz_1/dt)(t) = z_0(t)u\nu_1$. This is complemented by $z_0(0) = (1-y_0)f_0$ and $z_1(0) = y_0 f_0$. See also Thompson and McBride [TM74] for more details. The quantities z_0 and z_1 are usually referred to as *absolute frequencies* and $f(t)$ as the absolute size of the population at (forward) time t with initial population size f_0. We have explained in the discussion following (4.1) that setting the neutral reproduction rate to 0 in the underlying model still leads to the ODE (4.1). Then the transformation leads to a system that describes a population with variable population size in which reproduction is not coupled to the death of another individual and only the fit type reproduces. The population size satisfies the differential equation

$$\frac{d}{dt}(z_0(t)+z_1(t)) = s(1-y(t))f(t) = s(1-y(t))(z_0(t)+z_1(t)),$$

with $z_0(0) + z_1(0) = f_0$. Hence, $z_0(t) + z_1(t) = f_0 \exp(\int_0^t s(1-y(\xi))d\xi)$ and this explains the term *absolute size of the population* for $f(t)$. The solution to the original system is recovered by noting that $y(t) = z_1(t)/(z_0(t) + z_1(t))$. In our context, the key is to think in terms of absolute frequencies. Since the individuals unfit at time 0 neither reproduce nor mutate, their absolute frequency remains constant at $y_0 f_0$, while the population size grows by a factor of $f(t)/f_0$. It is therefore clear that the proportion of individuals at time t that have unfit ancestors at time 0 is $y_0 f_0/f(t)$, in line with (4.31).

4.5.2 Ancestral type distribution with interaction: stratified ASG with immune line ($\nu_0 = 0$)

In the remainder of this chapter, we assume $\nu_0 = 0$ (hence $\nu_1 = 1$). We first derive the ancestral type distribution in this case by means of the stratified ASG (Proposition 4.40 and Corollary 4.41). Finally, we consider the ancestral type distribution in an equilibrium population (Corollary 4.43).

The main argument that led to the representation of the ancestral type distribution in the context of (4.30) was that, if $\gamma = 0$, the ancestor of the root of the pLD-ASG is unfit if and only if all individuals in the pLD-ASG are unfit. Therefore, the line-counting process L is sufficient. For $\gamma > 0$, we can not expect a similar statement to hold. For example, already after a single trifurcation, the ancestor can be unfit, even though the individual on the incoming line is fit (an unfit continuing *and* checking line suffices). However, it is possible to generalise the new perspective mentioned before Proposition 4.32 and illustrated in Fig. 4.17, because also in the interactive case the ancestor must be on the immune line. We add trifurcations to the picture as pairs of ASGs emerging from the immune line and argue in the same way as before; that is, an ancestor is unfit if and only if 1) at the time of sampling, an unfit type is associated with the immune line, 2) all ASGs emerging from it via selection events are unfit at the time of their origin, and 3) at each interaction event on the immune line at least one of the origins of the ASGs arising from either the incoming or checking line is unfit. 2) and 3) are satisfied if none of the selective and interactive arrows hitting the immune line is used. Using the results from Section 4.4, we can validate 2) and 3) by means of stratified ASGs started in ① (corresponding to one line: the incoming line) and stratified ASGs started in $\Psi^\star$ (recall that $\Psi^\star$ is a stratified ASG of an ASG consisting of only a single interactive branching corresponding to one checking and one incoming line), respectively. Let us start to make this precise.

Definition 4.33 (Stratified ASG with immune line)**.** Consider $\mathcal{A}_t$, i.e. an ASG in the time interval $[0, t]$ starting with a single line at time 0. We call the *immune line* of $\mathcal{A}_t$ the line that is continuing at all the bifurcation and trifurcation events. The *stratified ASG with immune line* associated with $\mathcal{A}_t$ is defined by $S_\star(\mathcal{A}_t) := S(\mathcal{A}_t^\star)$, where $\mathcal{A}_t^\star$ is the ASG $\mathcal{A}_t$ after deletion of all the deleterious mutations on its immune line. By construction the immune line is always relevant and is associated with the leftmost leaf of the corresponding stratified ASG. We define $H_\star(\mathcal{A}_t, y_0)$ as the probability that the root of the ASG has an unfit ancestor at backward time t, given that the type distribution at backward time t is $(1 - y_0, y_0)$.

Proposition 4.34. *Consider $\mathcal{A}_t$, i.e. an ASG in $[0, t]$ started with a single line. We have*

$$H_\star(\mathcal{A}_t, y_0) = \mathcal{H}(S_\star(\mathcal{A}_t), y_0), \qquad \forall y_0 \in [0, 1], t > 0.$$

Proof. Note that a type assignment to the lines of $\mathcal{A}_t$ at time $r = t$ leads to an unfit ancestor at time t of the single individual at time 0 if and only if the same assignment of types to the lines of $\mathcal{A}_t^\star$ at time t leads to an unfit individual at time 0. Hence, $H_\star(\mathcal{A}_t, y_0) = H(\mathcal{A}_t^\star, y_0)$. Moreover, Theorem 4.15 yields $H(\mathcal{A}_t^\star, y_0) = \mathcal{H}(S(\mathcal{A}_t^\star), y_0)$ because $\mathcal{A}_t^\star$ is a realisation of an ASG. Since by definition $S_\star(\mathcal{A}_t) = S(\mathcal{A}_t^\star)$, the result follows. □

We can also define the stratified ASG with immune line in a Poissonian manner without a realisation of the ASG.

Definition 4.35 (Stratified ASG process with immune line)**.** We define the *stratified ASG process with immune line* as the continuous-time Markov chain $\mathscr{T}_\star = (\mathscr{T}_\star(r))_{r \geq 0}$ on Υ with transition rates, for $\mathcal{T} = (\tau, m_\mathcal{T}) \in \Upsilon$ and $\ell \in L_\mathcal{T}$,

$$q_{\mathscr{T}_\star}(\mathcal{T}, \mathcal{T}_\curlyvee^\ell) = s m_\mathcal{T}(\ell), \qquad q_{\mathscr{T}_\star}(\mathcal{T}, \mathcal{T}_\Psi^\ell) = \gamma m_\mathcal{T}(\ell), \qquad q_{\mathscr{T}_\star}(\mathcal{T}, \mathcal{T}_\times^\ell) = u\nu_1 \big(m_\mathcal{T}(\ell) - \mathbb{1}_{\ell_\mathcal{T}^1}(\ell)\big).$$

The process $\mathscr{T}_\star$ has no absorbing states. As in the case without interaction, we do not insist on starting the process with a single line, i. e. with state ①, but one should keep in mind that if we start the process in a state $\mathcal{T} \in \Upsilon$ with $M(\mathcal{T}) = n$ for some $n > 1$, the process does in general not describe the structure of the relevant potential influencers of n individuals.

Lemma 4.36. *Let $\nu_0 = 0$. For any $r \geq 0$ and $y_0 \in [0,1]$, we have*

$$g_r(y_0) = \mathbb{E}_{\textcircled{1}}[\mathcal{H}(\mathscr{T}_\star(r), y_0)].$$

Proof. Let $\mathcal{A}_r$ denote an ASG in $[0,r]$ starting from a generic individual at time 0. From the definition of g_r and the tower property for conditional expectations, we obtain

$$g_r(y_0) = \mathbb{P}_{y_0}(J_r = 1) = \mathbb{E}_{\textcircled{1}}[\mathbb{E}_{y_0}[\mathbb{1}_{\{J_r = 1\}} \mid \mathcal{A}_r]] = \mathbb{E}_{\textcircled{1}}[H_\star(\mathcal{A}_r, y_0)].$$

The result follows from Proposition 4.34 and the fact that by construction $S_\star(\mathcal{A}_r)$ and $\mathscr{T}_\star(r)$ have the same distribution. □

Let us now try to understand the ancestral type distribution from the perspective of the immune line (as in Section 4.5.1). Consider $\mathcal{A}_t$, i.e. an ASG in $[0,t]$ starting with a single line at $r = 0$. Note that the ancestor at time t of the single line at time 0 is unfit if and only if the ancestor is on the immune line and is unfit at time t. The ancestor is on the immune line if it succeeds at all the selective and interactive events it encounters, i.e. all the selective and interactive events on the immune line are not used. We know from Section 4.3 that in order to decide if the immune line succeeds at a given selective event it is enough to look at the stratified ASG arising from the corresponding incoming line. Similarly, the immune line succeeds at a given interactive event if either the incoming or checking line is unfit. Hence, we can verify if the incoming and checking line at a given interactive event on the immune line is unfit by looking at the ASG arising from the checking and incoming line. The associated stratified ASG starts in $\Psi^\star$, where the middle and right leaves of $\Psi^\star$ correspond to the checking and incoming line, respectively (recall that $\Psi^\star$ is the stratified ASG of an ASG consisting of only a single interactive branching). This idea motivates the following definition.

Definition 4.37 (Forest of stratified ASGs)**.** Consider $\mathcal{A}_t$, i.e. an ASG in the time interval $[0,t]$ starting with a single line at time 0. The *forest of stratified ASGs* associated with $\mathcal{A}_t$ is the collection

$$\mathcal{F}(\mathcal{A}_t) := \Big((S_i^{\textcircled{1}}(\mathcal{A}_t), T_i^{\textcircled{1}})_{i=1}^N, (S_i^{\Psi^\star}(\mathcal{A}_t), T_i^{\Psi^\star})_{i=1}^M \Big),$$

where

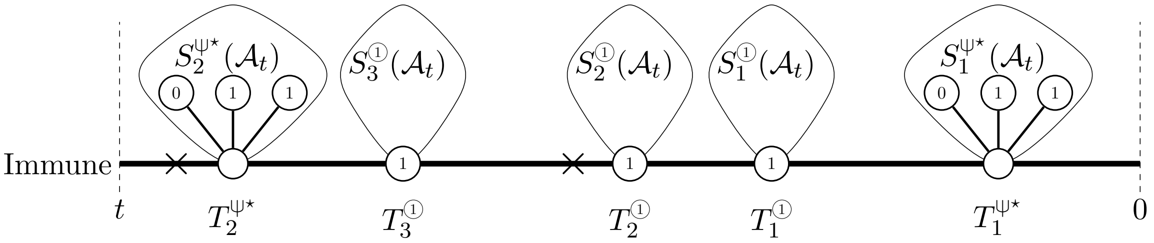


Figure 4.18. From the immune line there emerge independent stratified ASGs started either in ① or $\Psi^\star$. Here, $S_1^{\textcircled{1}}(\mathcal{A}_t)$, $S_2^{\textcircled{1}}(\mathcal{A}_t)$ and $S_3^{\textcircled{1}}(\mathcal{A}_t)$ emerge at times $T_1^{\textcircled{1}}, T_2^{\textcircled{1}}$, and $T_3^{\textcircled{1}}$, respectively. $S_1^{\Psi^\star}(\mathcal{A}_t)$ and $S_2^{\Psi^\star}(\mathcal{A}_t)$ emerge at times $T_1^{\Psi^\star}$ and $T_2^{\Psi\star}$, respectively.

(1) N is the number of selective events on the immune line of $\mathcal{A}_t$ and

$$0 \leq T_1^{\textcircled{1}} < \ldots < T_N^{\textcircled{1}} \leq t$$

are the successive times at which they occur.

(2) For $i \in \{1, ..., N\}$, $S_i^{\textcircled{1}}(\mathcal{A}_t)$ is the stratified ASG associated with the ASG (in the time interval $[T_i^{\textcircled{1}}, t]$) arising from the incoming line at the selective event occurring at time $T_i^{\textcircled{1}}$ in $\mathcal{A}_t$.

(3) M is the number of interactive events on the immune line of $\mathcal{A}_t$ and

$$0 \leq T_1^{\Psi^\star} < \ldots < T_M^{\Psi^\star} \leq t$$

are the successive times at which they occur.

(4) For $i \in \{1, ..., M\}$, $S_i^{\Psi^\star}(\mathcal{A}_t)$ is the stratified ASG associated with the ASG (in the time interval $[T_i^{\Psi^\star}, t]$) arising from the checking and incoming lines at the interactive event occurring at time $T_i^{\Psi^\star}$ in $\mathcal{A}_t$, with checking and incoming lines being assigned to the middle and right leaves of $\Psi^\star$, respectively.

See Fig. 4.18 for an illustration.

Proposition 4.38. *Let $\nu_0 = 0$. Using the notation from Definition 4.37, we have*

$$H_\star(\mathcal{A}_t, y_0) = y_0 \prod_{i=1}^{N} \mathcal{H}(S_i^{\textcircled{1}}(\mathcal{A}_t), y_0) \prod_{j=1}^{M} \mathcal{H}(S_j^{\Psi^\star}(\mathcal{A}_t), y_0), \quad y_0 \in [0,1]. \tag{4.34}$$

Proof. The ancestor at backward time t of the single lineage at time 0 is unfit if and only if the ancestor is on the immune line and it is unfit at time t; the latter is the case with probability y_0. In addition, the individual on the immune line is the ancestor at time t if it succeeds at all the selective and interactive events involved, i.e. all the selective and interactive events one the immune line are not used. The immune line succeeds at the selective event happening at time $T_i^{\textcircled{1}}$ if and only if the corresponding incoming line is of type 1; this occurs with probability $\mathcal{H}(S_i^{\textcircled{1}}(\mathcal{A}_t), y_0)$. The immune line succeeds at the interactive event $T_j^{\Psi^\star}$ if and only if either the checking or the incoming line is of type 1; this occurs with probability $\mathcal{H}(S_j^{\Psi^\star}(\mathcal{A}_t), y_0)$. The result follows from the independence of the corresponding stratified ASGs. □

We can also construct the forest of stratified ASGs in a Poissonian manner.

Definition 4.39 (Forest of stratified ASGs process)**.** Let $\mathcal{N} := (\mathcal{N}_r)_{r\geq 0}$ and $\mathcal{M} := (\mathcal{M}_r)_{r\geq 0}$ be two independent homogeneous Poisson processes with rate s and γ, respectively. Let $(T_i^{①})_{i\in\mathbb{N}}$ and $(T_i^{\Psi^\star})_{i\in\mathbb{N}}$ be the successive arrival times of $\mathcal{N}$ and $\mathcal{M}$, respectively. Furthermore, we invoke two independent collections of independent stratified ASG processes $(\mathscr{T}_i^{①})_{i\in\mathbb{N}}$ and $(\mathscr{T}_i^{\Psi^\star})_{i\in\mathbb{N}}$. In the first collection all the stratified ASGs start at ①, and in the second collection all the stratified ASGs start at $\Psi^\star$. The *forest of stratified ASGs process* $\mathscr{I} = (\mathscr{I}(r))_{r\geq 0}$ is then defined by setting

$$\mathscr{I}(r) := \Big((\mathscr{T}_i^{①}(r - T_i^{①}), T_i^{①})_{i=1}^{\mathcal{N}_r}, (\mathscr{T}_i^{\Psi^\star}(r - T_i^{\Psi^\star}), T_i^{\Psi^\star})_{i=1}^{\mathcal{M}_r} \Big), \quad r \geq 0.$$

Selective (resp. interactive) events on the immune line arrive at the same rate at which $\mathcal{N}$ (resp. $\mathcal{M}$) jumps. Similarly, the stratified ASG associated with the ASG arising from the incoming line at a selective (resp. interactive) event on the immune line has the same distribution as the the stratified ASG process that is started at the time of the selective (resp. interactive) event in ① (resp. $\Psi^\star$). In particular,

$$(\mathcal{F}(\mathcal{A}_r))_{r\geq 0} \overset{d}{=} (\mathscr{I}(r))_{r\geq 0}. \tag{4.35}$$

Theorem 4.40 (Ancestral type distribution)**.** *Let* $\nu_0 = 0$*. Then,*

$$g_r(y_0) = y_0 \exp\Big(- \int_0^r (1 - y(\xi; y_0))(s + \gamma(1 - y(\xi; y_0)))d\xi \Big), \qquad y_0 \in [0, 1]. \tag{4.36}$$

Theorem 4.40 is connected to the deterministic dynamics in the same way as (4.31). This time, $(1 - y(t; y_0))(s + \gamma(1 - y(t; y_0)))$ is the relative reproduction rate of the population (if the rate of neutral reproduction is 0), and

$$f(t) = f_0 \exp\Big(\int_0^t (1 - y(\xi; y_0)((s + \gamma(1 - y(\xi; y_0))))d\xi \Big)$$

is the size of the population at time t in terms of absolute frequencies. With the same argument as before, the proportion of unfit ancestors in the population at time t is $y_0 f_0 / f(t) = g_r(y_0)$, in line with the theorem. In particular, $g_r(y_0)$ describes the relative proportion of individuals at present with unfit ancestor at backward time r if the initial frequency of the unfit type is y_0.

Remark 4.25. Thompson's trick allows the transformation of the (cubic) mutation-selection equation into a system of equations the right hand side of which consists of quadratic polynomials. ◇

The function $g_r(y_0)$ has, in contrast to $y(t; y_0)$, a compact explicit expression, which can be derived via Theorem 4.40 by means of classical integration techniques. We provide the expression in the following corollary.

Corollary 4.41. *Let $\nu_0 = 0$, $\gamma > 0$, and $\bar{y}_1, \bar{y}_2, \bar{y}_3, \sigma$ be given as in* (4.4)*,* (4.6) *and* (4.7)*. For $y_0 \in \{\bar{y}_1, \bar{y}_2, \bar{y}_3\} \cap [0,1]$, we have*

$$g_r(y_0) = y_0 \exp\Big(- r(1-y_0)(s + \gamma(1-y_0))\Big).$$

For $y_0 \in [0,1] \setminus \{\bar{y}_1, \bar{y}_2, \bar{y}_3\}$ and

(i) $\sigma > 0$*,*

$$g_r(y_0) = y_0 \Big(\frac{\bar{y}_2 - y(r;y_0)}{\bar{y}_2 - y_0} \Big)^{\frac{\bar{y}_3}{\sqrt{\sigma}}} \Big(\frac{\bar{y}_3 - y_0}{\bar{y}_3 - y(r;y_0)} \Big)^{\frac{\bar{y}_2}{\sqrt{\sigma}}}, \tag{4.37}$$

(ii) $\sigma = 0$*,*

$$g_r(y_0) = y_0 \frac{y(r;y_0) - \bar{y}_2}{y_0 - \bar{y}_2} \exp\Big(- \bar{y}_2 \frac{y(r;y_0) - y_0}{(y(r;y_0) - \bar{y}_2)(y_0 - \bar{y}_2)} \Big), \tag{4.38}$$

(iii) $\sigma < 0$*,*

$$\begin{aligned} g_r(y_0) = y_0 & \sqrt{\frac{u - y(r;y_0)(s + \gamma(1 - y(r;y_0)))}{u - y_0(s + \gamma(1-y_0))}} \exp\Big(- \frac{1}{\sqrt{-\sigma}} \Big(1 + \frac{s}{\gamma}\Big) \\ & \times \Big[\arctan\Big(2 \frac{y(r;y_0) - \frac{1}{2}(1 + \frac{s}{\gamma})}{\sqrt{-\sigma}} \Big) - \arctan\Big(2 \frac{y_0 - \frac{1}{2}(1+\frac{s}{\gamma})}{\sqrt{-\sigma}} \Big) \Big] \Big). \end{aligned} \tag{4.39}$$

The proof of the corollary is postponed to Section 4.6.

Proof of Theorem 4.40. By Proposition 4.38, equation (4.35), and the independence of $\mathscr{T}_i^{\textcircled{1}}$ and $\mathscr{T}_j^{\Psi^\star}$,

$$\begin{aligned} g_r(y_0) &= y_0 \mathbb{E}\Big[\prod_{i=1}^{N} \mathcal{H}(S_i^{\textcircled{1}}(\mathcal{A}_r), y_0) \prod_{j=1}^{M} \mathcal{H}(S_j^{\Psi^\star}(\mathcal{A}_r), y_0) \Big] \\ &= y_0 \mathbb{E}\Big[\prod_{i=1}^{\mathcal{N}_r} \mathcal{H}(\mathscr{T}_i^{\textcircled{1}}(r - T_i^{\textcircled{1}}), y_0) \prod_{j=1}^{\mathcal{M}_r} \mathcal{H}(\mathscr{T}_j^{\Psi^\star}(r - T_j^{\Psi^\star}), y_0) \Big] \\ &= y_0 \mathbb{E}\Big[\prod_{i=1}^{\mathcal{N}_r} \mathcal{H}(\mathscr{T}_i^{\textcircled{1}}(r - T_i^{\textcircled{1}}), y_0) \Big] \mathbb{E}\Big[\prod_{j=1}^{\mathcal{M}_r} \mathcal{H}(\mathscr{T}_j^{\Psi^\star}(r - T_j^{\Psi^\star}), y_0) \Big] \end{aligned}$$

We begin by considering the first non-trivial factor. Then,

$$\mathbb{E}\Big[\prod_{i=1}^{\mathcal{N}_r} \mathcal{H}(\mathscr{T}_i^{\textcircled{1}}(r - T_i^{\textcircled{1}}), y_0) \Big] = \sum_{n=0}^{\infty} \mathbb{P}(\mathcal{N}_r = n) \mathbb{E}_n \Big[\mathbb{E}\Big[\prod_{i=1}^{n} \mathcal{H}(\mathscr{T}_i^{\textcircled{1}}(r - T_i^{\textcircled{1}}), y_0) \mid (T_i^{\textcircled{1}})_{i=1}^n \Big] \Big], \tag{4.40}$$

where by $\mathbb{E}_n[\cdot]$ we denote the expectation conditional on $\mathcal{N}_r = n$. Now, we use the well-known connection between Poisson processes and the uniform distribution. Conditional on $\mathcal{N}_r = n$, the arrival times of $\mathcal{N}$ have the same distribution as an ordered independent sample of size n from the uniform distribution on $[0,r]$ [Nor98, Thm. 2.4.6]. Let $U_1, \ldots, U_n$

be independent uniformly distributed random variables in $[0,r]$. Since $\mathbb{E}[\prod_{i=1}^{n}\mathcal{H}(\mathscr{T}_i^{①}(r-T_i^{①}),y_0)\mid(T_i^{①})_{i=1}^{n}]$ is a function that is symmetric in the arrival times of the Poisson process, we deduce that

$$\mathbb{E}_n\Big[\mathbb{E}\Big[\prod_{i=1}^{n}\mathcal{H}(\mathscr{T}_i^{①}(r-T_i^{①}),y_0)\mid(T_i^{①})_{i=1}^{n}\Big]\Big]=\mathbb{E}\Big[\prod_{i=1}^{n}\mathcal{H}(\mathscr{T}_i^{①}(U_i),y_0)\Big], \tag{4.41}$$

since $r-U_i$ is again uniform on $[0,r]$. Moreover, $(\mathscr{T}_i^{①}(U_i))_{i=1}^{n}$ are independent, and hence

$$\begin{aligned}\mathbb{E}\Big[\prod_{i=1}^{n}\mathcal{H}(\mathscr{T}_i^{①}(U_i),y_0)\Big]&=\mathbb{E}\Big[\mathcal{H}(\mathscr{T}_1^{①}(U_1),y_0)\Big]^n\\&=\Big(\frac{1}{r}\int_0^r\mathbb{E}\Big[\mathcal{H}(\mathscr{T}_1^{①}(\xi),y_0)\Big]d\xi\Big)^n\\&=\Big(\frac{1}{r}\int_0^r y(\xi,y_0)d\xi\Big)^n,\end{aligned} \tag{4.42}$$

where we used the duality result in Theorem 4.21. Combining (4.41) and (4.42) into (4.40) and using the fact that $\mathcal{N}_r$ is Poisson distributed with parameter sr yields

$$\begin{aligned}\mathbb{E}\Big[\prod_{i=1}^{\mathcal{N}_r}\mathcal{H}(\mathscr{T}_i^{①}(r-T_i^{①}),y_0)\Big]&=\sum_{n=0}^{\infty}\frac{(sr)^n}{n!}e^{-sr}\Big(\frac{1}{r}\int_0^r y(\xi,y_0)d\xi\Big)^n\\&=\exp\Big(-s\int_0^r(1-y(\xi,y_0))d\xi\Big).\end{aligned}$$

Next, we consider the second non-trivial factor. Applying the same techniques that lead to (4.40) and (4.41), we obtain

$$\mathbb{E}_m\Big[\prod_{j=1}^{m}\mathcal{H}(\mathscr{T}_j^{\Psi^\star}(r-T_j^{\Psi^\star}),y_0)\Big]=\mathbb{E}\Big[\mathcal{H}(\mathscr{T}_1^{\Psi^\star}(U_1),y_0)\Big]^m$$

By the tower property, the definition of $\mathscr{T}_1^{\Psi^\star}$, the duality relation of Theorem 4.21, and Definition 4.14,

$$\begin{aligned}\mathbb{E}\Big[\mathcal{H}(\mathscr{T}_1^{\Psi^\star}(U_1),y_0)\Big]^m&=\mathbb{E}\Big[\mathbb{E}\Big[\mathcal{H}(\mathscr{T}_1^{\Psi^\star}(U_1),y_0)\mid U_1\Big]\Big]^m\\&=\mathbb{E}\Big[\mathbb{E}_{\Psi^\star}\Big[\mathcal{H}(\mathscr{T}(U_1),y_0)\mid U_1\Big]\Big]^m\\&=\mathbb{E}\Big[\mathcal{H}(\Psi^\star,y(U_1;y_0))\Big]^m\\&=\Big\{2\mathbb{E}[y(U_1;y_0)]-\mathbb{E}[y(U_1;y_0)^2]\Big\}^m\end{aligned}$$

Using these calculations, we have

$$\mathbb{E}\Big[\prod_{j=1}^{\mathcal{M}_r} \mathcal{H}(\mathscr{T}_j^{\Psi\star}(r - T_j^{\Psi\star}), y_0)\Big] = \sum_{m=0}^{\infty} \frac{(\gamma r)^m}{m!} e^{-\gamma r} \Big(\frac{1}{r}\int_0^r y(\xi; y_0)(2 - y(\xi; y_0)) d\xi\Big)^m$$

$$= \exp\Big(- \gamma \int_0^r (1 - y(\xi; y_0))^2 d\xi\Big).$$

Altogether, we obtain (4.36). ☐

Taking the limit $r \to \infty$, we obtain the analogue to Theorem 3.16 in the case with interaction and in the absence of beneficial mutations.

Corollary 4.42. *Let $\nu_0 = 0$ and $\gamma > 0$. Then we have $g_\infty(1) = 1$. For $y_0 \in [0, 1)$ and*

(i) *$\sigma > 0$, we have*

$$g_\infty(y_0) = \mathbb{1}_{\{y_0 > \bar{y}_3\}} y_0 \Big(\frac{1 - \bar{y}_2}{y_0 - \bar{y}_2}\Big)^{\frac{\bar{y}_3}{\sqrt{\sigma}}} \Big(\frac{y_0 - \bar{y}_3}{1 - \bar{y}_3}\Big)^{\frac{\bar{y}_2}{\sqrt{\sigma}}}, \tag{4.43}$$

(ii) *$\sigma = 0$, we have*

$$g_\infty(y_0) = \mathbb{1}_{\{y_0 > \bar{y}_3\}} y_0 \frac{1 - \bar{y}_2}{y_0 - \bar{y}_2} \exp\Big(- \bar{y}_2 \frac{1 - y_0}{(1 - \bar{y}_2)(y_0 - \bar{y}_2)}\Big), \tag{4.44}$$

(iii) *$\sigma < 0$, we have*

$$g_\infty(y_0) = y_0 \sqrt{\frac{u - s}{u - y_0 s}} \cdot \exp\Bigg(- \frac{1}{\sqrt{-\sigma}}\Big(1 + \frac{s}{\gamma}\Big) \Bigg[\arctan\Bigg(\frac{1 - \frac{s}{\gamma}}{\sqrt{-\sigma}}\Bigg) - \arctan\Bigg(2 \frac{y_0 - \frac{1}{2}(1 + \frac{s}{\gamma})}{\sqrt{-\sigma}}\Bigg)\Bigg]\Bigg). \tag{4.45}$$

Proof. Combining Corollary 4.3 with Corollary 4.41 yields the result. ☐

Finally, we consider the ancestral type distribution at equilibrium, i. e.

$$(1 - g_\infty(y_\infty(y_0)), g_\infty(y_\infty(y_0))).$$

The following corollary extends the expression for $g_\infty(y_\infty(y_0))$ from Section 3.5 to the case $\gamma > 0$.

Corollary 4.43 (Ancestral type distribution at equilibrium)**.** *Let $\nu_0 = 0$ and $s, \gamma \geq 0$. Then we have $g_\infty(y_\infty(1)) = 1$. For all $y_0 \in [0, 1)$,*

(i) *if $s \geq \gamma$, then*

$$g_\infty(y_\infty(y_0)) = \begin{cases} 0, & \textit{if } u < \hat{u}, \\ 1, & \textit{if } u \geq \hat{u}. \end{cases} \tag{4.46}$$

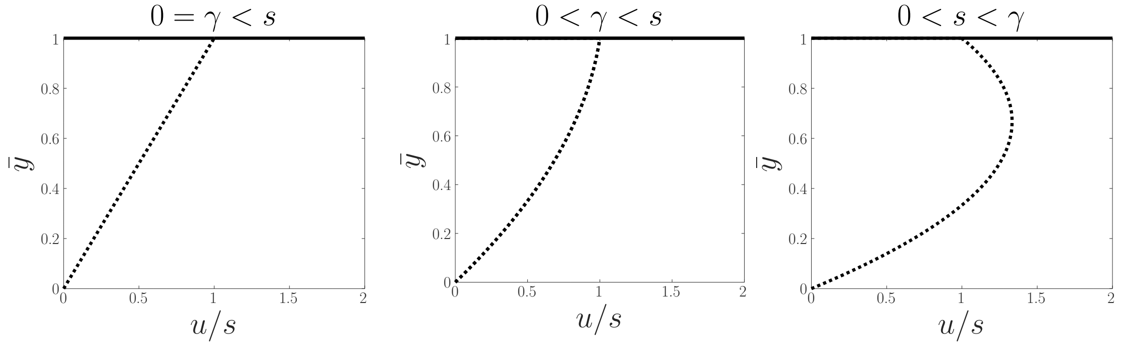


Figure 4.19. The ancestral type distribution at equilibrium for the parameter regimes from Fig. 4.1. Dotted line (resp. solid line): equilibria $\bar{y}$ s.t. $g_\infty(\bar{y}) = 0$ (resp. $g_\infty(\bar{y}) = 1$).

(ii) if $s < \gamma$ *and* $u < \hat{u}$, *then* $g_\infty(y_\infty(y_0)) \equiv 0$.

(iii) if $s < \gamma$ *and* $u \in [\hat{u}, \check{u}]$, *then*

$$g_\infty(y_\infty(y_0)) = \begin{cases} 0, & \textit{if } y_0 \leq \bar{y}_3, \\ 1, & \textit{if } y_0 > \bar{y}_3. \end{cases} \tag{4.47}$$

(iv) If $s < \gamma$ *and* $u > \check{u}$, *then* $g_\infty(y_\infty(y_0)) \equiv 1$.

Proof of Corollary 4.43. In the following, we throughout use Corollaries 4.3 and 4.42. For $y_0 = 1$, we have $y_\infty(1) = 1$ and hence $g_\infty(y_\infty(1)) = 1$. For the remainder, fix $y_0 \in [0, 1)$. If $s \geq \gamma$ and $u < \hat{u}$, we have that $y_\infty(y_0) < 1$ so that $g_\infty(y_\infty(y_0)) = 0$. For $s \geq \gamma$ and $u \geq \hat{u}$, $y_\infty(y_0) = 1$ and hence $g_\infty(y_\infty(y_0)) = 1$. This leads to (i). If $s < \gamma$ and $u < \hat{u}$, then $y_\infty(y_0) < 1$ and hence $g_\infty(y_\infty(y_0)) \equiv 0$. Similarly, $s < \gamma$ and $u > \check{u}$, then $y_\infty(y_0) = 1$ and hence $g_\infty(y_\infty(y_0)) \equiv 1$. Altogether, (ii) and (iv) follow. For (iii), note that when $s < \gamma$ and $u \in [\hat{u}, \check{u}]$, then $y_\infty(y_0) < 1$ for $y_0 \leq \bar{y}_3$ and $y_\infty(y_0) = 1$ for $y_0 > \bar{y}_3$. In particular, $g_\infty(y_\infty(y_0)) = 0$ for $y_0 \leq \bar{y}_3$ and $g_\infty(y_\infty(y_0)) = 1$ for $y_0 > \bar{y}_3$. ☐

Let us explain the underlying genealogical picture. Assume $y_0 \in [0, 1)$. Note that, since the immune line persists indefinitely, the number of lines does not absorb.

- If $u < \hat{u}$, each stratified ASG that emerges from the immune line grows to ∞ with probability $1 - w_1$ by Proposition 4.28. Note that, by the discussion at the end of Section 4.4, $w_1 < 1$. By Corollary 4.26, in this parameter regime a stratified ASG with infinite mass will always lead to a fit descendant provided the potential ancestors are sampled from a population with a positive frequency of fit types. In particular, one of the stratified ASGs emerging from the immune line grows to ∞ and then leads to a fit ancestor.
- If $u > \check{u}$ or ($u \in [\hat{u}, \check{u}]$ and $s \geq \gamma$), we have, by the discussion at the end of Section 4.4, that $y_\infty(y_0) \equiv 1$. In particular, all potential ancestors in the forest of stratified ASGs are unfit.

- If $u \in [\hat{u}, \check{u}]$ and $s < \gamma$, the situation is different. Here, each stratified ASG that emerges from the immune line may grow to ∞ with probability $1 - w_1$ by Proposition 4.28. Note that, by the discussion at the end of Section 4.4, $w_1 < 1$. But now the probability of an unfit ancestor depends crucially on y_0, and we recover the bistability from Section 4.1. The two equilibria of (4.1) lead to a long-term frequency of unfit types that is either < 1 or 1. If $y_0 \in [0, \bar{y}_3)$ (recall $\bar{y}_3 = \bar{y}_3(u, s, \gamma)$ is a function of the parameters) then $y_\infty(y_0) = w_1$, so that by Corollary 4.26 each of the stratified ASGs with infinite mass leads to a fit descendant. One of these descendants then is the fit ancestor. If $y_0 \in (\bar{y}_3, 1]$, then $y_\infty(y_0) = 1$ and all the potential ancestors are unfit. In particular, the total weight of the stratified ASG is irrelevant. In the case $y_0 = \bar{y}_3$, we have $\mathbb{E}[\mathcal{H}_\infty(y_0)] < 1$ by Corollary 4.29 so that one of the infinite stratified ASGs leads to a fit descendant and hence to a fit ancestor, see also Fig. 4.19 (right).

4.6 Remaining proofs

In this section, we provide the remaining proofs. We start with the proof that two total prunings of an ASG are either identical or both consist of a single line ending in the same type of mutation.

Proof of Lemma 4.10. We proceed by induction on the number of events in the ASG. The statement is clearly true for any ASG consisting of exactly one event. Let us assume that the statement holds true for any ASG consisting of at most n events. Let $\mathcal{A}_t$ be an ASG consisting of exactly $n+1$ events and let $T_\star$ be the time to the first event on $\mathcal{A}_t$.

If the event at time $T_\star$ is a mutation, then $\text{✂}(\mathcal{A}_t)$ is a single line ending at the corresponding mutation at time $T_\star$. Hence, $\text{✂}(\mathcal{A}_t)$ is the unique total pruning of $\mathcal{A}_t$. The result follows in this case.

For the remaining cases, denote by $\rho^1(\mathcal{A}_t)$ and $\rho^2(\mathcal{A}_t)$ two total prunings of $\mathcal{A}_t$. If the event at time $T_\star$ is an interactive branching, denote by $\mathcal{A}^2_{T_\star,t}$ and $\mathcal{A}^3_{T_\star,t}$ the ASGs arising at the first interactive event in $\mathcal{A}_t$ from the checking and incoming line, respectively. Moreover, denote by $\mathcal{A}^1_t$ the ASG obtained by removing $\mathcal{A}^2_{T_\star,t}$ and $\mathcal{A}^3_{T_\star,t}$ from $\mathcal{A}_t$. The total prunings ρ^1 and ρ^2 arise from a sequence of mutations $m^1_1, \dots, m^1_l$ and $m^2_1, \dots m^2_k$ such that the consecutive application of the 1-step prunings $\rho_{m^1_j}$ for $j \in [l]$ and $\rho_{m^2_j}$ for $j \in [k]$ lead to ρ^1 and ρ^2, respectively. Denote by $\rho^i(\mathcal{A}^1_t)$, $\rho^i(\mathcal{A}^2_{T_\star,t})$, and $\rho^i(\mathcal{A}^3_{T_\star,t})$ the generalised ASGs obtained by pruning $\mathcal{A}^1_t$, $\mathcal{A}^2_{T_\star,t}$, and $\mathcal{A}^3_{T_\star,t}$, respectively, according to ρ^i, for $i \in \{1,2\}$, ignoring the prunings associated with events that are not in $\mathcal{A}^1_t$, $\mathcal{A}^2_{T_\star,t}$, and $\mathcal{A}^3_{T_\star,t}$, respectively. This means that $\rho^i(\mathcal{A}^1_t)$, $\rho^i(\mathcal{A}^2_{T_\star,t})$, and $\rho^i(\mathcal{A}^3_{T_\star,t})$ arise by consecutively applying the 1-step prunings $\rho^i_{m_j}$ for each mutation m_j that is in $\mathcal{A}^1_t$, $\mathcal{A}^2_{T_\star,t}$, and $\mathcal{A}^3_{T_\star,t}$, respectively. (The pruning of a subtree according to a given total pruning of the entire tree is not always a total pruning, see Fig. 4.20.)

If $\rho^1(\mathcal{A}_t) \sim \circ$, then, by construction, we have $\rho^1(\mathcal{A}^1_t) \sim \circ$ or we have $\rho^1(\mathcal{A}^2_{T_\star,t}) \sim \circ \sim \rho^1(\mathcal{A}^3_{T_\star,t})$. In the first case, $\rho^1(\mathcal{A}^1_t)$ is a total pruning of $\mathcal{A}^1_t$, and by the induction hypothesis, if $\mathcal{A}'$ is a total pruning of $\mathcal{A}^1_t$, then $\mathcal{A}' \sim \circ$. Hence, if $\rho^2(\mathcal{A}^1_t)$ is a total pruning of $\mathcal{A}^1_t$, then $\rho^2(\mathcal{A}^1_t) \sim \circ$, so that also $\rho^2(\mathcal{A}_t) \sim \circ$. On the other hand, if $\rho^2(\mathcal{A}^1_t)$ is not a total pruning of

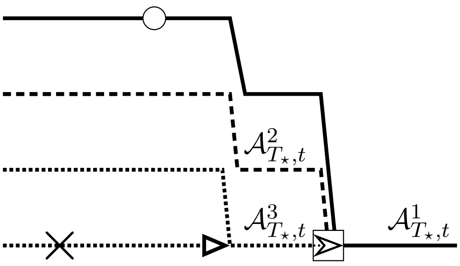


Figure 4.20. Consider the 1-step pruning $\rho_\circ$ and $\rho_\times$ of the beneficial and deleterious mutation, respectively, in the above ASG, which we abbreviate by $\mathcal{A}_t$. Define the composition of 1-step prunings $\rho'(\cdot) := \rho_\circ(\rho_\times(\cdot))$. $\rho_\circ(\mathcal{A}_t)$ and $\rho'(\mathcal{A}_t)$ are total prunings of $\mathcal{A}_t$. Using the same notation as in the proof of Lemma 4.10, we have $\rho'(\mathcal{A}^3_{T_\star,t})$ is a single line without a mutation such that it is in particular a total pruning of $\mathcal{A}^3_{T_\star,t}$. But $\rho_\circ(\mathcal{A}^3_{T_\star,t}) = \mathcal{A}^3_{T_\star,t}$. Since $\mathcal{A}^3_{T_\star,t}$ contains two lines and one mutation $\rho_\circ(\mathcal{A}^3_{T_\star,t})$ is not a total pruning of $\mathcal{A}^3_{T_\star,t}$.

$\mathcal{A}^1_t$, it contains at least two lines and at least one mutation. But then, $\rho^2(\mathcal{A}^2_{T_\star,t}) \sim \circ$ and $\rho^2(\mathcal{A}^3_{T_\star,t}) \sim \circ$, because otherwise ρ^2 is not a total pruning of $\mathcal{A}_t$. Hence, also $\rho^2(\mathcal{A}_t) \sim \circ$. For the second case, i.e. for $\rho^1(\mathcal{A}^2_{T_\star,t}) \sim \circ \sim \rho^1(\mathcal{A}^3_{T_\star,t})$, note that if $\rho^2(\mathcal{A}_t) \sim \circ$, then $\rho^2(\mathcal{A}_t) = \rho^1(\mathcal{A}_t)$. Now, assume $\rho^2(\mathcal{A}_t) \nsim \circ$. Then, either $\rho^2(\mathcal{A}^2_{T_\star,t})$ is a total pruning of $\mathcal{A}^2_{T_\star,t}$ or $\rho^2(\mathcal{A}^3_{T_\star,t})$ is a total pruning of $\mathcal{A}^3_{T_\star,t}$ (because otherwise $\rho^2(\mathcal{A}_t)$ is not a total pruning of $\mathcal{A}_t$). If $\rho^2(\mathcal{A}^2_{T_\star,t})$ is a total pruning of $\mathcal{A}^2_{T_\star,t}$ then, by the induction hypothesis, also $\rho^2(\mathcal{A}^2_{T_\star,t}) \sim \circ$. Then $\rho^2(\mathcal{A}^3_{T_\star,t})$ is also total pruning of $\mathcal{A}^3_{T_\star,t}$ and, by the induction hypothesis, $\rho^2(\mathcal{A}^3_{T_\star,t}) \sim \circ$, so that also $\rho^2(\mathcal{A}_t) \sim \circ$, which is a contradiction. Hence, $\rho^2(\mathcal{A}_t) \sim \circ$ also in this case.

For the other cases, assume $\rho^1(\mathcal{A}_t) \nsim \circ$. Then $\rho^1(\mathcal{A}^1_t)$ is a total pruning of $\mathcal{A}^1_t$ and, in addition, either $\rho^1(\mathcal{A}^2_{T_\star,t})$ is a total pruning of $\mathcal{A}^2_{T_\star,t}$ or $\rho^1(\mathcal{A}^3_{T_\star,t})$ is a total pruning of $\mathcal{A}^3_{T_\star,t}$. Assume without loss of generality that $\rho^1(\mathcal{A}^2_{T_\star,t})$ is a total pruning of $\mathcal{A}^2_{T_\star,t}$.
Assume $\rho^2(\mathcal{A}_t) \sim \circ$. Then either $\rho^2(\mathcal{A}^1_t) \sim \circ$ or $\rho^2(\mathcal{A}^2_{T_\star,t}) \sim \circ \sim \rho^2(\mathcal{A}^3_{T_\star,t})$. The former case is a contradiction to $\rho^1(\mathcal{A}_t)$ being a total pruning of $\mathcal{A}_t$ with $\rho^1(\mathcal{A}_t) \nsim \circ$. In the latter case, $\rho^2(\mathcal{A}^2_{T_\star,t}) = \rho^1(\mathcal{A}^2_{T_\star,t}) \sim \circ$ and therefore $\rho^1(\mathcal{A}^3_{T_\star,t})$ is a total pruning of $\mathcal{A}^3_{T_\star,t}$. But $\rho^1(\mathcal{A}_t) \nsim \circ$ by assumption and so $\rho^1(\mathcal{A}^3_{T_\star,t}) \nsim \circ$, which is a contradiction. Hence, $\rho^2(\mathcal{A}_t) \nsim \circ$.
If $\rho^2(\mathcal{A}_t) \nsim \circ$, then $\rho^2(\mathcal{A}^1_t)$ is a total pruning of $\mathcal{A}^1_t$ with $\rho^2(\mathcal{A}^1_t) = \rho^1(\mathcal{A}^1_t)$, by the induction hypothesis. In addition, either $\rho^2(\mathcal{A}^2_{T_\star,t})$ is a total pruning of $\mathcal{A}^2_{T_\star,t}$ or $\rho^2(\mathcal{A}^3_{T_\star,t})$ is a total pruning of $\mathcal{A}^3_{T_\star,t}$.
If $\rho^2(\mathcal{A}^2_{T_\star,t})$ is a total pruning of $\mathcal{A}^2_{T_\star,t}$, then, by the induction hypothesis, $\rho^2(\mathcal{A}^2_{T_\star,t}) = \rho^1(\mathcal{A}^2_{T_\star,t})$. Assume $\rho^1(\mathcal{A}^2_{T_\star,t}) \sim \circ$ or $\rho^1(\mathcal{A}^2_{T_\star,t})$ is a pruned ASG without mutations. Then $\rho^1(\mathcal{A}^3_{T_\star,t})$ and $\rho^2(\mathcal{A}^3_{T_\star,t})$ are total prunings of $\mathcal{A}^3_{T_\star,t}$, which are, by the induction hypothesis, identical. Hence, $\rho^1(\mathcal{A}_t) = \rho^2(\mathcal{A}_t)$. Now assume $\rho^1(\mathcal{A}^2_{T_\star,t}) \sim \times$. Then $\rho^1(\mathcal{A}_t) = \rho^1(\mathcal{A}^1_t)$ and hence, $\rho^2(\mathcal{A}_t) = \rho^2(\mathcal{A}^1_t)$ so that $\rho^1(\mathcal{A}_t) = \rho^2(\mathcal{A}_t)$.
If $\rho^2(\mathcal{A}^2_{T_\star,t})$ is not a total pruning, then $\rho^2(\mathcal{A}^3_{T_\star,t}) \sim \times$ (otherwise ρ^2 is not a total pruning of $\mathcal{A}_t$). But then, $\rho^2(\mathcal{A}_t) = \rho^2(\mathcal{A}^1_t) = \rho^1(\mathcal{A}^1_t)$. If $\rho^1(\mathcal{A}^2_{T_\star,t}) \sim \times$, then also $\rho^1(\mathcal{A}_t) = \rho^1(\mathcal{A}^1_t)$ and the result follows. If $\rho^1(\mathcal{A}^2_t) \sim \circ$ or $\rho^1(\mathcal{A}^2_{T_\star,t})$ is a total pruning of $\mathcal{A}^2_{T_\star,t}$ without mutations, then also $\rho^1(\mathcal{A}^3_{T_\star,t}) \sim \times$, by the induction hypothesis. The result follows. □

The following identities will be useful in what follows.

Lemma 4.44. *Let $\mathcal{T} = (\tau, m_\tau) \in \Upsilon$.*

(1) Let $k \in \mathbb{N}$ and $m_\tau^{(k)} : L_\tau \to \mathbb{N} \cup \{0\}$ defined by $m_\tau^{(k)}(\ell) := m_\tau(\ell) + \mathbb{1}_{\{\ell = \ell_\tau^1\}} k$. In addition, set $\mathcal{T}^{(k)} := (\tau, m_\tau^{(k)})$. Then,

$$\mathcal{H}(\mathcal{T}^{(k)}, y_0) = y_0^k \mathcal{H}(\mathcal{T}, y_0), \quad y_0 \in [0,1].$$

(2) For all $\widetilde{\mathcal{T}} \in \Upsilon$ and $y_0 \in [0,1]$, we have

$$\mathcal{H}(\mathcal{T} \otimes_{\ell_\tau^1} \widetilde{\mathcal{T}}, y_0) = \mathcal{H}(\mathcal{T}, y_0)\, \mathcal{H}(\widetilde{\mathcal{T}}, y_0).$$

Proof. (1) Fix $k \in \mathbb{N}$ and proceed by induction on $|L_\tau|$. If $\mathcal{T} = \textcircled{n}$, then $\mathcal{T}^{(k)} = \textcircled{j}$ with $j = n + k$, and the assertion follows from the definition of $\mathcal{H}$. Assume the assertion is true for all $\widetilde{\mathcal{T}} = (\tilde\tau, m_{\tilde\tau})$ with $|L_{\tilde\tau}| < |L_\tau|$. We have to show that the assertion remains true for $\mathcal{T}$. Let $\kappa^1, \kappa^2, \kappa^3$ be the left, middle, and right child of the root of $\mathcal{T}$. Since the leftmost leaf of $\mathcal{T}_{\kappa^1}$ is ℓ_τ^1, the induction hypothesis implies that the assertion holds true for $\mathcal{T}_{\kappa^1}$, i. e. $\mathcal{H}(\mathcal{T}_{\kappa^1}^{(k)}, y) = y^k\, \mathcal{H}(\mathcal{T}_{\kappa^1}, y)$. The result then follows via (4.12).

(2) Fix $\widetilde{\mathcal{T}} = (\tilde\tau, \tilde m) \in \Upsilon$ and proceed by induction on $|L_\tau|$. If $\mathcal{T} = \textcircled{n}$, then $\mathcal{T} \otimes_{\ell_\tau^1} \widetilde{\mathcal{T}} = \widetilde{\mathcal{T}}^{(n)}$, and the result follows from assertion (1). Assume the result is true for all $\bar{\mathcal{T}} = (\bar\tau, m_{\bar\tau})$ with $|L_{\bar\tau}| < |L_\tau|$. We have to show that the result remains true for $\mathcal{T}$. Let $\kappa^1, \kappa^2, \kappa^3$ be the left, middle, and right child of the root of $\mathcal{T}$. The induction hypothesis applied to $\mathcal{T}_{\kappa^1}$ yields $\mathcal{H}(\mathcal{T}_{\kappa^1} \otimes_{\ell_\tau^1} \widetilde{\mathcal{T}}, y_0) = \mathcal{H}(\mathcal{T}_{\kappa^1}, y_0)\, \mathcal{H}(\widetilde{\mathcal{T}}, y_0)$. Moreover, since $(\mathcal{T} \otimes_{\ell_\tau^1} \widetilde{\mathcal{T}})_{\kappa^1} = \mathcal{T}_{\kappa^1} \otimes_{\ell_\tau^1} \widetilde{\mathcal{T}}$, $(\mathcal{T} \otimes_{\ell_\tau^1} \widetilde{\mathcal{T}})_{\kappa^2} = \mathcal{T}_{\kappa^2}$ and $(\mathcal{T} \otimes_{\ell_\tau^1} \widetilde{\mathcal{T}})_{\kappa^3} = \mathcal{T}_{\kappa^3}$, the result follows by (4.12). □

Proof of Lemma 4.20. The claim is trivially true if $\mathcal{T} = \Delta$. Hence, fix $\mathcal{T} = (\tau, m) \in \Upsilon$ and $y_0 \in [0,1]$. First note that by Remark 4.17, $\mathcal{H}(\mathcal{T}, \cdot)$ is a polynomial. Hence, $\mathcal{H}(\mathcal{T}, \cdot) \in \mathcal{C}^1([0,1], \mathbb{R})$. We want to show that

$$\mathcal{G}_\Upsilon \mathcal{H}(\cdot, y_0)(\mathcal{T}) = -s y_0 (1 - y_0) \frac{\partial \mathcal{H}(\mathcal{T}, y)}{\partial y}(y_0), \tag{4.48}$$

$$\mathcal{G}_\Psi \mathcal{H}(\cdot, y_0)(\mathcal{T}) = -\gamma y_0 (1 - y_0)^2 \frac{\partial \mathcal{H}(\mathcal{T}, y)}{\partial y}(y_0), \tag{4.49}$$

$$\mathcal{G}_\times \mathcal{H}(\cdot, y_0)(\mathcal{T}) = (1 - y_0) u \nu_1 \frac{\partial \mathcal{H}(\mathcal{T}, y)}{\partial y}(y_0), \tag{4.50}$$

$$\mathcal{G}_\circ \mathcal{H}(\cdot, y_0)(\mathcal{T}) = y_0 u \nu_0 \frac{\partial \mathcal{H}(\mathcal{T}, y)}{\partial y}(y_0). \tag{4.51}$$

We proceed by induction on the size of the underlying tree. First note that, since $\mathcal{H}(\textcircled{n}, y_0) = y_0^n$, we have $(\partial \mathcal{H}(\textcircled{n}, \cdot)/\partial y)(y_0) = n y_0^{n-1}$. In addition,

$$\begin{aligned}
\mathcal{G}_\Upsilon \mathcal{H}(\cdot, y_0)(\textcircled{n}) &= n s (y_0^{n+1} - y_0^n) = -s y_0 (1 - y_0) n y_0^{n-1}, \\
\mathcal{G}_\Psi \mathcal{H}(\cdot, y_0)(\textcircled{n}) &= n \gamma (y_0^n (y_0 + y_0 - y_0^2) - y_0^n) = -\gamma y_0 (1 - y_0)^2 n y_0^{n-1}, \\
\mathcal{G}_\times \mathcal{H}(\cdot, y_0)(\textcircled{n}) &= n u \nu_1 (y_0^{n-1} - y_0^n) = (1 - y_0) u \nu_1 n y_0^{n-1},
\end{aligned}$$

$$\mathcal{G}_\circ\mathcal{H}(\cdot,y_0)(\textcircled{n}) = nu\nu_0(0-y_0^n) = -y_0u\nu_0ny_0^{n-1}.$$

Hence (4.48), (4.49), (4.50), and (4.51) hold true for $\mathcal{T} = \textcircled{n}$. Now, fix $\mathcal{T} = (\tau, m_\tau) \in \Upsilon$ and assume that (4.48), (4.49), (4.50) and (4.51) hold true for all $\widetilde{\mathcal{T}} = (\tilde\tau, \tilde m_{\tilde\tau}) \in \Upsilon$ with $|L_{\tilde\tau}| < |L_\tau|$. We aim to prove that they remain true for $\mathcal{T}$. Denote by κ^1, κ^2, and κ^3 the left, middle, and right child of the root of $\mathcal{T}$. From Definition 4.14 and the chain rule,

$$\begin{aligned}
&\frac{\partial\mathcal{H}(\mathcal{T},y)}{\partial y}(y_0)\\
&= \frac{\partial\mathcal{H}(\mathcal{T}_{\kappa^1},y)}{\partial y}(y_0)\Big[\mathcal{H}(\mathcal{T}_{\kappa^2},y_0)+\mathcal{H}(\mathcal{T}_{\kappa^3},y_0)-\mathcal{H}(\mathcal{T}_{\kappa^2},y_0)\mathcal{H}(\mathcal{T}_{\kappa^3},y_0)\Big]\\
&\quad+\mathcal{H}(\mathcal{T}_{\kappa^1},y_0)\Big[\frac{\partial\mathcal{H}(\mathcal{T}_{\kappa^2},y)}{\partial y}(y_0)\,(1-\mathcal{H}(\mathcal{T}_{\kappa^3},y_0))+\frac{\partial\mathcal{H}(\mathcal{T}_{\kappa^3},y)}{\partial y}(y_0)\,(1-\mathcal{H}(\mathcal{T}_{\kappa^2},y_0))\Big].
\end{aligned} \tag{4.52}$$

We claim that for each $\star \in \{\curlyvee, \Psi, \times, \circ\}$, we have

(1) for $\ell \in L_{\tau_{\kappa^1}}$

$$\begin{aligned}\mathcal{H}(\mathcal{T}^\ell_\star,y_0)-\mathcal{H}(\mathcal{T},y_0) &= \Big(\mathcal{H}((\mathcal{T}_{\kappa^1})^\ell_\star,y_0)-\mathcal{H}(\mathcal{T}_{\kappa^1},y_0)\Big)\\
&\qquad\times\Big[\mathcal{H}(\mathcal{T}_{\kappa^2},y_0)+\mathcal{H}(\mathcal{T}_{\kappa^3},y_0)-\mathcal{H}(\mathcal{T}_{\kappa^2},y_0)\mathcal{H}(\mathcal{T}_{\kappa^3},y_0)\Big],\end{aligned}$$

(2) for $\ell \in L_{\tau_{\kappa^2}}$,

$$\mathcal{H}(\mathcal{T}^\ell_\star,y_0)-\mathcal{H}(\mathcal{T},y_0)=\mathcal{H}(\mathcal{T}_{\kappa^1},y_0)(1-\mathcal{H}(\mathcal{T}_{\kappa^3},y_0))\Big(\mathcal{H}((\mathcal{T}_{\kappa^2})^\ell_\star,y_0)-\mathcal{H}(\mathcal{T}_{\kappa^2},y_0)\Big),$$

(3) for $\ell \in L_{\tau_{\kappa^3}}$,

$$\mathcal{H}(\mathcal{T}^\ell_\star,y_0)-\mathcal{H}(\mathcal{T},y_0)=\mathcal{H}(\mathcal{T}_{\kappa^1},y_0)(1-\mathcal{H}(\mathcal{T}_{\kappa^2},y_0))\Big(\mathcal{H}((\mathcal{T}_{\kappa^3})^\ell_\star,y_0)-\mathcal{H}(\mathcal{T}_{\kappa^3},y_0)\Big).$$

Assume that the claim is true. Denote by

$$q_\star = s\mathbb{1}_{\{\star=\curlyvee\}}+\gamma\mathbb{1}_{\{\star=\Psi\}}+u\nu_1\mathbb{1}_{\{\star=\times\}}+u\nu_1\mathbb{1}_{\{\star=\circ\}}.$$

Since

$$\mathcal{G}_\star\mathcal{H}(\cdot,y_0)(\mathcal{T}) = \sum_{i\in\{1,2,3\}}\sum_{\ell\in L_{\tau_{\kappa^i}}} q_\star m(\ell)(\mathcal{H}(\mathcal{T}^\ell_\star,y_0)-\mathcal{H}(\mathcal{T},y_0)),$$

we infer that

$$\begin{aligned}
&\mathcal{G}_\star\mathcal{H}(\cdot,y_0)(\mathcal{T})\\
&= \mathcal{G}_\star\mathcal{H}(\cdot,y_0)(\mathcal{T}_{\kappa^1})\Big[\mathcal{H}(\mathcal{T}_{\kappa^2},y_0)+\mathcal{H}(\mathcal{T}_{\kappa^3},y_0)-\mathcal{H}(\mathcal{T}_{\kappa^2},y_0)\mathcal{H}(\mathcal{T}_{\kappa^3},y_0)\Big]\\
&\quad+\mathcal{H}(\mathcal{T}_{\kappa^1},y_0)\Big[(1-\mathcal{H}(\mathcal{T}_{\kappa^3},y_0))\mathcal{G}_\star\mathcal{H}(\cdot,y_0)(\mathcal{T}_{\kappa^2})+(1-\mathcal{H}(\mathcal{T}_{\kappa^2},y_0))\mathcal{G}_\star\mathcal{H}(\cdot,y_0)(\mathcal{T}_{\kappa^3})\Big].
\end{aligned}$$

Applying the induction hypothesis to $\mathcal{T}_{\kappa^1}$, $\mathcal{T}_{\kappa^2}$ and $\mathcal{T}_{\kappa^3}$, the previous identity together with

(4.52) allow us to show that (4.48), (4.49), (4.50) and (4.51) hold true for $\mathcal{T}$. It remains to prove the claim.
For $\star \in \{\Upsilon, \Psi\}$, the claim follows easily by Definition 4.14 and noting that for $i \in \{1,2,3\}$ and $\ell \in L_{\mathcal{T}_{\kappa^i}}$, $\mathcal{T}^\ell_\star$ is constructed from $\mathcal{T}$ by replacing $\mathcal{T}_{\kappa^i}$ with $(\mathcal{T}_{\kappa^i})^\ell_\star$. But for the latter the induction hypothesis applies. In the following cases, the tree changes only in one subtree of one of the children of the root and therefore the same argument applies. We are in these cases if $\star = \times$ and

- if $m_\mathcal{T}(\ell) > 1$ or if $m_\mathcal{T}(\ell) = 1$ and ℓ is the left child of its parent,
- if $m_\mathcal{T}(\ell) = 1$, ℓ is not the left child of its parent, and b_ℓ is not the root

or if $\star = \circ$ and

- if ℓ is not a middle or right child of the root,
- if ℓ is a left child and $a_{a^\star_\ell}$ is not the root.

We treat the remaining cases separately. For $\star = \times$, $m_\mathcal{T}(\ell) = 1$, ℓ is not the left child of its parent, and b_ℓ is the root, we have that $\mathcal{T}^\ell_\times = \mathcal{T}_{\kappa^1}$. In particular, $\mathcal{H}(\mathcal{T}^\ell_\times, y_0) = \mathcal{H}(\mathcal{T}_{\kappa^1}, y_0)$ for all $y_0 \in [0,1]$. Furthermore, since b_ℓ is the root, $\ell \in T_{\kappa^2}$ or $\ell \in T_{\kappa^3}$, because otherwise κ^1 is an ancestor of ℓ (or $\ell = \kappa^1$) that is the left child of the root. Assume that $\ell \in \mathcal{T}_{\kappa^2}$. If b_ℓ is the root, then $(\mathcal{T}_{\kappa^2})^\ell_\times = ⓪$ and therefore also $\mathcal{H}((\mathcal{T}_{\kappa^2})^\ell_\times, y_0) = 1$. Again, by Definition 4.14,

$$\begin{aligned}
&\mathcal{H}(\mathcal{T}^\ell_\times, y_0) - \mathcal{H}(\mathcal{T}, y_0)\\
&= \mathcal{H}(\mathcal{T}_{\kappa^1}, y_0) - \mathcal{H}(\mathcal{T}_{\kappa^1}, y_0)[\mathcal{H}(\mathcal{T}_{\kappa^2}, y_0) + \mathcal{H}(\mathcal{T}_{\kappa^3}, y_0) - \mathcal{H}(\mathcal{T}_{\kappa^2}, y_0)\mathcal{H}(\mathcal{T}_{\kappa^3}, y_0)]\\
&= \mathcal{H}(\mathcal{T}_{\kappa^1}, y_0)(1 - \mathcal{H}(\mathcal{T}_{\kappa^3}, y_0))(1 - \mathcal{H}(\mathcal{T}_{\kappa^2}, y_0))\\
&= \mathcal{H}(\mathcal{T}_{\kappa^1}, y_0)(1 - \mathcal{H}(\mathcal{T}_{\kappa^3}, y_0))(\mathcal{H}((\mathcal{T}_{\kappa^2})^\ell_\times, y_0) - \mathcal{H}(\mathcal{T}_{\kappa^2}, y_0)).
\end{aligned}$$

We can proceed in a similar way if $\ell \in \mathcal{T}_{\kappa^3}$. The last case is $\star = \circ$ and ℓ is a child of the root. If ℓ is the middle child of the root, i. e. $\ell = \kappa^2$, then $(\mathcal{T}_\ell)^\ell_\circ = \Delta$. By the definition of $\mathcal{T}^\ell_\circ$, we have

$$\mathcal{T}^\ell_\circ = (\mathcal{T}^C_{\kappa_\tau} \otimes_{\kappa_\tau} \mathcal{T}_{\kappa^1}) \otimes_{\ell^1_{\mathcal{T}_{\kappa^1}}} \mathcal{T}_{\kappa^3} = \mathcal{T}_{\kappa^1} \otimes_{\ell^1_{\mathcal{T}_{\kappa^1}}} \mathcal{T}_{\kappa^3},$$

where κ_τ is the root of τ. By Lemma 4.44, $\mathcal{H}(\mathcal{T}_{\kappa^1} \otimes_{\ell^1_{\mathcal{T}_{\kappa^1}}} \mathcal{T}_{\kappa^3}, y_0) = \mathcal{H}(\mathcal{T}_{\kappa^1}, y_0)\mathcal{H}(\mathcal{T}_{\kappa^3}, y_0)$. Therefore, using Definition 4.14

$$\begin{aligned}
\mathcal{H}(\mathcal{T}^\ell_\circ, y_0) - \mathcal{H}(\mathcal{T}, y_0) &= \mathcal{H}(\mathcal{T}_{\kappa^1}, y_0)\mathcal{H}(\mathcal{T}_{\kappa^3}, y_0)\\
&\quad - \mathcal{H}(\mathcal{T}_{\kappa^1}, y_0)[\mathcal{H}(\mathcal{T}_{\kappa^2}, y_0) + \mathcal{H}(\mathcal{T}_{\kappa^3}, y_0) - \mathcal{H}(\mathcal{T}_{\kappa^2}, y_0)\mathcal{H}(\mathcal{T}_{\kappa^3}, y_0)]\\
&= \mathcal{H}(\mathcal{T}_{\kappa^1}, y_0)(1 - \mathcal{H}(\mathcal{T}_{\kappa^3}, y_0))[-\mathcal{H}(\mathcal{T}_{\kappa^2}, y_0)].
\end{aligned}$$

We can proceed in a similar way if $\ell \in \mathcal{T}_{\kappa^3}$. It remains to prove the case in which ℓ is the left child of its parent and $a_{a^\star_\ell}$ is the root. Assume $a^\star_\ell = \kappa^2$. Then, $(\mathcal{T}_{\kappa^2})^\ell_\circ = \Delta$. Again, $\mathcal{T}^\ell_\circ = (\mathcal{T}^C_{\kappa_\tau} \otimes_{\kappa_\tau} \mathcal{T}_{\kappa^1}) \otimes_{\ell^1_{\mathcal{T}_{\kappa^1}}} \mathcal{T}_{\kappa^3}$. Once more, we apply Lemma 4.44 and use Definition 4.14, so that

$$\begin{aligned}
\mathcal{H}(\mathcal{T}^\ell_\circ, y_0) - \mathcal{H}(\mathcal{T}, y_0) &= \mathcal{H}(\mathcal{T}_{\kappa^1}, y_0)\mathcal{H}(\mathcal{T}_{\kappa^3}, y_0)\\
&\quad - \mathcal{H}(\mathcal{T}_{\kappa^1}, y_0)[\mathcal{H}(\mathcal{T}_{\kappa^2}, y_0) + \mathcal{H}(\mathcal{T}_{\kappa^3}, y_0) - \mathcal{H}(\mathcal{T}_{\kappa^2}, y_0)\mathcal{H}(\mathcal{T}_{\kappa^3}, y_0)]\\
&= \mathcal{H}(\mathcal{T}_{\kappa^1}, y_0)(1 - \mathcal{H}(\mathcal{T}_{\kappa^3}, y_0))[-\mathcal{H}(\mathcal{T}_{\kappa^2}, y_0)].
\end{aligned}$$

We proceed in a similar way if $a_\ell^\star = \kappa^3$. Altogether, this proves the claim. $\square$

Proof of Corollary 4.41. Classical integration theory leads to the result. Recall that $\bar{y}_2$ and $\bar{y}_3$ are the roots of the polynomial $y \mapsto u - y(s + \gamma(1-y))$. First, consider the case $\sigma > 0$ in which $\bar{y}_2$ and $\bar{y}_3$ are both real (recall from (4.6)). We want to treat $u \le s$ in which case $\bar{y}_2 < \bar{y}_1 < \bar{y}_3$. For $y_0 < \bar{y}_2$, $y(r; y_0)$ is increasing. Substituting $\eta = y(\xi, y_0)$ and partial fraction expansion lead to

$$\begin{aligned}
&-\int_0^r (1 - y(\xi; y_0))(s + \gamma(1 - y(\xi; y_0)))d\xi \\
&= -\int_{y_0}^{y(r;y_0)} \frac{s + \gamma(1-\eta)}{u - \eta(s + \gamma(1-\eta))} d\eta \\
&= -\frac{1}{\gamma}\frac{s+\gamma(1-\bar{y}_2)}{\bar{y}_3 - \bar{y}_2}\int_{y_0}^{y(r;y_0)} \frac{1}{\bar{y}_2 - \eta} d\eta + \frac{1}{\gamma}\frac{s+\gamma(1-\bar{y}_3)}{\bar{y}_3 - \bar{y}_2}\int_{y_0}^{y(r;y_0)} \frac{1}{\bar{y}_3 - \eta} d\eta \\
&= \frac{1}{\gamma}\frac{s+\gamma(1-\bar{y}_2)}{\bar{y}_3 - \bar{y}_2}\log\left(\frac{\bar{y}_2 - y(r;y_0)}{\bar{y}_2 - y_0}\right) - \frac{1}{\gamma}\frac{s+\gamma(1-\bar{y}_3)}{\bar{y}_3 - \bar{y}_2}\log\left(\frac{\bar{y}_3 - y(r;y_0)}{\bar{y}_3 - y_0}\right).
\end{aligned}$$

Note that, $s/\gamma + (1-\bar{y}_2) = \bar{y}_3$, $s/\gamma + (1-\bar{y}_3) = \bar{y}_2$, and $\bar{y}_3 - \bar{y}_2 = \sqrt{\sigma}$ so that the claim follows. A similar argument applies, if $\bar{y}_1 > y_0 > \bar{y}_2$; only then $y(r; y_0)$ is decreasing. If $y_0 \in \{\bar{y}_1, \bar{y}_2\}$, then $y(\,\cdot\,; y_0) \equiv y_0$ so that

$$-\int_0^r (1 - y(\xi; y_0))(s + \gamma(1 - y(\xi; y_0)))d\xi = -r(s + \gamma(1-y_0))y_0.$$

If $\sigma > 0$ and $\gamma > s$, we can proceed similarly. The only subtlety lies in the monotonicity of $y(r; y_0)$ depending on y_0. For $\sigma = 0$, we have $\bar{y}_2 = \bar{y}_3$ and $y(r; y_0)$ is increasing for all $y_0 \in [0, 1]$. Hence,

$$\begin{aligned}
-\int_{y_0}^{y(r;y_0)} \frac{s + \gamma(1-\eta)}{u - \eta(s + \gamma(1-\eta))} d\eta &= \int_{y_0}^{y(r;y_0)} -\frac{s+\gamma(1-\bar{y}_2)}{\gamma(\eta - \bar{y}_2)^2} + \frac{1}{\eta - \bar{y}_2} d\eta \\
&= \bar{y}_2\left(\frac{1}{y(r;y_0) - \bar{y}_2} - \frac{1}{y_0 - \bar{y}_2}\right) + \log\left(\frac{y(r;y_0) - \bar{y}_2}{y_0 - \bar{y}_2}\right)
\end{aligned}$$

Finally, we treat the case $\sigma < 0$. Again, $y(r; y_0)$ is increasing. Here,

$$\begin{aligned}
&-\int_{y_0}^{y(r;y_0)} \frac{s + \gamma(1-\eta)}{u - \eta(s + \gamma(1-\eta))} d\eta \\
&= \frac{1}{2}\int_{y_0}^{y(r;y_0)} \frac{-(s+\gamma) + 2\gamma\eta}{u - (s+\gamma)\eta + \gamma\eta^2} - \frac{1}{2}\int_{y_0}^{y(r;y_0)} \frac{s+\gamma}{u - (s+\gamma)\eta + \gamma\eta^2} d\eta \\
&= \frac{1}{2}\log\left(\frac{u - y(r;y_0)(s + \gamma(1 - y(r;y_0)))}{u - y_0(s + \gamma(1-y_0))}\right) - \frac{1}{2}\int_{y_0}^{y(r;y_0)} \frac{s+\gamma}{u - (s+\gamma)\eta + \gamma\eta^2} d\eta
\end{aligned}$$

In the last term, we substitute $\mu = \phi(\eta) := 2(\eta - \frac{1}{2}(1+\frac{s}{\gamma}))/\sqrt{-\sigma}$ and we obtain

$$\begin{aligned}
-\frac{1}{2}\int_{y_0}^{y(r;y_0)} \frac{s+\gamma}{u-(s+\gamma)\eta+\gamma\eta^2}d\xi &= -\frac{1}{\sqrt{-\sigma}}\Big(1+\frac{s}{\gamma}\Big)\int_{\phi(y_0)}^{\phi(y(r;y_0))}\frac{1}{1+\mu^2}d\mu \\
&= -\frac{1}{\sqrt{-\sigma}}\Big(1+\frac{s}{\gamma}\Big)\Big[\arctan\big(\phi(y(r;y_0))\big) - \arctan\big(\phi(y_0)\big)\Big].
\end{aligned}$$

This ends the proof of Corollary 4.41. □

Lemma 4.45. *Let $Z = (Z_t)_{t\geq 0}$ be a binary Galton-Watson process with birth rate λ and death rate μ with $\mu < \lambda$. Let $\hat{Z} = (\hat{Z}_t)_{t\geq 0}$ be the same Galton-Watson process but conditioned to die out eventually. Then the transition rates of $\hat{Z}$ are given by*

$$q_{\hat{Z}}(n,n-1) = \lambda n, \qquad q_{\hat{Z}}(n,n+1) = \mu n$$

This is the continuous-time version of the classical discrete-time result (see [AN72, Thm. 3, Ch. I.12.3]).

Proof of Lemma 4.45. Consider a binary Galton-Watson process $Z = (Z_t)_{t\geq 0}$ with birth rate λ and death rate μ. The extinction probability of Z that is started at n is given by

$$h(n) := \mathbb{P}(Z_\infty = 0 \mid Z_0 = n) = (\mu/\lambda)^n.$$

By Doob's h-transform the rates of $\hat{Z}$ are given by

$$q_{\hat{Z}}(n,n-1) = q_Z(n,n-1)\frac{h(n-1)}{h(n)} = \lambda n, \qquad q_{\hat{Z}}(n,n+1) = q_Z(n,n+1)\frac{h(n+1)}{h(n)} = \mu n.$$

□

5

Summary and discussion

Summary

During the last decades, deterministic and stochastic population genetics have largely led separate lives. It is the purpose of this thesis to bring the two areas of research closer together by extending the backward point of view, so far reserved for stochastic models of population genetics, to deterministic mutation-selection equations. We establish the appropriate structures via the Moran model and its graphical representation as an interacting particle system. This leads to a probabilistic interpretation of the classical mutation-selection equation of population genetics. In particular, we equip the bifurcation structure of the corresponding ODEs with a genealogical picture.

The graphical representation naturally provides a way to establish ancestral structures in the Moran model. This framework gives rise to a *set-valued dual process.* Even though we do not make this construction precise, let us explain the basic idea. The starting point is a given typed sample of the population in the future. A configuration is an assignment of types to the lines of the ancestral structure, which is embedded into the graphical representation. The set-valued dual process traces back in the graphical representation all configurations that are consistent with the given sample from the future. For the Fleming-Viot process, which can be interpreted as the measure-valued analogue of the Moran model for large populations, Dawson and Greven [DG14, Ch. 5.5] explicitly construct such a dual process ([DG11, Sect. 3] already contains the idea and a formulation of the duality result). Since this kind of process may still contain a considerable amount of superfluous information, one might try to tailor the dual process in a way that allows the efficient inference of relevant information.

In this thesis, we demonstrate how to construct a dual process for the Moran model with mutation and selection. We achieve this by suitably tailoring the ancestral process that naturally arises from the graphical representation. Our goal is to provide expressions for the stationary type distribution, fixation probability, and ancestral type distribution by genealogical means. In the absence of pairwise interaction, we do this by lumping the genealogical information to just the number of lines in the ASG. The (dual) ancestral processes are tailored to the specific problems and lead to the line counting process of the ASG, of the killed ASG, and of the pLD-ASG, respectively.

We can embed the ASG into the graphical representation, but we can also construct the ASG without the entire graphical representation. This allows us to make sense of the ancestral processes in the large population limit. In particular, the above processes then translate to the smss-limit. A duality rigorously connects the forward model with the killed ASG in this limit. The representation for the stationary type distribution in the Moran model in terms of the ancestral structure carries over to a representation for the equilibria of the ODE. In particular, our approach explains the bifurcation phenomenon of the error-threshold in terms of the genealogy. The ancestral structures allow the notion of an ancestral type. The large population limit of the pLD-ASG for finite populations serves as our main tool to analyse its distribution also in this framework.

In the case of pairwise interaction, it is more challenging to tailor a suitable process. The reason is that it is not sufficient to count the number of lines in the ASG; their connection is also relevant. This requires a process that retains more information about the underlying ancestral structure than just the number of lines. In the smsssi-limit, the ancestral lines evolve independently, which allows us to exploit the inherent branching structure. A suitable ancestral process codes a partition of the lines in the ASG and the hierarchy of the partition elements. This leads to the stratified ASG, whose long-term behaviour is more complex than that of the killed ASG. This is not unexpected since the equilibria of the associated mutation-selection equation with pairwise interaction are also the roots of a cubic (as opposed to a quadratic) polynomial. Nevertheless, we recover the richer bifurcation structure via the stratified ASG. The genealogical approach allows us to characterise the ancestral distributions also in the case of pairwise interaction and univariate mutation. We do this by exploiting a special feature of the ancestral structure. This gives rise to the stratfied ASG with immune line and a sequence of stratified ASGs, which we call forest of stratified ASGs.

Discussion

The genealogical point of view provides tools that can be applied to population models in various parameter regimes. Hence, identifying the ancestral processes and understanding their suitable tailoring is crucial.

Once an ancestral process is tailored within the Moran model, it seems that this process is robust in the sense that it translates to most large population limits. The converse is not necessarily true. For instance, in the finite setting of the Moran model with pairwise interaction, ancestral lines are not independent and therefore we do not expect a similar simplification of the ASG as in the smsssi-framework. It would be interesting to see if this setting also allows for an efficient lumping. If it does, the corresponding dual process should be available in the classical large population limits. In particular, this should pave the way for an analysis of the diffusion limit with interaction by genealogical means.

There are models for more complex forms of frequency-dependent selection. One such generalisation corresponds to interactions that include a finite number of individuals. In the corresponding differential equation this leads to a drift term of the form $y(1-y)P(y)$ for general polynomial $P(y)$. It would be desirable to determine a systematic way to tailor and lump the set-valued dual process for these more complex forms of frequency-dependent selection. We are convinced that the ideas behind the killed and stratified ASG may help to identify the corresponding dualities.

In a next step, it is natural to study genealogies that arise for even more complex forward dynamics; for example, those that allow *large offspring events.* The analysis of such models is much more involved. In the classical large population limits, these kind of dynamics lead to jump-diffusion processes and most of the classical diffusion theory does not apply. In the corresponding retrospective approach, this implies the inclusion of Λ- *or* Ξ-*coalescent* structures. Loosely speaking, in the Λ-coalescent, multiple ancestral lines may at the same time coalesce into a single line. The rate of such coalescence events is described by a measure on $[0, 1]$, which is called the Λ-measure. See Pitman [Pit99] and Sagitov [Sag99] for details. In the Ξ-coalescent, blocks of ancestral lines coalesce into multiple ancestral lines simultaneously. The rate is then described by a more complicated measure. We refer to Möhle and Sagitov [MS01] and Schweinsberg [Sch00] for a thorough treatment.

There already exists some work that exploits the genealogical perspective to determine properties of such forward processes. For example, Baake et al. [BLW16] use a pLD-ASG to characterise the ancestral type distribution in the Λ-Wright-Fisher model with (frequency-independent) selection. Etheridge et al. [EGT10] provide another example of an analysis that is based on a genealogical approach of a more complex forward dynamic. They establish a duality between a coalescent process that is described by a finite measure on $[0, 1]$ and a Λ-Wright-Fisher process. As an application, they obtain a transition function expansion of the forward process as a mixture of the transition function of the dual process. Foucart [Fou13] and Bah and Pardoux [BP15] consider a Λ-Wright-Fisher model and use the genealogical structure to determine criteria for the almost sure absorption at (one of) the boundaries. Building up on this work, González Casanova and Spanò [GS18] use the ASG and its long-term behaviour to derive similar conditions for Ξ-Wright-Fisher models with a special form of frequency-dependent selection.

It would be desirable to generalise the existing ancestral processes to models with *multiple fitness types.* The problems to overcome are very similar in spirit to the ones that arise in the two-type case with pairwise interaction in the finite Moran model or in its diffusion limit: the valid configurations that arise in a dual process in the spirit of Dawson and Greven [DG14] are not a simple function of the number of lines in the ASG.

Nevertheless, for the finite Moran model with mutation, selection, and multiple fitness types, Etheridge and Griffiths [EG09] establish a relation of the corresponding forward and backward model in terms of a weighted moment duality.

Kimura [Kim55] describes a transition density expansion for the Wright-Fisher diffusion with (frequency-independent) selection via orthogonal polynomials. Recently, Mano [Man09] obtains a similar expansion for an associated ancestral process. Song and Steinrücken [SS12] and Steinrücken et al. [SWS13] describe a transition density expansion in terms of orthogonal polynomials for more complicated forward processes. This includes also the case of multiple (fitness) types. It remains to be established how these decompositions of the forward picture relate to the model inherent ancestral structures.

Bibliography

[AB05] Aldous, D. J. and Bandyopadhyay, A. A survey of max-type recursive distributional equations. *Ann. Appl. Probab.* 15 (2005), pp. 1047–1110.

[AN72] Athreya, K. B. and Ney, P. E. *Branching Processes.* Springer, New York, NY (1972).

[AS05] Athreya, S. R. and Swart, J. M. Branching-coalescing particle systems. *Probab. Theor. Relat. Fields* 131 (2005), pp. 376–414.

[AC74] Ayala, F. J. and Campbell, C. A. Frequency-dependent selection. *Annu. Rev. Ecol. Syst.* 5 (1974), pp. 115–138.

[BCH18a] Baake, E., Cordero, F., and Hummel, S. A probabilistic view on the deterministic mutation–selection equation: dynamics, equilibria, and ancestry via individual lines of descent. *J. Math. Biol.* 77 (2018), pp. 795–820.

[BCH18b] Baake, E., Cordero, F., and Hummel, S. Lines of descent in the deterministic mutation-selection model with pairwise interaction. *ArXiv e-prints* (2018). eprint: `1812.00872`. *Submitted for publication.*

[BLW16] Baake, E., Lenz, U., and Wakolbinger, A. The common ancestor type distribution of a Λ–Wright-Fisher process with selection and mutation. *Electron. Commun. Probab.* 21 (2016), no. 59, 16 pp.

[BW97] Baake, E. and Wiehe, T. Bifurcations in haploid and diploid sequence space models. *J. Math. Biol.* 35 (1997), pp. 321–343.

[BP15] Bah, B. and Pardoux, E. The Λ-lookdown model with selection. *Stoch. Proc. Appl.* 125 (2015), pp. 1089–1126.

[BES04] Barton, N. H., Etheridge, A. M., and Sturm, A. K. Coalescence in a random background. *Ann. Appl. Probab.* 14 (2004), pp. 754–785.

[Bre68] Breiman, L. *Probability.* Addison-Wesley, Reading, Mass. (1968).

[Bud07] Budan, F.-D. *Nouvelle Méthode pour la Résolution des Équations Numériques d'un Degré quelconque: d'après laquelle toute le Calcul Exigé pour cette Résolution se Réduit à l'emploi des deux Premières Règles de l'Arithmétique.* Courcier, Paris (1807).

[Bür00] Bürger, R. *The Mathematical Theory of Selection, Recombination, and Mutation.* Wiley, Chichester (2000).

[CS85] Clifford, P. and Sudbury, A. A sample path proof of the duality for stochastically monotone Markov processes. *Ann. Probab.* 13 (1985), pp. 558–565.

[Cor17a] Cordero, F. Common ancestor type distribution: A Moran model and its deterministic limit. *Stoch. Proc. Appl.* 127 (2017), pp. 590–621.

[Cor17b] Cordero, F. The deterministic limit of the Moran model: a uniform central limit theorem. *Markov Proc. Relat. Fields* 23 (2017), pp. 313–324.

[CM19] Cordero, F. and Möhle, M. On the stationary distribution of the block counting process for population models with mutation and selection. *J. Math. Anal. Appl.* 474 (2019), pp. 1049–1081.

[CK56] Crow, J. F. and Kimura, M. Some genetic problems in natural populations. In: *Proc. Third Berkeley Symp. on Math. Statist. and Prob.* Ed. by Neyman, J. Vol. 4. Univ. of Calif. Press, Berkeley, CA 1956, pp. 1–22.

[CK70] Crow, J. F. and Kimura, M. *An Introduction to Population Genetics Theory.* Harper & Row, New York, (1970).

[Dar59] Darwin, C. *On the Origin of Species by Means of Natural Selection, or the Preservation of Favoured Races in the Struggle for Life.* J. Murray, London (1859).

[DG11] Dawson, D. A. and Greven, A. Duality for spatially interacting Fleming-Viot processes with mutation and selection. *ArXiv e-prints* (2011). eprint: `1104.1099`.

[DG14] Dawson, D. A. and Greven, A. *Spatial Fleming-Viot Models with Selection and Mutation.* Springer, Cham (2014).

[DK99a] Donnelly, P. and Kurtz, T. G. Genealogical processes for Fleming-Viot models with selection and recombination. *Ann. Appl. Probab.* 9 (1999), pp. 1091–1148.

[DK99b] Donnelly, P. and Kurtz, T. G. Particle representations for measure-valued population models. *Ann. Probab.* 27 (1999), pp. 166–205.

[Dur08] Durrett, R. *Probability Models for DNA Sequence Evolution, 2nd edition.* Springer, New York, (2008).

[Eig71] Eigen, M. Self-organization of matter and the evolution of biological macromolecules. *Naturwissenschaften* 58 (1971), pp. 465–523.

[EMS89] Eigen, M., McCaskill, J., and Schuster, P. The molecular quasi-species. *Adv. Chem. Phys.* 75 (1989), pp. 149–263.

[Eth11] Etheridge, A. M. *Some Mathematical Models from Population Genetics: École D'Été de Probabilités de Saint-Flour XXXIX-2009.* Vol. 2012. Springer, Berlin Heidelberg (2011).

[EG09] Etheridge, A. M. and Griffiths, R. C. A coalescent dual process in a Moran model with genic selection. *Theor. Popul. Biol.* 75 (2009), pp. 320–330.

[EGT10] Etheridge, A. M., Griffiths, R. C., and Taylor, J. E. A coalescent dual process in a Moran model with genic selection, and the lambda coalescent limit. *Theor. Popul. Biol.* 78 (2010), pp. 77–92.

[EK86] Ethier, S. N. and Kurtz, T. G. *Markov Processes: Characterization and Convergence.* Wiley, Chichester (1986).

[Ewe04] Ewens, W. J. *Mathematical Population Genetics 1: Theoretical Introduction.* Springer, New York, (2004).

[Fea02] Fearnhead, P. The common ancestor at a nonneutral locus. *J. Appl. Probab.* 39 (2002), pp. 38–54.

[Fel51] Feller, W. Diffusion processes in genetics. In: Proceedings of second Berkeley symposium on mathematical statistics and probability. (1950). Ed. by Neyman, J. Univ. of Calif. Press, Berkeley, Los Angeles, CA 1951, pp. 227–246.

[Fis30] Fisher, R. A. *The Genetical Theory of Natural Selection.* Clarendon Press, Oxford, (1930).

[Fou13] Foucart, C. The impact of selection in the Λ–Wright-Fisher model. *Electron. Commun. Probab.* 18 (2013), no. 72, 10 pp.

[GB03] Georgii, H.-O. and Baake, E. Supercritical multitype branching processes: the ancestral types of typical individuals. *Adv. in Appl. Probab.* 35 (2003), pp. 1090–1110.

[GS18] González Casanova, A. and Spanò, D. Duality and fixation in Ξ–Wright–Fisher processes with frequency-dependent selection. *Ann. Appl. Probab.* 28 (2018), pp. 250–284.

[GT94] Griffiths, R. C. and Tavaré, S. Simulating probability distributions in the coalescent. *Theor. Popul. Biol.* 46 (1994), pp. 131–159.

[GH83] Guckenheimer, J. and Holmes, P. *Nonlinear Oscillations, Dynamical Systems, and Bifurcations of Vector Fields.* 4th printing 1993. Vol. 42. Springer, New York, NY (1983).

[Har78] Harris, T. E. Additive set-valued Markov processes and graphical methods. *Ann. Probab.* 6 (1978), pp. 355–378.

[Hén15] Hénard, O. The fixation line in the Λ-coalescent. *Ann. Appl. Probab.* 25 (2015), pp. 3007–3032.

[HS98] Hofbauer, J. and Sigmund, K. *Evolutionary Games and Population Dynamics.* Cambridge Univ. Press, Cambridge (1998).

[Jag89] Jagers, P. General branching processes as Markov fields. *Stoch. Proc. Appl.* 32 (1989), pp. 183–212.

[Jag92] Jagers, P. Stabilities and instabilities in population dynamics. *J. App. Prob.* 29 (1992), pp. 770–780.

[JK14] Jansen, S. and Kurt, N. On the notion(s) of duality for Markov processes. *Probab. Surveys* 11 (2014), pp. 59–120.

[JH85] Johnson, C. R. and Horn, R. A. *Matrix analysis.* Cambridge Univ. Press, Cambridge (1985).

[KDH88] Kaplan, N. L., Darden, T., and Hudson, R. R. The coalescent process in models with selection. *Genetics* 120 (1988), pp. 819–829.

[KM57] Karlin, S. and McGregor, J. The classification of birth and death processes. *Trans. Amer. Math. Soc.* 86 (1957), pp. 366–400.

[KP91] Kelley, W. G. and Peterson, A. C. *Difference Equations.* 2nd ed. Academic Pr., New York, NY (1991).

[KP04] Kelley, W. G. and Peterson, A. C. *The Theory of Differential Equations: Classical and Qualitative.* Pearson Educ., Upper Saddle River, NJ (2004).

[Kel11] Kelly, F. P. *Reversibility and Stochastic Networks.* Cambridge Univ. Press, New York, NY (2011).

[Kim55] Kimura, M. Stochastic processes and distribution of gene frequencies under natural selection. In: *Cold Spring Harbor Symposia on Quantitative Biology.* Vol. 20. 1955, pp. 33–53.

[Kin82a] Kingman, J. F. C. Exchangeability and the evolution of large populations. In: North-Holland, Amsterdam, 1982, pp. 97–112.

[Kin82b] Kingman, J. F. C. On the genealogy of large populations. *J. Appl. Probab.* 19, A (1982), pp. 27–43.

[Kin82c] Kingman, J. F. C. The coalescent. *Stoch. Process. Appl.* 13 (1982), pp. 235–248.

[KHB13] Kluth, S., Hustedt, T., and Baake, E. The common ancestor process revisited. *Bull. Math. Biol.* 75 (2013), pp. 2003–2027.

[KN97] Krone, S. M. and Neuhauser, C. Ancestral processes with selection. *Theor. Popul. Biol.* 51 (1997), pp. 210–237.

[Kur70] Kurtz, T. G. Solutions of ordinary differential equations as limits of pure jump Markov processes. *J. Appl. Probab.* 7 (1970), pp. 49–58.

[Len+15] Lenz, U., Kluth, S., Baake, E., and Wakolbinger, A. Looking down in the ancestral selection graph: A probabilistic approach to the common ancestor type distribution. *Theor. Popul. Biol.* 103 (2015), pp. 27–37.

[Mac17] Mach, T. Dualities and genealogies in stochastic population models. Dissertation. Universität Göttingen, 2017.

[MSS18a] Mach, T., Sturm, A., and Swart, J. M. A new characterization of endogeny. *Math. Phys. Anal. Geom.* 21 (2018), 30 pp.

[MSS18b] Mach, T., Sturm, A., and Swart, J. M. Recursive tree processes and the mean-field limit of stochastic flows. *ArXiv e-prints* (2018). eprint: `1812.10787`.

[Mal48] Malécot, G. *Les Mathématiques de l'Hérédité.* Masson, Paris, (1948).

[Man09] Mano, S. Duality, ancestral and diffusion processes in models with selection. *Theor. Popul. Biol.* 75 (2009), pp. 164–175.

[Men66] Mendel, G. Versuche über Pflanzen-Hybriden. *Verhandlungen des naturforschenden Vereines in Brünn Bd. IV für das Jahr 1865* (1866), pp. 3–47.

[MS01] Möhle, M. and Sagitov, S. A classification of coalescent processes for haploid exchangeable population Models. *Ann. Probab.* 29 (2001), pp. 1547–1562.

[Mor58] Moran, P. A. P. Random processes in genetics. In: *Math. Proc. Cam. Phil. Soc.* Ed. by Neyman, J. Vol. 54. Cambridge Univ. Press, Cambridge 1958, pp. 60–71.

[Mül78] Müller, F. Ueber die Vortheile der Mimicry bei Schmetterlingen. *Zool. Anz.* 1 (1878), pp. 54–55.

[Mül79] Müller, F. Ituna and Thyridia; a remarkable case of mimicry in butterflies (transl. by R. Meldola from the original German article in Kosmos, May 1879, 100). *Trans. R. ent. Soc. Lond.* (1879), pp. 20–29.

[Neu94] Neuhauser, C. A long range sexual reproduction process. *Stoch. Proc. Appl.* 53 (1994), pp. 193–220.

[Neu99] Neuhauser, C. The ancestral graph and gene genealogy under frequency-dependent selection. *Theor. Popul. Biol.* 56 (1999), pp. 203–214.

[NK97] Neuhauser, C. and Krone, S. M. The genealogy of samples in models with selection. *Genetics* 145 (1997), pp. 519–534.

[Nob92] Noble, C. Equilibrium behavior of the sexual reproduction process with rapid diffusion. *Ann. Probab.* 20 (1992), pp. 724–745.

[Nor98] Norris, J. R. *Markov Chains, 2nd ed.* Cambridge Univ. Press, Cambridge (1998).

[PW06] Pfaffelhuber, P. and Wakolbinger, A. The process of most recent common ancestors in an evolving coalescent. *Stoch. Process. Appl.* 116 (2006), pp. 1836–1859.

[Pit99] Pitman, J. Coalescents with multiple collisions. *Ann. Probab.* 27 (1999), pp. 1870–1902.

[PP13] Pokalyuk, C. and Pfaffelhuber, P. The ancestral selection graph under strong directional selection. *Theor. Popul. Biol.* 87 (2013), pp. 25–33.

[Sag99] Sagitov, S. The general coalescent with asynchronous mergers of ancestral lines. *J. Appl. Probab.* 36 (1999), pp. 1116–1125.

[Sch00] Schweinsberg, J. Coalescents with simultaneous multiple collisions. *Electron. J. Probab.* 5 (2000), no. 12, 50 pp.

[Shi81] Shiga, T. Diffusion processes in population genetics. *J. Math. Kyoto Univ.* 21 (1981), pp. 133–151.

[SU86] Shiga, T. and Uchiyama, K. Stationary states and their stability of the stepping stone model involving mutation and selection. *Probab. Theor. Relat. Fields* 73 (1986), pp. 87–117.

[SC74] Shivakumar, P. N. and Chew, K. H. A sufficient condition for nonvanishing of determinants. *Proc. Amer. Math. Soc.* (1974), pp. 63–66.

[SS12] Song, Y. S. and Steinrücken, M. A simple method for finding explicit analytic transition densities of diffusion processes with general diploid selection. *Genetics* 190 (2012), pp. 1117–1129.

[SWS13] Steinrücken, M., Wang, Y. X. R., and Song, Y. S. An explicit transition density expansion for a multi-allelic Wright-Fisher diffusion with general diploid selection. *Theor. Popul. Biol.* 83 (2013), pp. 1–14.

[SS15] Sturm, A. and Swart, J. M. A particle system with cooperative branching and coalescence. *Ann. Appl. Probab.* 25 (2015), pp. 1616–1649.

[SS18] Sturm, A. and Swart, J. M. Pathwise duals of monotone and additive Markov processes. *J. Theor. Probab.* 31 (2018), pp. 932–983.

[Tay07] Taylor, J. E. The common ancestor process for a Wright-Fisher diffusion. *Electron. J. Probab.* 12 (2007), no. 28, 808–847.

[TM74] Thompson, C. J. and McBride, J. L. On Eigen's theory of the self-organization of matter and the evolution of biological macromolecules. *Math. Biosci.* 21 (1974), pp. 127–142.

[Wak09] Wakeley, J. *Coalescent Theory: An Introduction.* Roberts & Co, Greenwood Village, (2009).

[Wri31] Wright, S. Evolution in Mendelian populations. *Genetics* 16 (1931), pp. 97–159.

[Wri49] Wright, S. Adaptation and selection. In: *Genetics, Paleontology, and Evolution.* Ed. by Jepson, G. L., Simpson, G. G. and Mayr, E. Princeton Univ. Press, Princeton 1949, pp. 365–389.

[Zan+17] Zanini, F., Puller, V., Brodin, J., Albert, J., and Neher, R. *In vivo* mutation rates and the landscape of fitness costs of HIV-1. *Virus Evolution* 3 (2017), vex003.