Evolution of Highly Fecund Organisms

Luke Rex Miller

Department of Statistics

University of Oxford

A thesis submitted for the degree of

Doctor of Philosophy

Hilary Term 2015

Statement of Authorship

I declare that the material in this thesis is my own work, with the exception of the contents of Section 3.5, which is a collaborative effort between myself and Helmut Pitters, a fellow D.Phil. candidate at the University of Oxford.

Acknowledgements

First and foremost, I would like to thank my supervisor, Alison Etheridge, for her support, faith, kindness and brilliant intuition, which have guided me through the doctorate process to this point. Her boundless patience allowed her to cope with the seemingly endless lists of questions that I brought to our weekly meetings; her answers taught me a lot.

I would also like to extend my thanks to Helmut Pitters and Bob Griffiths, with whom I have had many fruitful meetings and learning experiences. The few discussions that I had with Matthias Birkner and Tom Kurtz were enlightening, and I am grateful for those. I extend my appreciation to Christina Goldschmidt and James Martin for providing helpful comments during my transfer and confirmation of status examinations. Finally, my studies would not have been possible without the funding body EPSRC, to whom I am very grateful.

I would not have been able to study for this doctorate degree without those figures who shaped my earlier academic career at Southend High School for Boys and the University of Leeds. Out of all my teachers, I would particularly like to thank Edward Mohan, Stuart Barber and Evy Kersalé.

Then the people who were around me on a daily basis. I thank my office mates, Daniel, Anna, Dominic, Sarah, Mitch and Imre, who were always willing to act as a backboard for stupid questions and silly jokes. The necessary and enjoyable tea breaks served as useful distractions. Thanks to DW, an honorary member of our office, for the stimulating conversation, and for the Frisbee. Thanks to Hannah, James, Eleanor, Rey, Henry, Mireille, Jamie, JP, Claire, Jin, Ali, Charlotte and other members of OPIG past and present for supplying a constant stream of lunchtime laughter, birthday cake and answered crossword clues. From

the wider community at the Department of Statistics, I would like to thank Nick Altemose, Jerome Kelleher, Cyril Labbé, Habib Saadi, Mike Salter-Townshend, Alex Bolton and Ella Kaye. A treasured memory of mine will be the Tuesday coffee mornings, and the delicious cookies that they brought. For those, I am indebted to Beverley Lane.

I would like to acknowledge and appreciate those friends who helped to take my mind off work during social and sporting events: Imogen, Stephen, Pedro, Alex, Jed, Mo, Aisling, Laura and Arnie – your friendship undoubtedly made the work much easier. Thanks to Exeter College and the MCR for providing a thriving social atmosphere and an unconditional friendship group. From the friends who I have made before my time in Oxford, I would like to express my thanks in particular to Amanda, Hannah Gillen, John, Peter, Hannah Beirne, Emmap, Chris, Cat, Sarah and the friendly bunch who lived with me at 52C.

I would like to give special thanks to my family, especially Mum, Dad, Ben, Lauren, Chris, Stanley, Nan and Grandad, who have always been there for encouragement. To Maren, I would like to say thank you for the support, the advice, the love, the adventure and, most importantly, the balloon fights.

Abstract

We develop and study the high-density limit of various new models in mathematical population genetics. These models extend the Λ-Fleming–Viot process when there are two genetic types at the locus of study. Given a finite sample from a population undergoing these dynamics, a key tool for understanding the corresponding genealogy is the method of duality.

We introduce the reproduction-linked mutation mechanism and consider how this affects the process of relative allelic frequencies and the genealogy.

The second generalization incorporates two forms of natural selection – differential killing and differential birth. We contrast the structure of their genealogies.

Several properties of the block size spectra of the Kingman and Beta coalescents are also investigated, including their behaviour as they come down from infinity.

Contents

List of Figures

CHAPTER 1

Introduction

The study of mathematical population genetics focuses on understanding the major evolutionary forces that drive patterns of genetic variation in a set of organisms. It is a challenging subject for two reasons. Firstly, there is a vast number of variables which one could conceivably think of that might affect the mating behaviour of the organism(s) being studied. For example, disease is certainly something which affects the fitness of the carrier – perhaps through inability, the individual cannot reproduce as effectively as if it were not blighted. One need look no further than the catalogue of illnesses which affect humans, each to a varying degree of prevalence and severity, to see that modelling even this aspect is a formidable test. Then there are, of course, many other significant factors, such as climate, other organisms (predators, for example), and even evolution itself, which could influence the overall success of the

population of interest.

In order to have a hope of finding a meaningful model, one must be willing to take on some simplifying assumptions, and concentrate on the most significant features. To that end, the inclusion of some sort of *randomness* is a method which is very often employed to deal with the fine details of the model. Much of the early modelling effort, pioneered by John Haldane, Sewall Wright and Ronald Fisher in the first half of the twentieth century, assumed that the population is in a state of *panmixia* – also known as random mating – whereby each individual is able to mate with any other. This asserts that geographical structure is non-existent or not important which, whilst an oversimplification in many situations, has been validated in many others. The point is that much of the unimportant demography is hidden away inside the randomness. There have, of course, been a number of studies exploring the mathematics of populations living with geographical structure, such as Wright [1943], Kimura [1953] and Barton *et al.* [2010]. For the majority of this thesis, we shall consider models that do not have a spatial component, though we do briefly explore the works of the latter authors in Chapter 3.

Perhaps the simplest and most familiar model in population genetics that assumes panmixia, named after two of the aforementioned pioneers of the subject, is that of the Wright–Fisher model. Roughly speaking, it assumes that the population is of constant size, evolves randomly in discrete generations, and the gene of an individual in one generation is 'chosen' from the *entire* pool of genes of the parents in the previous generation. This random process is Markovian, and will be defined more carefully in Chapter 2. In its most basic form, this model has further simplifying assumptions: firstly, each individual inherits exactly one gene because the population is haploid – that is, they possess exactly one copy of each chromosome; secondly, the process is defined only for a single locus, which is a short region of the genome encoding some sequence of base pairs or a particular gene. Furthermore, it is assumed that there exist exactly two alleles which can reside at the locus in question – this has the benefit of making the associated mathematics much simpler, as the resultant processes are one-dimensional.

The outlined assumptions, whilst unrealistic in many cases, will be used in each of the models that we study in this exposition. We do this in the knowledge that some of them can be easily be adapted to more general situations, and in the hope that we can study new phenomena without unnecessary complications.

The human population, in which each individual possesses two copies of each chromosome, is diploid, not haploid. In the Wright–Fisher setting, this can be corrected by simply doubling the population size in the haploid model, for a hermaphroditic population at least (the effects of gender subdivision are also rectifiable – see Wakeley [2009], Section 6.1).

Often there exist many alleles that befit the locus of interest. After many decades of research, the scientific field can now deal with an arbitrary number of allelic types, even infinitely many. This is to accommodate the fact that a locus may correspond to a region of the genome which is thousands of base pairs long, and could reasonably receive a large number of mutations, thereby creating a complicated set of alleles. In fact, the 'infinitely many alleles assumption' states that, whenever a mutation occurs, it creates an allele never before seen in the population. Tracking such a large number of potential genetic types requires the machinery of measure-valued processes as demonstrated, for example, by Donnelly and Kurtz [1999]. The idea of using a small number of allelic types emerged before the advent of large-scale genetic sequencing, at a time when only phenotypical traits were observable. Retrospectively, it is the case that many of the mutations that do arise yield very few functional differences, thus many of these synonymous mutations map into the same phenotypical classes. It is with this in mind that we restrict our attention to biallelic populations for the remainder of this thesis, though it should be a somewhat straightforward matter to generalize our models to allow for an arbitrary but finite number of alleles (except, perhaps, for models with selection, the subject of Chapter 5). One further scenario for which a two-type regime is appropriate is when there is one genetic type that carries an advantage over all the remaining types; in this case, the beneficial allele can be placed into one class, and the rest can be amalgamated into the other.

The Wright–Fisher model, with all the simplifying assumptions previously discussed, does highlight the role of *genetic drift*, which is the randomness due to reproduction in a finite population. Much of the demographic, social and geographical structure is captured by this one important and easy-to-model phenomenon. Just how important it is has been the subject of a long-standing debate between two schools of thought. The neutral theory, proposed by Kimura [1985], states that, with the help of mutations, genetic drift is the main driver of observable genetic variation. This theory opposed the much

earlier ideas of Darwin [1859], who suggested that certain genes provided the carrier with a fitness, which made it more likely to reproduce than those with lower fitness. This is the famous theory of *natural selection*. There has been much opposition to the neutral theory, with perhaps the most famous retort coming from Gillespie [1984].

The early version of the Wright–Fisher model has since been modified to include other important evolutionary forces. Of these, perhaps *mutation* is the most fundamental, as without it, there could be no adaptation. Genome duplication is not a perfect process and hence, every so often, an offspring will receive a different copy of an allele to that of its parent(s). This mutation event, which is random in nature, can be seen as an exploration of the space of possible genomes, an opportunity to improve the probability of survival of the carrier and its descendants. Of course, there remains the fact that if you make a random change to a machine, you are very unlikely to improve it. For this reason, the effect of most mutations may be unnoticeable or worse, disabling. In the latter case, such a mutation is unlikely to be passed on to subsequent generations. For modelling purposes in a biallelic context, mutation is incorporated by assuming that an offspring may mutate to the genetic type not of the parent with a very small probability. That is, mutation is a very slow process. We shall study it further in Chapters 2 and 4.

Selection, the notion that one allele may be more likely to succeed than the other, can also be incorporated quite easily into models such as that of Wright and Fisher. We shall study the effects of selection further in Chapter 5, but in the context of a more complicated population process.

The assumption that our genetic locus of study evolves independently of others in the genome is quite a heavy one, one that cannot be overcome so easily. Whole copies of chromosomes are passed from one generation to the next, except for the effect of *recombination*, which jumbles up the genetic material of the parent(s). Generally speaking, however, there is a positive probability that two adjacent loci on the genome may continue to be neighbours even after recombination takes place. For this reason, the independence of our locus to all others in the genome may be questionable. For example, suppose that there is a mutation which arises in the neighbourhood of our locus which boosts the fitness of the carrier. It follows that the carrier is more likely to reproduce, and therefore more likely to pass on both the advantageous gene *and* its neighbours, one of which we are interested in. Just how likely our gene is

to benefit depends on how tightly linked it is to the advantageous allele, or the strength of recombination between the two genes. This effect is known as *genetic hitchhiking*, and is addressed by Gillespie [2000] and in the 'hitchhiker's guide to the genealogy' of Berestycki [2009]. The effects of recombination provide a very interesting avenue of research, but this is not the focus of this thesis.

The second reason that mathematical population genetics is a difficult topic is that most of the data is missing. That is, we only have access to genomic data at the present time. The fact of the matter is that evolution occurs on such a slow time scale that recent advances in sequencing technology essentially only capture a snapshot of the evolutionary process as a whole. For example, the generation time of humans is approximately thirty years, and since whole-genome sequencing has only been around for fifteen years, we can have access to at most three or four generations. This is a mere drop in the ocean compared to the number of generations that have elapsed since the speciation of humans, some 200,000 years ago. However, the generation time strongly depends on the organism, and for some species of bacteria and virus, time effectively runs much faster.

For the analysis of populations of a fixed size, denote the number of individuals by $N \in \mathbb{N}$. One common approach employed by population geneticists is that of approximation by (effective) population size. The idea is to find some model which approximates the discrete version. Obtaining the limit often requires a rescaling, but since the population size is typically large, the hope is that the discrepancy between the two versions is small. By looking at, say, the *relative* frequencies – relative to N – of a certain allele in the population, rather than the absolute frequencies, there is a hope that the process will converge as N tends to infinity. (We will often use 'proportion' as a synonym for 'relative frequency'.) To achieve a non-trivial limit, one also has to speed up time by a factor of N, but this makes intuitive sense as evolution runs on a slow time scale. We will discuss this rescaling paradigm for the Wright–Fisher model in more detail in Chapter 2, and state that the limit is given by a continuous diffusion (the aptly-named Wright–Fisher diffusion). The limit of the individual-based models need not always be continuous, however, as is the case for the class of Λ-Fleming–Viot models that we introduce in Chapter 3. The use of the infinite population models will play a central role in this thesis.

For a long time, much of the attention of the field was focused on forward-in-time properties of the

population models. Despite the early work of Wright [1943] and Malécot [1948] on the probability of identity by descent, it was not until the seminal paper of Kingman [1982] that coalescent theory really got going. Suppose that we select $k \in \mathbb{N}$ individuals uniformly without replacement from the current generation of a population of size N undergoing the dynamics of the Wright–Fisher model. This is a *sample* of size k. If we think of the time at which this sample was taken as the present, then we must look *backwards* in time to discern if there are any ancestral relationships between the members of the sample. (The actual process of looking into the past is, for the most part, not possible – it is just a mathematical abstraction. That the coalescent is random represents the fact that it cannot be known exactly.) It is certainly true that each sampled individual at the present time has exactly one parent in the preceding generation, which is in turn true for each of their parents, and so on. This introduces the abstract notion of a *lineage*: a path tracing the genetic material of a sampled individual into the past via its ancestors. What is of interest to population geneticists is how these lineages interact with one another. Each sampled individual (or ancestor) in a given generation has one parent, and very often, but not always, it will have siblings living alongside it, parented by the same individual in the previous generation. When that sibling (or perhaps siblings) is also in the sample, a *coalescence* of the corresponding lineages occurs because the genes carried by those individuals were necessarily transmitted by the same parent. One can envisage this process as a tree of branches, where the branches merge at coalescence events.

Kingman's k-coalescent serves as very good approximation to the behaviour of the gene *genealogy* of this k-sample, provided k is far smaller than N (otherwise we shall observe the distortion that Bhaskar *et al.* [2014] warn of). In fact, as $N \to \infty$, the approximation becomes exact. Omitting some details, the k-coalescent can be stated quite simply in words: it is a continuous-time Markov process on the partitions of $[k]$, in which two sets in the current partition may merge together at rate 1. If we initiate the process from the partition of singletons, then each set (or block) represents a lineage in the genealogy, and its elements represent descendant lineages. The merging of two blocks is a coalescence. Notice that only pairwise mergers are allowed despite the fact that, for finite N, three or more sampled individuals have the same parent with positive probability. However, the consequence of the rescaling is such that the probabilities of coalescences of this kind are vanishingly small, and pairwise mergers dominate in the

limit. (The exact mathematical details of this phenomenon, and further properties of the k-coalescent, are explained more carefully by Wakeley [2009].) This happens provided that the variance of the number of offspring that a parent can produce in the Wright–Fisher model is not too large. Weakening this assumption leads to coalescent processes in which mergers of size three and higher are possible – the so-called Λ-coalescents, and variants thereof, will feature strongly throughout Chapters 3-5.

One important observation is that, provided that time extends far enough back into the past, all the individuals in any sample will have originated from just one ancestor, the *most recent common ancestor* (MRCA). Exactly who this individual was and when they lived will depend both on the size and composition of the sample, and the realization of the gene genealogy. To emphasize this point, we shall briefly outline an example. Suppose that you have a sibling, and we concentrate on a specific region of mitochondrial deoxyribonucleic acid (DNA). The MRCA of a sample consisting of you and your sibling is your mother. The MRCA of you and your cousin, should you have one, would be your mother's mother (as mitochondrial DNA is transmitted by the maternal parent). However, the chance of taking a random sample of size two from the human population and obtaining two close relatives is very small. What is much more likely is that you get picked alongside a citizen of China, and in this scenario you would have to trace the lineages a very long way back, tens of thousands of years, before they are likely to coalesce. Whether this coalescent is an appropriate model for the entire human population is another question entirely. Properties of the Kingman coalescent are well-studied, and hence much is known about its MRCA – see Etheridge [2011], Section 2.6, for a brief survey.

Closely linked to the genealogy is the concept of *duality*. This notion is a rather general one which connects two Markov processes together through a set of duality functions. It is quite commonly applied to problems in population genetics in the hope that properties of complicated forward-in-time processes can be expressed more simply by its relation to the dual process, if one exists. It is in fact the case that the Wright–Fisher diffusion is dual to the Kingman coalescent, as we shall demonstrate in Section 2.3. This is just one example in which the genealogy can be thought of as a dual process. This particular duality relationship has been exploited by Ethier and Griffiths [1993] where they show that the transition distribution of the Wright–Fisher diffusion can be expressed as a mixture distribution, where the mixing

weights are the transition rates of the Kingman coalescent. The context of that paper is much more general; the statement follows as a corollary. (Results about the transition distribution were known much earlier, even as early as Malécot [1948], but only in the 1993 paper was the connection to duality made explicit).

One of the aims of this thesis is to explore the link between duality and the genealogy, which we address in Chapters 2, 4 and 5. A detailed and technical account of duality can be found in Ethier and Kurtz [1986].

We would like to understand the behaviour of the genealogy of a *typed* sample, that is when each of the allelic types of the sampled individuals is known. The importance of this problem is justified by the fact that genomic data is naturally presented in this form. Specifying the composition of the sample gives some information about the how the population must have unfolded forwards in time – for example, if your sample consists entirely of one type, then you know that this type must have high frequency in the current population, and in previous generations, to some extent. Therefore, any method that you use to study the genealogy of a typed sample is going to have some dependence on the forwards-in-time process.

In order to make sense of this conditioning, it is very commonly assumed that there is a mutation process operating in the population, which induces a stationary distribution on the frequencies of types. For the two-type regime of interest to us, it is enough to assume that each individual has a positive probability to mutate to the type which its parent does not have (which we often refer to as 'the other type'). A second simplifying assumption is that the MRCA was alive when the population had reached this stationarity, and so its type is a random draw from the stationary distribution. If there were no mutation process then, by the second assumption, there would be no genetic variation during the generation of the MRCA, and certainly no further mutations could occur between then and the present day. Hence, the only samples compatible with this model would be those consisting of entirely one type.

We shall now describe in broad terms three simulation paradigms under which one can study the typed genealogy.

(1) Rejection sampling. Generate a realization of the coalescent process up to the MRCA *without* knowledge of the types: this step is often quite simple. The mutation process is usually superimposed afterwards by running a Poisson point process along each of the lineages in the generated tree (though this does depend on the model) – see Chapter 2 for details. Choose the allelic type of the MRCA by drawing once from the stationary distribution of types. Once this is decided, traverse back down the tree from the root to the leaves, marking the type of each lineage, which only changes through a mutation event. This procedure will provide a sample configuration which may or, more often than not, may not match the data. We reject the realization of the genealogy if there is not a match, but otherwise accept it. This method is deeply inefficient since, if the size of the sample is quite large, the rejection rate will be very high. In addition, it requires being able to sample from the stationary distribution, which might not always be known.

(2) Coalescent in a random background. The method of Barton *et al.* [2004] outlines a coalescent process which is conditioned on knowing the types in the sample – that is, it is possible to specify the numbers of both alleles in the sample before running the process, which effectively solves the rejection problem of (1). The rates of transitions in this coalescent are conditional on a (time-reversed) path of the forwards-in-time process of allele frequencies. Their result shows that (a) only lineages of the same type can coalesce together pairwise, (b) such a coalescence will happen *faster* when there are *fewer* individuals of that type in the population and (c) lineages may switch between backgrounds through mutation events. Unfortunately, obtaining a time-reversed path may not be that simple in a more general setting, and so it is not desirable to have such a strong dependence on the allele frequencies. Note, however, that there is no explicit dependence on any detail of the stationary distribution.

(3) Conditioned genealogy. Stephens and Donnelly [2003] outline an algorithm for simulating the conditioned genealogy (in which the types in the sample are known) without knowledge of the forward-in-time allele frequency process. The rates of the transitions in their process only depend on the *moments* of the stationary distribution. The order of the moment will depend on the state

of the coalescent at that time. In fact, since these moments only ever appear as ratios (Bayes factors), one need only know them up to a constant, provided that constant is independent of the numbers of lineages. Conditional on having these moments available, this method of simulating the genealogy is very efficient indeed. We shall use this paradigm (but often in the context of duality – see Barbour *et al.* [2000] or Chapter 2) frequently throughout this thesis, for a variety of models.

Like the Wright–Fisher diffusion, the class of population models known as the Λ-Fleming–Viot processes, introduced by Bertoin and Le Gall [2003], provides high-population limits of the allele frequencies of certain individual-based models. They are, however, discontinuous in nature, being pure-jump Markov processes. We shall not discuss the details of the model until Chapter 3, but we shall say a little about their motivation. The random discontinuities in the process represent very large changes in the population, large enough to affect a significant fraction of the individuals. This could be interpreted in several ways. For example, it has been observed that some marine species, such as Pacific oysters, exhibit a reproductive behaviour that Hedgecock [1994] refers to as 'reproduction sweepstakes', in which parents either give birth to one individual, or a brood which is on the order of the population size (this typically happens with low probability). The reproductive patterns of this highly fecund organism could reasonably be modelled by a particular Λ-Fleming–Viot process, and this is the subject of Eldon and Wakeley [2006]. The authors also discuss details of the convergence of an individual-based model to the limiting process (in a special case).

An alternative scenario is to posit that, every so often, the population experiences natural disasters, such as disease or drought, which affects a significant portion of the population. Shortly after the disaster, the genetic landscape would change dramatically from its previous state as the population regains normality. The size of the discontinuity between the allele frequencies before and after the event would somehow represent its severity or magnitude. Assuming that the frequency of such events is very low, we can model the extinction and subsequent recolonization as instantaneous.

The Fleming–Viot process, which we shall not discuss here, is a generalization of the Wright–Fisher diffusion to the space of measures on an arbitrary type space. The infinitely many alleles assumption

that we spoke of earlier could be handled in this setting. The Λ-Fleming–Viot process, or generalized Fleming–Viot process as it is sometimes known, is a further extension to include discontinuities. The appendage of Λ comes as a result of its association to the Λ-coalescent, which we shall discuss shortly. Despite the fact that there exists machinery to deal with an arbitrary number of types, we shall again focus our attention to the more simplistic two-type case.

The Λ-Fleming–Viot process (Chapter 3), and extensions thereof (Chapters 4 and 5), that we consider in this thesis will all assume panmixia – the lack of significant geographical structure. We shall note briefly, however, that there is a significant body of work devoted to studying populations which undergo these dynamics whilst residing on spatial continua. Such models are naturally called *spatial* Λ-Fleming–Viot processes, and are reviewed by Barton *et al.* [2013].

In contrast to the Kingman coalescent, the Λ-coalescent was studied *before* the corresponding forwards-in-time population model was defined. This general class of coalescent processes was discovered independently by Pitman [1999], Sagitov [1999] and Donnelly and Kurtz [1999], and its duality with the Λ-Fleming–Viot process was proved by Bertoin and Le Gall [2003]. The prominent feature is that, whilst pairwise mergers are still possible, now a coalescence of more than two lineages is permissible. The measure Λ, as well as being responsible for the name of this process, also dictates the rates of the multiple merger events. There is a constraint on Λ that prevents certain pathologies such as simultaneous multiple mergers (dealt with by the Ξ-coalescent of Schweinsberg [2000]), namely that it must be a finite measure; we will discuss this condition further in Chapter 3. The class of Λ-coalescents is broad and flexible enough to include other important processes, such as the Kingman coalescent and the family of Beta coalescents.

From our intuitive explanation of the Λ-Fleming–Viot process, a coalescent process with multiple mergers should be a biologically-reasonable candidate for the genealogy: in the extinction-recolonization scenario, we might expect that a large number of offspring might be parented by a single individual as strong newborns take advantage of the void left by the disaster. This might apply particularly well to certain species of tree, for example. Ordinarily, the trees in question would be in fierce competition for resources – space, light, nutrients – but after a severe storm or forest fire, the damage may significantly

reduce the struggle, and many seeds from a single parent could conceivably root in the empty patch and survive.

We should point out that it is possible to superimpose an ordinary reproduction mechanism onto the rare sequence of extreme events. In the literature, this has come to be known as a 'Kingman component' of a population process. Forwards in time, it would add a diffusive component corresponding to genetic drift; backwards in time, there would be an extra Poisson point process of pairwise mergers, making them more common than they would otherwise be.

Mutations drive genetic diversity and are necessary for evolution to occur. However, they cannot appear too fast, otherwise DNA replication would be too unstable. That is why mutation rates are very small in nature – on the order of 10^{-10} per base pair per replication in humans and other eukaryotic species. Accordingly, it is modelled as a very weak force: in the Wright–Fisher model, it must be on the order of N^{-1}, proportional to the inverse of the population size, in order to yield a sensible limit as $N \to \infty$. Upon doing so, one can incorporate mutation into the limiting genealogy (the Kingman coalescent) by superimposing independent Poisson point processes along the lineages, the points of which correspond to mutation events. Whenever mutation appears in this form, we shall refer to it as *weak* mutation. These issues are discussed further in Chapter 2.

In Chapter 4, we shall study a different infinite population model in which mutation occurs *during* the large-scale events that lead to the discontinuities in the Λ-Fleming–Viot process, i.e. mutation is linked to reproduction. We shall refer to this mechanism as *strong* mutation. Very roughly speaking, the class of Λ-coalescents can be split into two parts – those for which $\Lambda(du)/u$ is a finite measure, and those for which it is not. In the former case, very large mergers are commonplace. We shall show that, in order for this model to be well-defined, we must restrict the class of Λ to those for which very large mergers occur in the Λ-coalescent. We shall not focus on the details of the individual-based model that would lead to the limit that we study (except for a brief discussion), but it would be an interesting open problem to understand these details, to see if a stronger mutation rate is required to obtain such a limit, and to evaluate the biological implications that follow.

Natural selection has, for a long time, been established as the evolutionary force that drives pop-

ulations toward their optimal existence given the environment in which they reside. The idea is quite simple: whilst most mutations will be non-beneficial or even deleterious, every so often there will be one that arises that provides the carrier with a higher probability of surviving or reproducing. If this mutation enjoys some early success, then it will quickly spread through the population until most, if not all, individuals possess the allele. (Quickly, in this context, refers to hundreds of generations.)

From a mathematical standpoint, selection introduces many complications that do not exist in the neutral theory. For example, *genic selection* is perhaps the most simple formulation of this concept, which incorporates selection by superimposing an extra Poisson point process of events onto, say, the Moran [1958] model, whose points provide events in which individuals with the favoured allele outcompete those without it. Krone and Neuhauser [1997] studied the genealogy of the diffusive limit of the Moran model with genic selection and called it the *ancestral selection graph* (ASG). This object, again a dual process, essentially generalizes the Kingman coalescent, but the presence of selection makes it somewhat more complicated. For example, not only does coalescence occur, but also branching, at the behest of the reversed point process of selective events. The ancestry of a k-sample is no longer represented by the entire ASG, but rather by a tree that is contained within it. The lineages can be classed as either *real* or *virtual*, where those of the former type really do represent the true genealogy, and those of the latter constitute a mechanism to keep track of unresolved selective events. As a consequence, the first time that the ASG contains exactly one lineage is not necessarily the MRCA; instead it is the so-called ultimate ancestor.

Viability and fecundity selection are yet more complicated forms of natural selection. They can be incorporated quite naturally into the framework of the Λ-Fleming–Viot process. The former represents a difference in the probability that individuals would be killed during a natural disaster or the extinction phase, caused by resistance to a certain disease or plague, for example. The latter represents a difference in the probability of reproducing during the recolonization phase – for example, some pyrophilic species of plant would fare better after a forest fire. These selective mechanisms turn out to be closely linked, and we shall explore this further in Chapter 5.

The outline of this thesis is as follows. In Chapter 2, we define duality, the machinery that will help us to study genealogical processes in a variety of situations. Section 2.1 will host these definitions along with a discussion of how the duality between two processes may be inferred by establishing a relationship between their infinitesimal generators. Section 2.2 will focus on some important special cases under assumption (2.11, p.24), a condition on the generators which is often satisfied. Of particular interest is the normalized duality derived by Barbour *et al.* [2000], which is the duality argument related to paradigm (3) above that will be used frequently in the subsequent chapters. We shall take an in-depth look at how duality can be applied to population genetics in Section 2.3, through the Wright–Fisher model (with and without weak mutation), its limiting diffusion and genealogy (the Kingman coalescent). This example will be discussed from both mathematical and biological standpoints.

In Chapter 3, we formally introduce the Λ-Fleming–Viot process, defining both the measure-valued and two-type versions in Section 3.1. In order to facilitate the intuition behind the model, we describe one possible pre-limiting version of this infinite population limit in Section 3.2. We scratch the surface of a large body of work in Section 3.3 by briefly outlining the spatial Λ-Fleming–Viot process, and examples thereof. Section 3.4 provides definitions, properties and examples relating to the aforementioned Λ-coalescent. In Section 3.5 we consider various problems associated to the *block size spectrum*, which is vector-valued continuous-time process that tracks the number of blocks of any given size in a Λ-coalescent. Such an object offers more-refined information than the block-counting process of a coalescent, which tracks only the total number of blocks. We begin by deriving a formula for the falling factorial moments of the n-coalescent's block size spectrum in Section 3.5.1. This novel result is described in Theorem 3.10 and Lemma 3.12. The Smoluchowski coagulation equations [Smoluchowski, 1916], with a constant kernel, can be seen as the limiting version of the rescaled block size spectrum of Kingman's coalescent at small times and will be discussed in Section 3.5.2. Whilst this deterministic diffusion limit may be of little interest to population geneticists (it is rare and ill-advised to take a large, let alone infinite, sample size), it may be of interest to mathematicians. We use this to motivate Section 3.5.3, an investigation into similar small-time behaviour of the (generalized) Beta coalescent. Equation (3.39, p.74) is a generalization of the Smoluchowski differential equation for this particular multiple-

merger coalescent; its solution is given by a complete Bell polynomial in Theorem 3.18. This result only applies to those Beta coalescents which come down from infinity. This is joint research with Helmut Pitters.

Chapter 4 investigates a new modification of the Λ-Fleming–Viot process to include a strong form of mutation. This model, defined in Section 4.1, links the mutation and reproduction mechanisms together, in contrast to weak mutation models. We have to impose stronger conditions on Λ – namely that $\Lambda(du)/u$ be a finite measure – in order for the process to be well-defined. Though we shall focus more on the properties of the infinite population limit, Section 4.2 outlines one potential individual-based model which may serve as a pre-limit. In Section 4.3, and in particular Proposition 4.3, we show that, through a certain transformation, the generator of the Λ-Fleming–Viot process with strong mutation can be decomposed into two other linear operators, one of which is precisely the generator of the ordinary Λ-Fleming–Viot process (the other is not a generator and deals with mutation). It turns out that this result is key to understanding the stationary distribution of the process when specializing to the Beta coalescent. This case is dealt with explicitly in Section 4.4 by adapting the method of Handa [2014]: a linear second order differential equation can be derived, and solved, for the generalized Stieltjes transform of the density function of the stationary distribution. Whilst the author has made limited progress on the inversion problem (Section 4.4.3), the density function in Stieltjes space does allow one to find an explicit recursion for the moments of the stationary distribution – see Section 4.4.2. Section 4.5 investigates the unconditioned (paradigm (1)) and conditioned (paradigm (3)) genealogy of the Λ-Fleming–Viot process with strong mutation, focusing on the latter. This discussion takes on a probabilistic angle, with intuitive explanations of the rates in the backwards-in-time jump process, before using the duality arguments of Barbour *et al.* [2000] to fully characterize the genealogy in Theorem 4.7. One interesting property of this genealogy is that, due to the mutation mechanism, multiple lineages of *different* types can merge into one. We remark that, since we have derived a recursion for the moments of the stationary distribution of the Beta coalescent, this theorem could easily be applied to simulate realizations of the conditioned genealogy. Some generalizations of the model are suggested in Section 4.6 – it may be possible to obtain the full range of Λ-coalescents by allowing the mutation rates to be random rather than fixed.

Chapter 5 discusses two further modifications of the Λ-Fleming–Viot process to include two forms of reproduction-linked selection: the first is a viability selection which we refer to as 'differential killing' and define in Section 5.1; the second incorporates fecundity selection which we refer to as 'differential birth' and define in Section 5.2. These models can be defined in some generality, but for most of the chapter we shall focus on a specialized version of the processes. The first reason for doing so is that the specialization has a rather intuitive interpretation in terms of a Poisson point process of events which throws down 'strong' and 'weak' circles. It brings the model more into align with the ordinary Λ-Fleming–Viot process and introduces one important parameter, $r \in (0, 1]$, which measures the selective difference in these differential killing and birth models. As in Chapter 4, we have to restrict the class of measures for which these models work. In order to study the genealogies under paradigm (3) above using the methods of Chapter 2, we introduce an additional, independent weak mutation mechanism to induce a stationary distribution (Section 5.4). In Section 5.3, we prove that the Λ-Fleming–Viot processes with differential killing/birth and weak mutation have the very same process of allele frequencies and therefore the same generator. We shall study the *conditioned* genealogies of these models, their similarities and differences, in some detail in Section 5.5. We define the *simple* genealogy to be the backwards-in-time process which describes the number of lineages of each type for a sample in which the types are known. The simple genealogies of these two selective models (Section 5.5.2) are precisely the same when they are coupled in the natural way. They are complex branching-coalescing systems in which multiple mergers are permissible, and are characterized through Theorem 5.3, Proposition 5.5 and Lemma 5.6. In Section 5.5.3, we show that the *real* genealogies, those which keep track of both the real and virtual lineages of each type, do in fact differ between the models due to the presence of 'silent events', which cannot be detected by the simple genealogy. This has the consequence that, in this case, duality has failed to truly capture the genealogical process of the sample. (A similar observation was noted by Taylor [2009], but for a modified Wright–Fisher diffusion model.) Theorems 5.10 and 5.11 provide characterizations of the real genealogies for the two processes. In principle, provided one could derive the moments of the corresponding stationary distribution, these results would allow one to simulate properties of the time to most recent common ancestor in the two models, and numerically

check if they are equal in distribution. This is not possible using the simple genealogy alone. Section 5.6 suggests an alternative model which interpolates between the differential killing and birth models, and is free of silent events. Finally, we remark that, as a corollary of everything in this chapter, results concerning the neutral version, that is the Λ-Fleming–Viot process with weak mutation, are implicitly available by removing selection (which amounts to setting $r = 1$).

Appendix A provides a summary of the notational conventions that are used throughout this thesis. Appendix B gives a brief outline of some of the preliminary topics that will be referred to without proof. Finally, Appendix C outlines the statements and proofs of the more technical lemmas required by the arguments that follow, ordered by chapter number.

CHAPTER 2

Duality

Duality is a powerful concept which connects two Markov processes through a triple of functions. This connection can be extremely useful for making statements about one process in terms of the other, especially if one is easier to analyse, and we will often rely on it throughout this thesis. There are no hard and fast rules for concocting duality functions. Some well-known examples do exist however, like those of moment duality functions in population genetics – these will be discussed in Section 2.3.2.

The treatment of duality can be made very formal, and consequently very general. Ethier and Kurtz [1986] define duality in the context of martingale problems, and this type of thinking allows the method to be very useful there. In this exposition, we choose a different route: our goal is to be less general, but recast duality in terms of the infinitesimal generators of the processes in question, via Proposition 2.6.

The layout of this chapter is as follows. We begin with the definition of the infinitesimal generator of a time-homogeneous Markov process, the notion(s) of duality, and the link between them in Section 2.1. Proposition 2.6 will be useful for identifying special cases of duality, such as those in Section 2.2. In fact, the purpose of that section is to recognize how to form a duality should the generator of the process of interest satisfy condition (2.11, p.24). The method of Section 2.2.1 is rather important in the context of this thesis, as it will be employed several times. We will show that, for a given process, the duality functions are not necessarily unique, although they will yield different dual processes. Finally, in Section 2.3, we take an extensive look at the classical example of duality in population genetics, that of the Wright–Fisher diffusion and the Kingman coalescent. We will extend it to include weak mutation in Section 2.3.3, and see how this fits in to the framework of Section 2.2. We complete the chapter with a biological discussion of these models, interpreting duality as a genealogical relationship.

2.1 Definitions

We begin this chapter with a few necessary definitions associated with Markov processes. Many of these notions are standard – here we leave out many of the technical details which can be found, for example, in Ethier and Kurtz [1986].

Definition 2.1 (Generator)**.** Let $(X_t)_{t\geq 0}$ be a time-homogeneous Markov process on a state space S. For a function $f : S \to \mathbb{R}$ and initial value $x \in S$, we define the *infinitesimal generator* of X acting on f to be

$$\mathscr{L}f(x) = \lim_{\delta t \downarrow 0} \frac{\mathbb{E}_x[f(X_{\delta t}) - f(x)]}{\delta t}, \tag{2.1}$$

where $\mathbb{E}_x[\,\cdot\,] = \mathbb{E}[\,\cdot\,|X_0 = x]$. The *domain* of the generator, $\mathscr{D}(\mathscr{L})$, is the set of functions for which this limit exists.

Such an object often characterizes the nature of a Markov process, and we now present some common examples which will feature in the sequel.

Example 2.2 (Diffusion on an interval)**.** Let $I \subseteq \mathbb{R}$ be an interval on the real line. Let $f, \mu : I \to \mathbb{R}$ be continuous functions and $\sigma^2 : I \to \mathbb{R}_+$ be a positive continuous function (though σ may vanish at the boundaries of I). Then the differential operator

$$\mathscr{L}f(x) = \mu(x)f'(x) + \tfrac{1}{2}\sigma^2(x)f''(x) \tag{2.2}$$

corresponds to the diffusion process X on the interval I given by the stochastic differential equation (SDE)

$$dX_t = \mu(X_t)dt + \sigma(X_t)dW_t$$

where W is a one-dimensional Brownian motion. It is not difficult to extend this example to diffusions on higher-dimensional spaces – see Ethier and Kurtz [1986], p.368, for example.

Example 2.3 (Jump process)**.** Suppose that X is continuous-time Markov jump process on the countable state space K with rate matrix $Q = (q(k, k'))_{k,k' \in K}$. Then, the generator for X is given by the operator

$$\mathscr{L}f(k) = \sum_{k' \in K} q(k, k') \left(f(k') - f(k) \right). \tag{2.3}$$

So far, we have only discussed individual Markov processes; we will now explore a concept that links two Markov processes together. The notion of duality can be extremely useful in making statements about one process in terms of the other one. Before demonstrating this, we shall define it in the narrow and broad senses below.

Definition 2.4. Let $(X_t)_{t \geq 0}, (Y_t)_{t \geq 0}$ be two independent Markov processes with respect to the probability measures $\mathbb{P}^X$ and $\mathbb{P}^Y$ on the state spaces S^X and S^Y, respectively. Suppose that there is a measurable function $h(x, y)$ on the product space $S^X \times S^Y$ such that

$$\mathbb{E}_x^X[h(X_t, y)] = \mathbb{E}_y^Y[h(x, Y_t)] \tag{2.4}$$

$\forall x \in S^X,\ y \in S^Y$ and $t \geq 0$. Then, X, Y are *dual* with respect to h.

We will refer to this as an 'exact' duality, or duality 'in the narrow sense'. Whilst this definition is satisfying and useful in many cases, there is in fact a broader definition that can be utilized.

Definition 2.5 (Duality)**.** Suppose that, in addition to X, Y and h of Definition 2.4, we have the measurable functions $\alpha : S^X \to \mathbb{R}$ and $\beta : S^Y \to \mathbb{R}$ which satisfy $\int_0^t \alpha(X_s)ds < \infty$ $a.s.$ and $\int_0^t \beta(Y_s)ds < \infty$ $a.s.$ $\forall x \in S^X$, $y \in S^Y$ and $t \geq 0$. If instead of condition (2.4), we have that

$$\mathbb{E}_x^X \left[h(X_t, y) \exp\left(\int_0^t \alpha(X_s)\, ds \right) \right] = \mathbb{E}_y^Y \left[h(x, Y_t) \exp\left(\int_0^t \beta(Y_s)\, ds \right) \right], \tag{2.5}$$

then X, Y are *dual* with respect to the triple of functions (h, α, β).

It is clear that Definitions 2.4 and 2.5 are the same when $\alpha = \beta = 0$, in which case we simply replace the triple $(h, 0, 0)$ by h. We note that it is important that the expectations in (2.5), and therefore (2.4), must satisfy some boundedness conditions, namely

$$\mathbb{E}_x^X \left[\left| h(X_t, y) \exp\left(\int_0^t \alpha(X_s)\, ds \right) \right| \right] < \infty, \tag{2.6}$$

$$\mathbb{E}_y^Y \left[\left| h(x, Y_t) \exp\left(\int_0^t \beta(Y_s)\, ds \right) \right| \right] < \infty. \tag{2.7}$$

We shall briefly recall Theorem 5.1 from Varadhan [2007] which reviews the Feynman–Kac formula for Brownian motion, W_t, with initial value $W_0 = x$. Suppose that $V(x)$ is a real-valued bounded continuous function. Then,

$$u(t, x) = \mathbb{E}_x^W \left[h(W_t) \exp\left(\int_0^t V(W_s)\, ds \right) \right] \tag{2.8}$$

solves the partial differential equation

$$\frac{\partial u}{\partial t} = \frac{1}{2} \frac{\partial^2 u}{\partial x^2} + V(x) u(t, x) \tag{2.9}$$

with initial condition $u(x, 0) = h(x)$. In fact, a more general result holds for diffusions: replace the Brownian motion by the diffusion, and the Laplacian term in (2.9) by its corresponding generator given by (2.2). So when either X_t or Y_t in Definition 2.5 is a diffusion, the left- or right-hand side of (2.5) satisfies a Feynman–Kac formula. We will often be working with processes that are not diffusions, but we shall refer to duality in the broader sense (as in Definition 2.5) as *Feynman–Kac duality*, because of this close connection.

Often it is easier to work with the infinitesimal generators of processes rather than conditional expectations. There is an important connection between generators and duality, which is given by the following result.

Proposition 2.6 (Ethier and Kurtz [1986], Corollary 4.13)**.** *Suppose that for two time-homogeneous Markov processes X, Y, we have the triple of bounded functions (h, α, β) satisfying*

$$\mathscr{L}^X h(x,y) + \alpha(x)h(x,y) = \mathscr{L}^Y h(x,y) + \beta(y)h(x,y) \tag{2.10}$$

for each $x \in S^X, y \in S^Y$, where $\mathscr{L}^X$ is the generator of X acting on the first coordinate, and $\mathscr{L}^Y$ is the generator of Y acting on the second coordinate. Then X and Y are dual with respect to (h, α, β).

In summary, this proposition says that if we would like to check if two processes are dual with respect to (h, α, β), we need only inspect their generators. We will rely heavily on this result in the remainder of this exposition.

2.2 Special Cases

There are some important cases of duality that will be useful in later chapters. The two subsets that we explore here demonstrate methods of how to obtain duality from processes whose generators satisfy condition (2.11) below. As shown throughout the sequel, this will work for a large class of processes. We shall utilize the link between generators and duality given by Proposition 2.6.

2.2.1 Normalized Duality

The method that we discuss here was observed by Barbour *et al.* [2000], Section 2. Here, we shall restate the procedure in detail, but with the use of slightly different notation (which is more akin to that used by Etheridge and Griffiths [2009]).

Let S^K be some countable state space. Suppose that one can find a positive function $f : S^X \times S^K \to$

$\mathbb{R}$ such that $f(\cdot, k) \in \mathscr{D}(\mathscr{L}^X)$ for each $k \in S^K$ and, when applied to the generator for X, satisfies

$$\mathscr{L}^X f(x,k) = \sum_{k' \neq k} c(k,k') f(x,k') - c(k,k) f(x,k), \tag{2.11}$$

for all $x \in S^X$ and $k \in S^K$, where the collection $\{c(k,k') : k, k' \in S^K\}$ is a set of non-negative constants. The range of the sum is $k' \in S^K$ such that $k' \neq k$. (We have purposefully replaced y by k in the second coordinate to emphasize the discrete nature of the 'Y process'.) Suppose further that X_t admits a stationary distribution, X, for which the generalized moments

$$m_k := \mathbb{E}^X \left[f(X,k) \right], \tag{2.12}$$

exist for each $k \in S^K$. We can now take the expectation of (2.11) under the stationary distribution to obtain

$$\mathbb{E}^X \left[\mathscr{L}^X f(X,k) \right] = \sum_{k' \neq k} c(k,k') m_{k'} - c(k,k) m_k.$$

It follows from Ethier and Kurtz [1986], Proposition 4.9.2, that, so long as f is in $\mathscr{D}(\mathscr{L}^X)$ for each $k \in S^K$, the left-hand side vanishes. Thus,

$$c(k,k) = \sum_{k' \neq k} c(k,k') \frac{m_{k'}}{m_k}. \tag{2.13}$$

This equation is the key step which allows us to construct a valid rate matrix for a Markov jump process using the constants $c(\cdot,\cdot)$. For the off-diagonal terms, set

$$q(k,k') = c(k,k') \frac{m_{k'}}{m_k}, \qquad k' \neq k, \tag{2.14}$$

which are all non-negative (recall that f is positive). For the diagonal terms, we invoke (2.13) to obtain

$$q(k,k) = -\sum_{k' \neq k} q(k,k') = -\sum_{k' \neq k} c(k,k') \frac{m_{k'}}{m_k} = -c(k,k). \tag{2.15}$$

This confirms that $Q = (q(k,k'))_{k,k' \in S^K}$ is indeed a valid rate matrix for some Markov jump process since each row sums to zero, and all entries are non-negative except the diagonal ones, which are

negative. Define the normalized duality function

$$g(x,k) := \frac{f(x,k)}{m_k}. \tag{2.16}$$

We have yet to show that g is indeed a duality function, but applying it to the generator of X_t yields

$$\begin{aligned}\mathscr{L}^X g(x,k) = \frac{1}{m_k}\mathscr{L}^X f(x,k) &= \frac{1}{m_k}\sum_{k'\neq k} c(k,k')f(x,k') - \frac{1}{m_k}c(k,k)f(x,k) \\ &= \sum_{k'\neq k}\frac{1}{m_{k'}}q(k,k')f(x,k') + q(k,k)\frac{f(x,k)}{m_k} \\ &= \sum_{k'\neq k} q(k,k')\left[g(x,k') - g(x,k)\right].\end{aligned}$$

In the first line we used assumption (2.11) and the fact that m_k is a constant with respect to $\mathscr{L}^X$. In the second, we used the definition of the c's in (2.14) and (2.15), and in the final line we used the zero-row-sum property of Q and definition (2.16). We now recognize the right-hand side as the generator (in the second coordinate) of a continuous-time pure-jump Markov process with rate matrix Q, as in Example 2.3. We shall call this process K_t and its generator $\mathscr{L}^K$. As long as g is bounded on $S^X \times S^K$, we can apply Proposition 2.6 to deduce that X_t and K_t are dual with respect to the normalized duality function g, since α and β are exactly zero. That is, since $\mathscr{L}^X g(x,k) = \mathscr{L}^K g(x,k)$,

$$\mathbb{E}_x^X[g(X_t,k)] = \mathbb{E}_k^K[g(x,K_t)], \qquad \forall x \in S^X,\ k \in S^K,\ t \geq 0. \tag{2.17}$$

This is a special case of (2.4), the definition of exact duality.

Despite the fact that K_t must be a jump process, we have made no comment thus far about the nature of X_t. In fact, in Etheridge and Griffiths [2009] and the follow-up paper Etheridge *et al.* [2010], the authors give examples where X_t is either a jump process on a countable or uncountable state space, or a diffusion. We shall discuss an example of the latter in Section 2.3 through the Wright–Fisher diffusion.

This method could quite easily have been called 'duality with a stationary distribution' as there is a strong dependence on the existence of one. The form of duality in the next section, however, does not have such a dependence.

2.2.2 A Special Feynman–Kac Duality

The algebraic manipulation required to obtain this form of duality is conceptually much simpler than that of normalized duality, but later in this section we shall demonstrate a link between these two forms. For now, assume that once again we can find a function $f(x,k)$ such that (2.11) is satisfied. By a simple addition-subtraction trick,

$$\mathscr{L}^X f(x,k) = \sum_{k' \neq k} c(k,k') \left[f(x,k') - f(x,k) \right] + \left(\sum_{k' \neq k} c(k,k') - c(k,k) \right) f(x,k). \tag{2.18}$$

With a view to making this conform to (2.10), let

$$\tilde{q}(k,k') := c(k,k'), \qquad k' \neq k \tag{2.19}$$

be the off-diagonal terms of another rate matrix, $\tilde{Q}$. It follows that

$$\tilde{q}(k,k) = -\sum_{k' \neq k} \tilde{q}(k,k') = -\sum_{k' \neq k} c(k,k'), \tag{2.20}$$

by the zero-row-sum property of rate matrices. We can then rewrite (2.18) as

$$\mathscr{L}^X f(x,k) = \mathscr{L}^{\tilde{K}} f(x,k) - (\tilde{q}(k,k) + c(k,k)) f(x,k), \tag{2.21}$$

where $\tilde{K}_t$ is another pure-jump Markov process on the same space S^K with rate matrix $\tilde{Q}$. We see that this is indeed a special case of (2.10), and hence we conclude that X_t and $\tilde{K}_t$ are dual with respect to the *un*normalized duality function f, along with the auxiliary functions $\alpha = 0$ and $\tilde{\beta}(k) = -(\tilde{q}(k,k) + c(k,k))$.

To arrive at (2.21), we made no assumption about the process X_t having a stationary distribution, unlike in the previous section. This method therefore works for a larger class of processes. However, for those processes which do admit such a stationary measure, we can make a link between the normalized and unnormalized dualities. In particular, (2.15) holds, transforming (2.21) into

$$\mathscr{L}^X f(x,k) = \mathscr{L}^{\tilde{K}} f(x,k) + q_d(k) f(x,k), \tag{2.22}$$

where $q_d(k) = q(k,k) - \tilde{q}(k,k)$ and $q(k,k)$ is the diagonal entry of the rate matrix Q belonging to the process K_t from Section 2.2.1. We remark that, whilst (2.22) does depend on the existence of the stationary distribution X, it does not depend in any way upon the generalized moments given by (2.12).

The interesting observation is that such processes admit two forms of dualities: X_t is dual to K_t in the narrow sense with respect to the normalized duality function g, and to $\tilde{K}_t$ in the broad sense with respect to the functions $(f, 0, q_d)$. In fact, as we will show in Section 2.2.3, there is a rich class of dualities for X_t.

In the Section 2.3, we shall give an explicit example of what these dualities look like in relation to the Wright–Fisher diffusion.

2.2.3 Many Dualities

As mentioned previously, if we can write $\mathscr{L}^X f(x,k)$ in the form of (2.11), then duality is not unique. We will now show the connection between Sections 2.2.1 and 2.2.2, and in the process uncover a rich class of dualities to the process X_t. In addition, we shall demonstrate that the duality function of Section 2.2.1, $g(x,k)$, is a canonical choice.

Once again, assume (2.11). Then, for each $k \in S^K$, define the strictly positive function $a(k)$ which is constant in x, but evidently not in k. Using the fact that $\mathscr{L}^X$ is a linear operator,

$$\begin{aligned}
\mathscr{L}^X a(k) f(x,k) &= a(k)\mathscr{L}^X f(x,k) \\
&= \sum_{k' \neq k} a(k)c(k,k')f(x,k') - a(k)c(k,k)f(x,k) \\
&= \sum_{k' \neq k} \frac{a(k)}{a(k')} c(k,k') \left[a(k')f(x,k')\right] - a(k)c(k,k)f(x,k) \\
&= \sum_{k' \neq k} \frac{a(k)}{a(k')} c(k,k') \left[a(k')f(x,k') - a(k)f(x,k)\right] \\
&\quad + \left(\sum_{k' \neq k} \frac{a(k)}{a(k')} c(k,k') - c(k,k) \right) a(k)f(x,k) \\
&= \mathscr{L}^{\hat{K}} a(k) f(x,k) - (\hat{q}(k,k) + c(k,k)) a(k) f(x,k),
\end{aligned}$$

where $\mathscr{L}^{\hat{K}}$ is the generator of the jump process $\hat{K}_t$ with state space S^K and transition rates

$$\hat{q}(k,k') = \frac{a(k)}{a(k')}c(k,k'), \quad k' \neq k, \quad \text{and} \quad \hat{q}(k,k) = -\sum_{k' \neq k} \frac{a(k)}{a(k')}c(k,k').$$

Hence, given any set of strictly positive constants $\{a(k) : k \in S^K\}$, we can construct a duality between X_t and $\hat{K}_t$ in the broad sense (with respect to the functions $(af, 0, \hat{\beta})$ where $\hat{\beta}(k) := -(c(k,k) + \hat{q}(k,k))$) using this method.

One question we might like to answer is how could we make $\hat{\beta}(k) = 0$, and consequently find a duality in the narrow sense? If X_t admits a stationary distribution then (2.13) holds. Invoking this assumption,

$$\hat{\beta}(k) = \sum_{k' \neq k} \hat{q}(k,k') - c(k,k) = \sum_{k' \neq k} \frac{a(k)}{a(k')}c(k,k') - \sum_{k' \neq k} \frac{m_{k'}}{m_k}c(k,k') = \sum_{k' \neq k} c(k,k') \left(\frac{a(k)}{a(k')} - \frac{m_{k'}}{m_k} \right).$$

Then it is clear that we can make this vanish by setting

$$a(k) = \frac{1}{m_k} \tag{2.23}$$

for all $k \in S^K$. This is precisely the choice we made in Section 2.2.1. In fact, that section is simply a special case of this one, but it just happens to be the choice that yields duality in the narrow sense. In Section 2.2.2, $a(k) \equiv 1$.

2.3 Example: the Wright–Fisher Diffusion and Kingman Coalescent

This section aims to exemplify the notions of duality previously discussed, particularly those in Section 2.2. We shall discuss at length a canonical example from mathematical population genetics: the Wright–Fisher diffusion, both with and without weak mutation. We first take a look at the mathematics in Sections 2.3.1-2.3.3, before bringing it into a biological context in Section 2.3.4.

2.3.1 Wright–Fisher Model

In this section, we shall define and discuss the classical Wright–Fisher model, along with its infinite population limit in the absence of mutation. We shall make explicit the duality between the diffusion limit and the Kingman coalescent.

Definition 2.7. Suppose that we have $N \in \mathbb{N}$ haploid individuals in our population of interest, and each of them may carry one of two allelic types, a_1 or a_2, at the locus of study. The *Wright–Fisher model* advances in discrete generations, which we index by $t \in \mathbb{N}_0$. At generation t, define N_t^1 to be the number of individuals carrying allele a_1; similarly, define N_t^2 to be those members carrying allele a_2. The population size N does not depend on time, so $N_t^1 + N_t^2 \equiv N$, $\forall t \in \mathbb{N}_0$. The population in generation $t+1$ is constructed from the population at generation t thus: each child 'chooses' its parent independently *with* replacement from the individuals in generation t with probability $1/N$ (uniformly at random). Once all N children in generation $t+1$ have made their random selections, all the individuals from the previous generation die and the process begins anew.

The configuration of the population at some time $t+1$ is created only through the state of the population at the previous generation t. As such, the Wright–Fisher model evolves as a Markov process, and this property is also inherited by the summary processes N_t^1 and N_t^2. We are interested in the rescaling that will yield a diffusion limit for the *proportion* of type-a_1 individuals as $N \to \infty$. Define this proportion to be $X_t = \frac{N_t^1}{N} \in [0,1]$ with initial condition $X_0 = x$. Supposing that each generation actually corresponds to a time-step of length $\frac{1}{N}$, the generator of the jump chain process converges to the differential operator

$$\mathscr{L}_{WF} = \frac{1}{2}x(1-x)\frac{d^2}{dx^2}. \tag{2.24}$$

as $N \to \infty$. (The details of this calculation are reviewed by Etheridge [2011].) We can see that this is in the form of (2.2) with $\mu = 0$ and $\sigma(x) = \sqrt{x(1-x)}$, and is therefore the generator of a diffusion satisfying the SDE

$$dX_t = \sqrt{X_t(1-X_t)}\, dW_t, \quad X_0 = x, \tag{2.25}$$

where W_t is a standard Brownian motion.

2.3.2 Moment Duality

The Wright–Fisher diffusion, characterized by (2.24), exhibits a *moment duality* relationship with a jump process. This duality is realized through the moment duality function

$$h(x,k) = f_k(x) := x^k \tag{2.26}$$

in the narrow sense with $\alpha = \beta = 0$. To see this, a simple calculation reveals that

$$\mathscr{L}_{WF} f_k(x) = \binom{k}{2}(f_{k-1}(x) - f_k(x)) =: \mathscr{L}_K f_k(x), \tag{2.27}$$

$\forall x \in [0,1], k \in \mathbb{N}$ (with the convention that $\binom{1}{2} = 0$). The generator on the right-hand side, which acts on the second coordinate (now expressed as a subscript), is that of a pure-death process starting with k lineages, which decreases by one at rate $\binom{l}{2}$ when there are l lineages. In fact, it characterizes the process that captures the total number of blocks in the k-coalescent (a restriction of the Kingman coalescent). We shall refer to this as the *block-counting process*.

Because of this duality relationship, we may conclude, through Proposition 2.6, that

$$\mathbb{E}_x^X[X_t^k] = \mathbb{E}_k^K[x^{K_t}], \tag{2.28}$$

$\forall x \in [0,1], k \in \mathbb{N}, t \geq 0$, where X_t is the Wright–Fisher diffusion as above, and K_t is the block-counting process of the k-coalescent. This enables us to make statements about one process from knowledge of the other: the moments of X_t are specified by the k-coalescent for various values of k (hence the name *moment* duality); the probability generating function of K_t is specified by the Wright–Fisher diffusion with various initial conditions.

2.3.3 Adding Mutation

One feature of the Wright–Fisher diffusion in the absence of mutation is that it will hit one of the absorbing boundaries in finite time *a.s.* [Feller, 1951]. The effect of including mutation between the two types in the model changes this phenomenon radically: provided the mutation rates are strictly positive, this modification will induce a stationary distribution such that the process will never be absorbed at the

boundaries.

To incorporate this form of transverse mutation, we simply add another stage to the Wright–Fisher model of Definition 2.7. Specifically, once the child has chosen its parent, which is of type a_i, say, then it may mutate to the other type with probability $\mu_i \in (0,1)$.

In order to model *weak* mutation, one can let the mutation probabilities depend upon the population size, N. Setting $\mu_i = \frac{\nu_i}{N}$, for some $\nu_i > 0$, is the correct scaling of the mutation rates in order to see a non-trivial mutation component in the diffusion limit. Details of this calculation can be found in Feller [1951] or Karlin and Taylor [1981], Section 15.2.F. Having passed to the limit, the generator of the Wright–Fisher diffusion with weak mutation is given by

$$\mathscr{L}_{WFm} = \frac{1}{2}x(1-x)\frac{d^2}{dx^2} + (\nu_2(1-x) - \nu_1 x)\frac{d}{dx}. \tag{2.29}$$

We shall also denote this process by X_t throughout the chapter with the understanding that we recover the ordinary Wright–Fisher diffusion upon setting $\nu_1 = \nu_2 = 0$. In comparison with (2.24), the similarity between this process and the ordinary Wright–Fisher diffusion is that of the noise term. What is different is that a linear drift term has been introduced, which is the effect of the mutation. Note that the subscript 'm' will always refer to this type of mutation mechanism – in Chapter 4 we shall consider a different mutation mechanism, for which reserve the capitalized 'M'.

We now attempt to find a dual process for the Wright–Fisher diffusion with weak mutation. Using the moment dual function (2.26), we find that

$$\mathscr{L}_{WFm} f_k(x) = \left(\binom{k}{2} + k\nu_2\right)(f_{k-1}(x) - f_k(x)) - k\nu_1 f_k(x) = \mathscr{L}^{\tilde{M}} f_k(x) - k\nu_1 f_k(x). \tag{2.30}$$

After a small amount of rearranging, we see that this equation is a special case of (2.11) (and also (2.10)) with

$$c(k,k-1) = \binom{k}{2} + k\nu_2, \qquad c(k,k) = \binom{k}{2} + k(\nu_1+\nu_2),$$

and therefore fits into the framework of Section 2.2. In this instance, whilst the Feynman–Kac duality procedure of Section 2.2.2 is valid, it does not help us much as this example is rather simple – for a given k, there is only one state that the dual process, $\tilde{M}_t$, can jump to: $k-1$. In fact, we can see immediately

from (2.30) that $\tilde{q}(k,k) + c(k,k) = k\nu_1 = -\tilde{\beta}(k)$. (In more complicated situations, the method of Section 2.2.1 will be far more useful.) The dual process is similar to Kingman's coalescent, except that there is an additional $k\nu_2$ term in the rate of loss of a single lineage – compare (2.27) and (2.30). Using Proposition 2.6, we may deduce the duality

$$\mathbb{E}_x^X[X_t^k] = \mathbb{E}_k^{\tilde{M}}\left[x^{\tilde{M}_t}\exp\left(-\nu_1\int_0^t \tilde{M}_s\,ds\right)\right], \tag{2.31}$$

$\forall x \in [0,1], k \in \mathbb{N}, t \geq 0$, where X_t is the Wright–Fisher diffusion with weak mutation, and $\tilde{M}_t$ is the modified Kingman's coalescent with mutation. Comparing this expression to (2.28), we see that there is an extra exponential term inside the expectation on the right-hand side.

That there exists a stationary distribution for X_t was originally proved by Wright [1949], along with the exact form of this distribution. Stated in the following lemma, we offer an alternative proof which makes use of the method of duality.

Lemma 2.8. *Let X_t be the Wright–Fisher diffusion with weak mutation, with parameters $\nu_1, \nu_2 > 0$. Then, X_t admits a stationary distribution $X \overset{d}{=}$ Beta $(2\nu_2, 2\nu_1)$.*

Proof. Our starting point is the duality relationship that we developed in (2.31). We shall remove the superscripts from the expectations in order to simplify the notation. We would like to take limits in t, and prove that

$$\mathbb{E}[X^k] = \mathbb{E}_x\left[\lim_{t\to\infty} X_t^k\right] = \lim_{t\to\infty}\mathbb{E}_x[X_t^k].$$

At this point we make use of the duality relationship to show that

$$\lim_{t\to\infty}\mathbb{E}_x[X_t^k] = \lim_{t\to\infty}\mathbb{E}_k\left[x^{\tilde{M}_t}\exp\left(-\nu_1\int_0^t \tilde{M}_s\,ds\right)\right].$$

Since $x \in [0,1]$, $\nu_1 > 0$, and $\tilde{M}_t$ is *a.s.* non-negative, the expression inside the expectation on the right-hand side is dominated by the constant 1, so we may apply the Dominated Convergence Theorem with dominating function $g(\cdot) = 1$ to justify passing the limit inside the expectation. So long as the resulting expectation is finite, we may do the same on the left-hand side too. This will follow in the calculation to

come, and leads to

$$\mathbb{E}[X^k] = \mathbb{E}_k\left[\lim_{t\to\infty} x^{\tilde{M}_t}\exp\left(-\nu_1\int_0^t \tilde{M}_s\,ds\right)\right] = \mathbb{E}_k\left[\exp\left(-\nu_1\int_0^\infty \tilde{M}_s\,ds\right)\right].$$

It remains to show that the moments are finite and coincide with those of the Beta distribution. Define the stopping times

$$\tau_i := \inf_{t\geq 0}\{\tilde{M}_t = i\},\ \ i\in\{0,1,\ldots,k\}$$

which satisfy $0 = \tau_k < \tau_{k-1} < \ldots < \tau_1 < \tau_0 < \infty$ *a.s.* Recalling that $\tilde{M}_t = 0\ \forall t \geq \tau_0$, and that $\tilde{M}_t = i$ for $t\in[\tau_i,\tau_{i-1})$, $i = 1,\ldots,k$,

$$\begin{aligned}\mathbb{E}[X^k] &= \mathbb{E}_k\left[\exp\left(-\nu_1\left(\int_0^{\tau_0}\tilde{M}_s\,ds + \int_{\tau_0}^\infty \tilde{M}_s\,ds\right)\right)\right] = \mathbb{E}_k\left[\exp\left(-\nu_1\sum_{i=1}^k\int_{\tau_i}^{\tau_{i-1}}\tilde{M}_s\,ds\right)\right]\\ &= \mathbb{E}_k\left[\prod_{i=1}^k \exp\left(-\nu_1 i(\tau_{i-1}-\tau_i)\right)\right] = \prod_{i=1}^k \mathbb{E}\left[e^{-\nu_1 i E_i}\right],\end{aligned}$$

where the $E_i = \tau_{i-1}-\tau_i$ are independent $\text{Exp}\left(\binom{i}{2} + i\nu_2\right)$ random variables (this follows from the Markov jump nature of $\tilde{M}_t$). Using simple integration,

$$\mathbb{E}\left[e^{-\nu_1 i E_i}\right] = \frac{\binom{i}{2}+i\nu_2}{\binom{i}{2}+i(\nu_1+\nu_2)},$$

from which it follows that

$$\mathbb{E}[X^k] = \prod_{i=1}^k \frac{\frac{1}{2}(i-1)+\nu_2}{\frac{1}{2}(i-1)+\nu_1+\nu_2} = \prod_{j=0}^{k-1}\frac{2\nu_2+j}{2\nu_1+2\nu_2+j} = \frac{(2\nu_2)^{\overline{k}}}{(2(\nu_1+\nu_2))^{\overline{k}}},$$

where $z^{\overline{k}} := z(z+1)\cdot\ldots\cdot(z+k-1)$ is the kth rising factorial power of z. Using Lemma C.2.1, we recognize these as the moments of the Beta$(2\nu_2, 2\nu_1)$ distribution. That the moments uniquely determine this distribution completes the proof. □

Equation (2.30) can be generalized to the bivariate case, as demonstrated in the following. For this, we need some more notation. Let $\boldsymbol{k} := (k_1, k_2) \in \mathbb{N}_0^2$ and $|\boldsymbol{k}| := k_1 + k_2$. Then, as pointed out by Etheridge and Griffiths [2009], though with slightly different notation,

$$\begin{aligned}\mathscr{L}_{WFm} x^{k_1}(1-x)^{k_2} = &\left(\binom{k_1}{2} + k_1\nu_2\right) x^{k_1-1}(1-x)^{k_2} + \left(\binom{k_2}{2} + k_2\nu_1\right) x^{k_1}(1-x)^{k_2-1} \\ &- \left(\binom{|\boldsymbol{k}|}{2} + |\boldsymbol{k}|(\nu_1+\nu_2)\right) x^{k_1}(1-x)^{k_2}. \end{aligned} \tag{2.32}$$

Take $k_2 = 0$ and $k_1 = k$ to see that this generalizes (2.30). One may observe that this is also a special case of (2.11), and since the Wright–Fisher diffusion with weak mutation has a stationary distribution, we can form a normalized duality using the methodology of Section 2.2.1. In contrast to the ordinary Kingman coalescent, the dual is a death process on the space of non-negative two-dimensional vectors, with *two* possible transitions. Defining $\boldsymbol{\nu} = (\nu_1, \nu_2)$, the constants are given by

$$c(\boldsymbol{k}, \boldsymbol{k} - \boldsymbol{e}_1) = \binom{k_1}{2} + k_1\nu_2 \qquad c(\boldsymbol{k}, \boldsymbol{k} - \boldsymbol{e}_2) = \binom{k_2}{2} + k_2\nu_1 \qquad c(\boldsymbol{k}, \boldsymbol{k}) = \binom{|\boldsymbol{k}|}{2} + |\boldsymbol{k}||\boldsymbol{\nu}|. \tag{2.33}$$

The unnormalized function is given by

$$f(x, k) = f_{\boldsymbol{k}}(x) := x^{k_1}(1-x)^{k_2} \tag{2.34}$$

and, for the Beta$(2\nu_2, 2\nu_1)$ distribution, the moments (2.12) are explicitly given by (Lemma C.2.1)

$$m_{\boldsymbol{k}} = \mathbb{E}^X\left[f_{\boldsymbol{k}}(X)\right] = \frac{(2\nu_2)^{\overline{k_1}}\,(2\nu_1)^{\overline{k_2}}}{(2|\boldsymbol{\nu}|)^{\overline{|\boldsymbol{k}|}}}, \tag{2.35}$$

where $z^{\overline{k}}$ is the rising factorial. It then follows from our general construction that the Wright–Fisher diffusion with mutation is dual to the process $\boldsymbol{M}_t$ in the narrow sense with respect to the normalized duality function,

$$g_{\boldsymbol{k}}(x) = \frac{f_{\boldsymbol{k}}(x)}{m_{\boldsymbol{k}}}.$$

That is,

$$\mathbb{E}_x^X[g_{\boldsymbol{k}}(X_t)] = \mathbb{E}_{\boldsymbol{k}}^{\boldsymbol{M}}[g_{\boldsymbol{M}_t}(x)], \tag{2.36}$$

$\forall x \in [0,1], \boldsymbol{k} \in \mathbb{N}_0^2, t \geq 0$. The dual process is a two-dimensional pure-death process which, when in

state $\boldsymbol{k}$, has the following rates of transitions,

$$q(\boldsymbol{k}, \boldsymbol{k}-\boldsymbol{e}_1) = \frac{2|\boldsymbol{\nu}| + |\boldsymbol{k}| - 1}{2\nu_2 + k_1 - 1}\, c(\boldsymbol{k}, \boldsymbol{k}-\boldsymbol{e}_1) \qquad q(\boldsymbol{k}, \boldsymbol{k}-\boldsymbol{e}_2) = \frac{2|\boldsymbol{\nu}| + |\boldsymbol{k}| - 1}{2\nu_1 + k_2 - 1}\, c(\boldsymbol{k}, \boldsymbol{k}-\boldsymbol{e}_2).$$

Notice that $\boldsymbol{M}_t$, when initialized from $(k, 0)$ with $k \in \mathbb{N}$, is essentially a *one*-dimensional death process, M_t say, as there is no way for the second coordinate to change. M_t is the 'untilded' counterpart to $\tilde{M}_t$ of the special Feynman–Kac duality, and they are linked together in the same way as remarked upon at the end of Section 2.2.2.

2.3.4 Biological Discussion

The aim of this chapter is to define the notion of duality and demonstrate its utility. Sections 2.1-2.2 introduced it as quite a general concept, which indeed it is. In this section, however, we have placed duality in the context of population genetics without, so far at least, mentioning the motivation behind these models. The purpose of this subsection is to discuss the bridge between the mathematics and the biological interpretation.

In Definition 2.7, we defined the two-type Wright–Fisher model for a fixed parameter N which represents the (finite) number of individuals in the population. When it was introduced, the model supplanted many of its competitors which were deterministic in nature. On the one hand it included some sophistication which made it more realistic, capturing the important effects of random mating; on the other hand, such a model is necessarily more complex to analyse. Whilst some results may be gleaned in the discrete case (such as the probability that the a_1-individuals eventually fixate (i.e. N_t^1 reaches N for some generation t and remains in that state thereafter) which is stated in Feller [1951]), they are typically harder to derive than for the Wright–Fisher diffusion. Since populations are typically rather large, one can approximate the discrete model by this diffusion by sending $N \to \infty$, provided the rescaling is done in the right way. This rescaling roughly says that, if we consider the proportion of one genetic type, it behaves as a continuous, diffusive variable provided we look over long enough time scales. That is, the change from one generation to the next is relatively unimportant in the context of evolutionary history.

The Wright–Fisher diffusion has been identified as a universality class, being the infinite limit of various individual-based models whose dynamics differ from that of the Wright–Fisher model. The most notable of these was perhaps given by Moran [1958], who devised a continuous-time model with overlapping generations.

In order to commence a discussion about the meaning of duality in the context of population genetics, we must first define some terminology. Suppose that we have a finite collection of N individuals undergoing some dynamics (in discrete or continuous time) in which each child has exactly one parent, and the parent dies upon giving birth. There may be more than one parent giving birth at any given time. A *sample* of size $k \leq N$ at time t is when one selects k individuals from the population at time t independently *without* replacement. (One can think of taking a finite sample from an infinite number of individuals in a similar way, except that there is always an infinite number left over after each choice.) This will also be referred to as a k-sample.

Once one has a sample, one can ask: how are they related? We call this process the *genealogical relationship*, or simply the *genealogy*. This process has its time indexed in the opposite direction to that of the model defining the dynamics – we shall refer to this as *backwards* or *genealogical* time. It may be that all the sampled individuals had the same parent, but it may also happen that none of them have a parent in common, nor do any of their parents. By the time we reach two or more generations in genealogical time, some or all of the sampled individuals will have died (depending on the model dynamics), but their ancestry lives on through their parents and parents' parents, and so on. We refer to this ancestry (of a single individual) as a *lineage* or *line of descent*. We emphasize that a single lineage is represented by several individuals throughout its history. For simplicity, we shall say that the line of descent begins at the sampled individual itself.

The point of asking about the genealogy in the first place is that, unless the dynamics of the forwards-in-time model is trivial, it will undergo its own set of non-trivial dynamics, backwards in time. The genealogical process is driven by *coalescence events* which cause lineages to merge together. A coalescence happens when, at some point in the past, two or more of the individuals representing the sampled lineages had the same parent. An interesting event in the genealogy of a sample is its *most recent com-*

mon ancestor, which occurs when there is only one lineage remaining. The properties of this event depend strongly on the dynamics of the process: for example, it may not even exist if your population resides on two separate islands between which migration is impossible (but whether this makes a sensible population model is another question). Indeed, it may also not exist if the sample descended from more than one individual in the initial population.

In Section 2.3.2 we showed that, through (2.27), there exists a moment duality between the Wright–Fisher diffusion and the block-counting process of the Kingman coalescent (with a finite number of lineages). We may use this to interpret the k-coalescent as the genealogical relationship between k individuals independently sampled from the infinite number of individuals undergoing the Wright–Fisher diffusion dynamics. Actually, we have to be somewhat careful, since it is not exactly clear what is meant by this. In fact, what we mean is that the k-coalescent serves as a good approximation to the genealogy of a k-sample from a population undergoing the N-Wright–Fisher model (provided we adhere to the warning of Bhaskar *et al.* [2014] by ensuring the sample size is small enough relative to N). The approximation becomes exact as $N \to \infty$. At first sight, this may seem odd as the Kingman coalescent admits only pairwise mergers, whereas in the Wright–Fisher model, it is possible that parents give birth to more than two children. It turns out that mergers of three or more lineages are rare relative to the rate of pairwise mergers, and hence will not happen with high probability. The details of this approximation are reviewed on p.6 of Etheridge [2011].

The moment duality that we observe is an example where the temporal interpretation of duality is very natural: the Wright–Fisher diffusion models the evolution of the population *forwards* in time, whilst the Kingman coalescent describes the genealogical relationship of a sample of k individuals *backwards* in time.

The left-hand side of the duality equation (2.28), $\mathbb{E}_x^X[X_t^k]$, has the following natural interpretation: assuming that you take a k-sample at time t in which you do not know the allelic types, it is the probability that all k lineages are of type a_1 at this time. This is because, given X_t, the proportion of type a_1s in the population at time t, the probability of picking an individual of type a_1, is simply X_t. Since the population is infinite, sampling extra individuals is just like making independent draws from the

same distribution, and hence we multiply these probabilities together: for a k-sample it is X_t^k, given X_t. Finally, partition over the law of X_t to obtain the desired quantity. Note that, when conditioning on K_t, the number of lineages at genealogical time t, x^{K_t} is the probability that each lineage is of type a_1 (since t in genealogical time is the same as $t = 0$ forwards in time, when $X_0 = x$). Then, partitioning over the law of K_t gives the right-hand side of (2.28).

We could equally have asked what the probability of obtaining k type-a_2 individuals at time t is. Indeed, the duality function $\hat{f}_k(x) := (1-x)^k$ is also an exact duality function for the Wright–Fisher diffusion in the narrow sense, and yields the very same dual process. However, $f_{\boldsymbol{k}}(x) = x^{k_1}(1-x)^{k_2}$ is not an exact dual function so long as both k_1 and k_2 are positive (but can be regarded as a dual function strictly in the broad sense, along with an auxiliary function).

Perhaps the most striking difference between the Wright–Fisher diffusion models with ((2.29)) and without ((2.24)) weak mutation is the existence of a stationary distribution. We have already seen that, for the ordinary Wright–Fisher diffusion, one of the types will die out *a.s.*, whereas when we include the two-way mutation mechanism of Section 2.3.3, long-term coexistence is guaranteed and distributed according to a Beta law (Lemma 2.8). In light of DNA sequencing and the discovery of the abundance of mutations in many organisms, one could question why we would even consider this model of mutation. Whilst there are potentially millions of mutations at a single locus, many of them are synonymous and will tend to map into a small number of phenotypical traits. We can consider these as the genetic types. In particular, this model is appropriate if there are only two observable phenotypes at the locus of interest. Alternatively, if we are really interested in the mutations at the base level, this model can track exactly one mutation (a_1 say), whilst aggregating all the other possible mutations into the a_2 category (in which case we can think of a_2 as 'not a_1').

In the following we shall attempt to understand duality equation (2.31). We shall use arguments similar to those used by Tavaré [1984], whose lucid explanation of one-way mutation models clarifies many ideas. We should think of the dual process $\tilde{M}_t$, starting with k lineages, as a k-coalescent with coalescences occurring at rate $\binom{k}{2}$ except that, through mutation, each lineage is 'cut off' from the rest of the tree at an additional rate ν_2. For this reason, we actually have a forest of trees. Overall, the rate

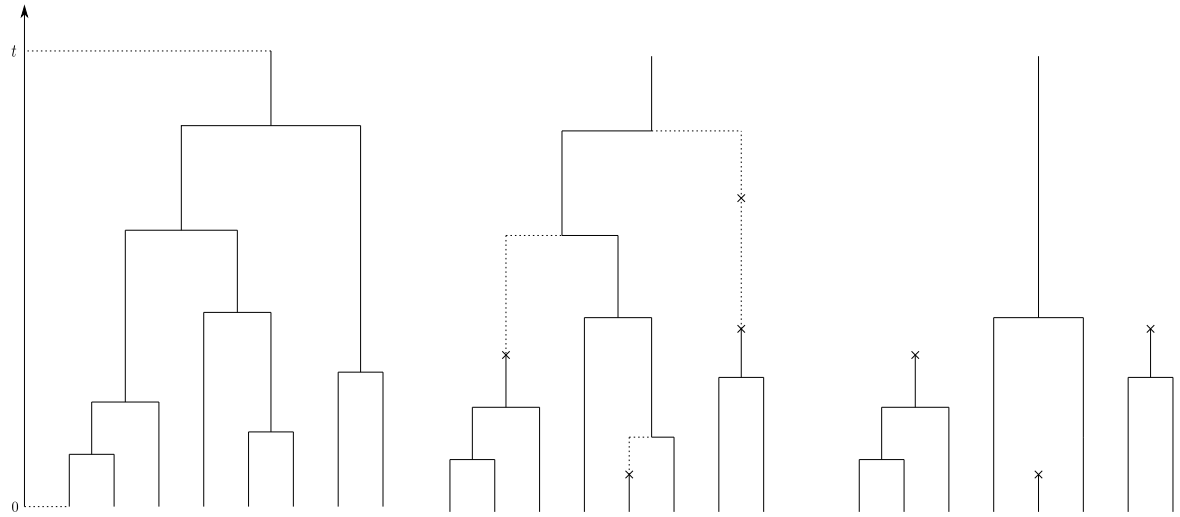


Figure 2.1: An example of how to construct a Kingman forest with $k = 8$ leaves. (a) Run an ordinary Kingman coalescent until genealogical time t. (b) Run an independent Poisson point process along each branch with mutation intensity $\nu_2 dt$ (mutations are marked by crosses). (c) Remove the (dashed) lines above the mutations closest to the leaves until you reach a lineage that has not been lost through mutation, or time t

of loss of lineages through mutation when there are k lineages is $k\nu_2$, which highlights the Poissonian nature of mutation. To emphasize this further, we could construct this forest in an alternative manner: start by generating an ordinary k-coalescent run until time t; run an independent Poisson point process of intensity $\nu_2 dt$ along each branch; finally, for each mutation, remove the branch segment connecting it to the root or next branch in the direction of the root. This is demonstrated pictorially in Figure 2.1. For this reason, we could think of this as a 'Kingman forest' with rate ν_2.

Why is ν_2 the important parameter in the dual process? In this situation we are concerned with finding the probability that k lineages are each of type a_1 at the time of sampling (at time t forwards in time, or time 0 in genealogical time). Therefore, going forwards in time, we only care about mutations which convert type-a_2 individuals into those of type a_1 (*provided* that that is the final mutation that occurs to that individual and its offspring), so that we see the relevant type when we sample. This happens at rate ν_2. Forwards in time, this final mutation will create a tree in our forest; backwards in time, the opposite happens: this *first* mutation that strikes the lineage annihilates it. We will see shortly that ν_1 also plays a role in the duality equation.

On the right-hand side of (2.31), we once again have the term $x^{\tilde{M}_t}$ which, given $\tilde{M}_t$, corresponds to the probability that each of the lineages at genealogical time t is of type a_1. In addition, we require that there are no mutations along $\tilde{M}_t$. Mutations of this type ($a_1 \to a_2$ forwards in time) occur at rate ν_1 independently for each lineage. Since the total length forest, given the full history $(\tilde{M}_u)_{0\le u\le t}$, is simply $\int_0^t \tilde{M}_u du$, this gives us the overall factor $\exp\left(-\nu_1 \int_0^t \tilde{M}_u du\right)$ as the probability of not seeing any mutations on $\tilde{M}_t$, using the additive nature of Poisson processes. Mutations in the other direction ($a_2 \to a_1$ forwards in time) along the forest are impossible since the individuals representing those lineages are all of type a_1. As usual, we need to partition over the law of $(\tilde{M}_u)_{0\le u\le t}$ to complete the argument.

When we consider the more general duality function $f_{\boldsymbol{k}}(x) = x^{k_1}(1-x)^{k_2}$, we take a $|\boldsymbol{k}|$-sample and ask what is the probability that there are k_1 type-a_1 lineages and k_2 type-a_2 lineages? We shall now apply the methods of Section 2.2.2 in this more general setting. Using the collection of constants $\{c(\boldsymbol{k}, \boldsymbol{k}')\}$ defined by (2.33), we can deduce that there is a dual process[1], $\tilde{\boldsymbol{M}}_t = (\tilde{M}_1(t), \tilde{M}_2(t))$, to the Wright–Fisher diffusion with weak mutation, X_t, such that the following duality equation holds

$$\mathbb{E}_x^X\left[X_t^{k_1}(1-X_t)^{k_2}\right] = \mathbb{E}_{\boldsymbol{k}}^{\tilde{\boldsymbol{M}}}\left[x^{\tilde{M}_1(t)}(1-x)^{\tilde{M}_2(t)} e^{-\nu_1\int_0^t \tilde{M}_1(u)du} e^{-\nu_2\int_0^t \tilde{M}_2(u)du} e^{-\int_0^t \tilde{M}_1(u)\tilde{M}_2(u)du}\right], \tag{2.37}$$

where $\tilde{\boldsymbol{M}}_t$ is jump process that performs the transitions $\boldsymbol{k} \to \boldsymbol{k} - \boldsymbol{e}_i$ at rate $\binom{k_i}{2} + k_i\nu_{!i}$, where $!i := \mathbb{1}_{\{i=2\}} + 2\cdot\mathbb{1}_{\{i=1\}}$ is the complimentary element of i in the two-point set $\{1,2\}$. This process behaves exactly like $\tilde{M}_t$ except that we now have a pair of collections of *typed* lineages. Type-a_2 lineages can be lost through a mechanism similar to that which removes a_1-lineages. Equation (2.37) has all the factors of (2.31) (think of $\tilde{M}_t$ as $\tilde{M}_1(t)$), plus some others. The factor $(1-x)^{\tilde{M}_2(t)}$ is the probability that, given $\tilde{\boldsymbol{M}}_t$, the $\tilde{M}_2(t)$ lineages at genealogical time t are chosen to be of type a_2. The factor $\exp\left(-\nu_2\int_0^t \tilde{M}_2(u)du\right)$ is the probability that no unhelpful mutations appear on the entire forest for type a_2 (that is, forwards in time, we do not observe any $a_2 \to a_1$ transitions after the final $a_1 \to a_2$ transition for that tree). The final factor, $\exp\left(-\int_0^t \tilde{M}_1(u)\tilde{M}_2(u)du\right)$, regulates the correlation between

[1] We once again use the tilde convention to mean that the dual process forms a duality in the broad sense, and in particular in the sense of Section 2.2.2

the two forests, and is the probability that no two individuals of different types coalesce at any point in the typed forest.

Notice that there are a number of differences between the Feynman–Kac duality and the normalized duality given by (2.36). Firstly, by design, we have an exact duality in the latter but not in the former. In general, from a practical standpoint, this exactness is quite useful because if you would like to estimate the moments of the forward process via simulation, say, then you do not need to compute functionals of the entire Kingman forest, as these moments depend only on $\tilde{\boldsymbol{M}}_t$ rather than $(\tilde{\boldsymbol{M}}_u)_{0\leq u\leq t}$. However, knowledge of the moments of the stationary distribution of the forwards-in-time process is then required, which may be difficult to calculate. (For the example we describe here, this is not a problem as they are given explicitly by (2.35).)

Another key difference is in the dual processes themselves: whilst $\boldsymbol{M}_t$ and $\tilde{\boldsymbol{M}}_t$ can jump to the same set of states from $\boldsymbol{k}$, they do so at different rates. In fact, as we have seen already, these rates are related: to obtain the rates for $\boldsymbol{M}_t$ from a given state $\boldsymbol{k}$, take the corresponding rate for $\tilde{\boldsymbol{M}}_t$, and multiply by the ratio $\frac{m_{\boldsymbol{k}'}}{m_{\boldsymbol{k}}}$, where $\boldsymbol{k}'$ is the new state. This ratio, however, is in general a non-trivial function of $\boldsymbol{k}$, and therefore difficult to interpret. Stephens and Donnelly [2003] explain why Equation (2.36) must correspond to *conditioning* on the types in the sample – that is, for a $|\boldsymbol{k}|$-sample, the composition into k_i type-a_i lineages is known. This is useful for most applications since we often know what the genotypes of the sequenced individuals are.

Equation (2.36) can be written differently as

$$\mathbb{E}_x^X\left[X_t^{k_1}(1-X_t)^{k_2}\right] = \mathbb{E}_{\boldsymbol{k}}^{\boldsymbol{M}}\left[\frac{m_{\boldsymbol{k}}}{m_{\boldsymbol{M}_t}}x^{M_1(t)}(1-x)^{M_2(t)}\right],$$

since $m_{\boldsymbol{k}}$ is a constant independent of X_t. We can now once again observe the left-hand side as the probability that our $|\boldsymbol{k}|$-sample contains k_i type-a_i individuals. The probability of drawing a particular sample $\boldsymbol{k}$ from the stationary distribution is represented by $m_{\boldsymbol{k}}$. So, given $\boldsymbol{M}_t$, $\mathbb{E}_x^X[f_{\boldsymbol{k}}(X_t)]$ depends on the ratio

$$\frac{\mathbb{P}(\text{draw the exact sample } \boldsymbol{M}_t \text{ from the initial population (at } t=0))}{\mathbb{P}(\text{draw the exact sample } \boldsymbol{M}_t \text{ from the stationary distribution } X \text{ (at } t=\infty))},$$

and then this is multiplied further by $\mathbb{P}$(draw the exact sample $\boldsymbol{k}$ from the stationary distribution X).

The method of renormalization by a stationary distribution may heuristically explain why one cannot form an exact duality from the function $f_{\boldsymbol{k}}(x)$ when both k_1 and k_2 are positive, and the mutation rates are zero. Whilst in the absence of mutation the Wright–Fisher diffusion does not have a stationary distribution, it does have a stationary measure which we can think of as $\pi(dy) = (1-x)\delta_0(dy) + x\delta_1(dy)$. We can then integrate against this measure to find $m_{\boldsymbol{k}} = \int_0^1 y^{k_1}(1-y)^{k_2}\pi(dy)$, which is a non-zero constant, independent of $\boldsymbol{k}$, if either $k_1 = 0$ or $k_2 = 0$, but vanishes if both are positive. Of course, the probability of drawing at least one individual of each type from a population which has fixated (or, said another way, has reached its stationary measure) is zero. For this reason, perhaps we can interpret the familiar duality equation (2.28) as

$$\mathbb{E}_x^X\left[\frac{X_t^k}{m_{(k,0)}}\right] = \mathbb{E}_k^K\left[\frac{x^{K_t}}{m_{(K_t,0)}}\right],$$

since the normalizing constants simply cancel out. A similar argument could be made for the function $f_{(0,k)}(x) = (1-x)^k$, but as soon as both $k_1, k_2 > 0$, it no longer makes sense to divide by $m_{\boldsymbol{k}}$, because it equals zero.

To end the discussion, we shall observe one last way of forming a duality for X_t. Let $\tilde{\boldsymbol{N}}_t = (\tilde{N}_1(t), \tilde{N}_2(t))$ be a jump process with transitions $\boldsymbol{k} \to \boldsymbol{k} - \boldsymbol{e}_i$ at rate $\binom{k_i}{2}$ (coalescence) and $\boldsymbol{k} \to \boldsymbol{k} - \boldsymbol{e}_i + \boldsymbol{e}_{!i}$ at rate $k_i\nu_{!i}$ (mutation). Then, X_t and $\tilde{\boldsymbol{N}}_t$ are dual in the broad sense with respect to the functions $(f_{\boldsymbol{k}}, 0, (\nu_2-\nu_1)(k_1-k_2) - k_1k_2)$. That is to say

$$\mathbb{E}_x^X\left[X_t^{k_1}(1-X_t)^{k_2}\right] = \mathbb{E}_{\boldsymbol{k}}^{\tilde{\boldsymbol{N}}}\left[x^{\tilde{N}_1(t)}(1-x)^{\tilde{N}_2(t)}e^{(\nu_2-\nu_1)\int_0^t \tilde{N}_1(u)-\tilde{N}_2(u)du}e^{-\int_0^t \tilde{N}_1(u)\tilde{N}_2(u)du}\right]. \quad (2.38)$$

This is somewhat difficult to interpret since $(\nu_2-\nu_1)(k_1-k_2)$ is neither positive nor negative in general. However, this problem disappears if we set the mutation parameters to be equal.

CHAPTER 3

The Λ-Fleming–Viot Process and Its Dual

Just before the turn of the twenty-first century, there was a rapid burst of papers generalizing the Kingman coalescent. Independently, and in the same year, Donnelly and Kurtz [1999], Pitman [1999] and Sagitov [1999] published descriptions of the coalescent with multiple mergers (or the 'Λ-coalescent' as it is now known – the naming convention due to Pitman's notation). The main feature of this model is that *more than* two lineages may coalesce together within a single event, unlike the Kingman coalescent which admits pairwise mergers only.

There are situations in which a Λ-coalescent may be more suitable for modelling purposes. Firstly, there is the important phenomenon of the selective sweep: when a beneficial allele arises in the pop-

ulation it may, depending on the strength of the advantage, quickly[1] 'sweep' through to fixation (or near-fixation) so that every (or nearly every) individual possesses that mutation. Thinking backwards-in-time, and because the sweep was so fast (or instantaneous, in fact, in Gillespie's hitchhiking model (Gillespie [2000])), it may appear that the original recipient of the mutation mothered many children. This phenomenon, it is argued, is well-modelled by a multiple merger.

Another scenario in which the Λ-coalescent might be a suitable model for the genealogy of a population is one which undergoes repeated extinction and subsequent recolonization events. Consider, for example, a forest of trees which experiences frequent destructive fires. During the event many trees will die out, and the individual which colonizes the area the quickest will have a large family. This example, however, may be better-suited to a *spatial* extinction-recolonization model such as that described in Section 3.3.

Finally, there have recently been a number of papers studying the genealogies of marine organisms, such as Pacific oysters (Eldon and Wakeley [2006]) and Atlantic cod (Steinrücken *et al.* [2013]), which are believed to use a 'sweepstake reproduction' mechanism. This means that, from time to time, highly fecund individuals give birth to large number of offspring, constituting a large fraction of the entire population. More details can be found in Example 3.7.

Bertoin and Le Gall [2003] formally introduced the corresponding forwards in time process, the Λ-Fleming–Viot process. They also proved that it is dual to the Λ-coalescent and that this is the correct model for the genealogy of the population undergoing such dynamics forwards-in-time. In this section, in a somewhat chronologically-flawed way, we shall define this process in Section 3.1, in general and in the simplified two-type case, before discussing the duality in Section 3.4. Prior to that, we shall consider a pre-limiting version of this model in Section 3.2, and briefly review the *spatial* Λ-Fleming–Viot process in Section 3.3. We finish the chapter in Section 3.5 by studying the *block size spectra* of the Kingman and Beta coalescents, and their scaling limits. There we derive an expression for the falling factorial moments of the block size spectrum of the Kingman coalescent, and describe how its scaling limit can be related to the solution of a certain Smoluchowski coagulation equation. In addition, we

[1]The use of the word 'quickly' here is relative to evolutionary timescales.

study the manner in which the block size spectrum of the Beta coalescent behaves as it comes down from infinity. To this end, we derive a system of ordinary differential equations that the hydrodynamic limit satisfies, generalizing the Smoluchowski coagulation equations in this direction. Our main result, Theorem 3.18, states that the solution of this system can be written as a complete Bell polynomial. (For a brief introduction to Bell polynomials, we refer the reader to Appendix B.)

3.1 Definition

The Λ-Fleming–Viot process is a general model capable of describing populations undergoing potentially large restructuring events. It is most easily defined via a Poissonian construction (Bertoin and Le Gall [2003]). This section is only intended to be a brief introduction; much more detail can be found in the review by Berestycki [2009].

Definition 3.1 (Λ-Fleming–Viot process)**.** Let $\rho_t(\cdot)$ be a time-indexed family of measures on a compact type space K. Let Π be a Poisson point process on $\mathbb{R}_+ \times (0,1]$ with intensity $dt \otimes u^{-2}\Lambda(du)$, where Λ is a finite measure on $(0,1]$ satisfying $\Lambda(\{0\}) = 0$. Suppose initially that the distribution of types is given by the measure $\rho_0(\cdot)$. For $(t,u) \in \Pi$, if $\rho_{t-}(\cdot)$ represents the distribution of types at time $t-$, then after the event we have

$$\rho_t(\cdot) = (1-u)\rho_{t-}(\cdot) + u\delta_k, \tag{3.1}$$

where k, a variate from the measure $\rho_{t-}(\cdot)$, is the chosen parental type.

The measure Λ parametrizes the process, and is responsible, in part, for the name that ρ bears. We shall sometimes refer to Λ as the *driving measure*. We can extend it further by allowing it to contain a point mass at zero, i.e. $\Lambda(\{0\}) > 0$. This would be referred to as the 'Kingman component' of the driving measure, but would require broadening Definition 3.1 to something more subtle, and perhaps less enlightening. In fact, the Λ-Fleming–Viot process is an extension of the classical Fleming–Viot process, and corresponds exactly to the special case $\Lambda(du) = \delta_0(du)$ where the driving measure is exclusively a Kingman component. The Fleming–Viot process is studied at length in Etheridge [2000].

Let the generator of $\rho_t(\cdot)$ be denoted $\mathscr{L}_\Lambda$. Its domain $\mathscr{D}(\mathscr{L}_\Lambda)$ consists of the functions

$$G(\rho) = \int_{K^n} g(x_1, \ldots, x_n)\rho(x_1)\ldots\rho(x_n),$$

where $n \in \mathbb{N}$ and $g : K^n \to \mathbb{R}$ is a bounded measurable function. It may be written

$$\mathscr{L}_\Lambda G(\rho) = \int_{(0,1]} \int_K \left(G((1-u)\rho + u\delta_k) - G(\rho)\right) \rho(dk) \frac{\Lambda(du)}{u^2}. \tag{3.2}$$

This generator is akin to that of a jump process, though defined on the space of probability measures on K. It is a special case of (3.1), p.376, in Ethier and Kurtz [1986].

The point about defining the process ρ on K is that it may cope with an arbitrary number of allelic types, including *infinitely many* types. This may be an important approximation for DNA sequences, where the type space is finite but very large. On the other hand, we can also specialize to the more familiar two-type case – this was the setting for the Wright–Fisher model in Section 2.3. To do this, simply let K to be the two-point set $K = \{1, 2\}$, and associate 1 with allele a_1 and 2 with a_2. Let X_t be the proportion of a_1-alleles. (This convention will be used throughout the remainder of this thesis, which focuses on the two-type Λ-Fleming–Viot process and extensions thereof.) In this case, the generator in (3.2) has a much simpler form:

$$\mathscr{L}_\Lambda f(x) = \int_0^1 \left[xf((1-u)x + u) + (1-x)f((1-u)x) - f(x)\right] \frac{\Lambda(du)}{u^2}, \tag{3.3}$$

where f is a bounded, measurable function. This time, the process takes its values on the one-dimensional space $[0, 1]$, as it tracks the proportion of a_1-alleles (although the notion of proportion is unclear without an individual-based model – see Section 3.2). The generator in (3.3) has an intuitive explanation, which we postpone until Section 3.2.

In the generator $\mathscr{L}_\Lambda$, we integrate against the measure $\nu(du) := u^{-2}\Lambda(du)$, which may well be infinite (which has the consequence that an uncountable number of events may fall on a compact time interval). We must impose, however, in order for the process to be well-defined, that Λ be finite. To begin to see why this is true, take a smooth function f and perform a Taylor expansion about x to obtain

$$\mathscr{L}_\Lambda f(x) = \frac{1}{2}x(1-x)f''(x)\int_0^1 \Lambda(du) + \sum_{i=1}^{\infty} c_i(x) f^{(i+2)}(x) \int_0^1 u^i \Lambda(du). \tag{3.4}$$

In particular, the zeroth and first derivatives of f cancel exactly, and consequently the terms corresponding to $i = -1, -2$ vanish. That $\int_0^1 \Lambda(du) < \infty$ is equivalent to ν having a finite second moment ensures that the generator remains bounded when acting on smooth functions in its domain.

3.2 Individual-Based Model

The pre-limiting model of the Λ-Fleming–Viot process described in this section will shed light on the dynamics of the limiting model. We shall define the extinction-recolonization process for a biallelic haploid population with allelic types a_1 and a_2. Fixing $\lambda > 0$, define X_t^λ to be the proportion of a_1-alleles at time t in a population of size N_t^λ. Suppose that initially the number of individuals in this population, $N_0^\lambda \stackrel{d}{=} \text{Po}(\lambda)$, is taken to be a Poisson variate with mean λ; set $X_0^\lambda = \frac{n}{N_0^\lambda}$, $n \in [N_0^\lambda]_0 := \{0, 1, \ldots, N_0^\lambda\}$. Now, let Π be a Poisson point process on the space $\mathbb{R}_+ \times (0, 1]$ with intensity $dt \otimes \nu(du)$, where ν is some measure on $(0, 1]$. The points $(t, u) \in \Pi$ parametrize the events of the process, which occur instantaneously and consist of three stages:

(1) Select a parent *uniformly* at random from the N_{t-}^λ members of the population.

(2) Every member of the population dies independently probability u (extinction).

(3) A Po(λu) number of children with the same allelic type as the parent are born into the population (recolonization).

In contrast to the Wright–Fisher model of Section 2.3.1, the population size may vary. In particular, there is some positive probability that the process dies out. However, Berestycki *et al.* [2009] show that, in a spatial context, there exists some $\lambda_0 \in (0, \infty)$ such that, with probability one, for each $\lambda > \lambda_0$, the population survives forever; for $\lambda \in [0, \lambda_0)$ the population will eventually die out. Moreover, Etheridge and Kurtz [2014], Section 4.1.2, prove that this individual-based model converges to the Λ-Fleming–Viot process of Section 3.1 as $\lambda \to \infty$. That is, $(X_t^\lambda)_{t\geq 0} \xrightarrow{d} (X_t)_{t\geq 0}$ whilst $\frac{N_t^\lambda}{\lambda} \xrightarrow{\mathbb{P}} 1$ as $\lambda \to \infty$.

The results of that paper, however, are far more general than we have described here – for example, convergence is shown for the more general type space, K, and when the population is distributed over a geographical space (for a brief introduction to the *spatial* Λ-Fleming–Viot process, see Section 3.3).

In the limiting process, there is no concrete notion of what an individual really is, but it is helpful to think back to the pre-limiting model to understand how individuals behave. Purely at an intuitive level, it is also helpful to think of a collection of individuals as a mass: for example, the collection of a_1-individuals has mass x. Indeed, as $\lambda \to \infty$, the population 'size' stabilizes to a constant, so that there is *no* change in total population mass in the limiting model[1].

The correspondence between the dynamics of the individual-based model and the components of (3.3) can now be explained. Firstly, the integral against the measure ν is how the event parameter u is picked, and determines the 'impact' of the event; this is the same in both models. We can match the different terms that compose the generator to the stages outlined in the above definition.

(1) One can think of x as representing the proportion or mass of 'individuals' of type a_1. Since we pick the parent uniformly, x is also the limiting probability that the parent is of type a_1 (and $1-x$ is the complimentary probability of picking an a_2-parent). These prefactors determine the probabilities of where the process could jump to during an event.

(2) Each individual is killed with probability u, and so the proportion of remaining a_1-individuals is approximately thinned by a factor $1-u$ no matter which parent was picked. In the limit, this fraction stabilizes to exactly $(1-u)x$.

(3) The mass of newborns replaces the mass lost through killing, which is exactly u. They each receive the same allelic type as their parent, and so the mass can be assigned either type a_1 or a_2, depending on the parental type (the respective probabilities are x and $1-x$). If it is a_1, then the value u is added, otherwise it is not.

[1]This aspect of the model has further been generalized by Etheridge and Kurtz [2014], Section 4.2. In order to obtain a limit whose population density varies, one should use an offspring distribution with a stronger variance, replacing the Po(λu) by a Geom(u/λ) random variable, say.

Overall, during a u-event, the process may jump to the new value $(1-u)x+u$ (with probability x), or to $(1-u)x$ (with probability $1-x$). The $-f(x)$ term in the generator follows from the jump nature of the process.

Remark 3.2. The Λ-Fleming–Viot process may also be obtained from an individual-based model with a fixed population size N. This will be discussed in a generalized setting in Section 4.2.

3.3 Spatial Λ-Fleming–Viot Process

The dynamics of the Λ-Fleming–Viot process can naturally be extended over a continuous geographical space in such a way that the large-scale events have a spatial dependence, along with the measures ρ_t. Such a model was introduced by Etheridge [2008] in order to overcome 'the pain in the torus' (Felsenstein [1975] coined this phrase to refer to the problems with the models studied by Wright [1943] and Malécot [1948] in which reproduction is driven by individuals, *à la* classical Wright–Fisher model), her main innovation being that reproduction is now driven by an *external* source associated with the space itself, rather than the members of the population residing on it. This mechanism is set up to regulate the population at a local level, and prevent the clumping phenomenon exhibited by the early models. The purpose of this section is not to introduce any new results, but to inform the reader of generalizations that are being studied, and how to visualize them.

The spatial Λ-Fleming–Viot model provides a way to represent the dynamics of gene flow in a continuous metric space (E,d), where $E\subset\mathbb{R}^d$. As stated in Barton *et al.* [2010], the main purpose of the model is to capture these dynamics over a two-dimensional continuum, such as $\mathbb{R}^2$, but there is no harm in defining it more generally. There is an analogous individual-based model in the spatial setting, which is described in Berestycki *et al.* [2009], and the convergence of this model to the spatial Λ-Fleming–Viot process is proved in Etheridge and Kurtz [2014].

Suppose once again that at a particular locus we have a two-point type space, $K=\{a_1,a_2\}$, and we ignore the possibility of mutations within or out of K. Evolution is again driven by extinction-recolonization events, but this time governed by an *extended* Poisson point process, Π_S, on $\mathbb{R}_+\times E\times$

$\mathbb{R}_+ \times (0,1]$ with intensity $dt \otimes dx \otimes \mu(dr)\nu_r(du)$. For the two-type spatial Λ-Fleming–Viot process, it is common to denote the proportion of type-a_1 individuals at time $s \in \mathbb{R}_+$ and position $y \in E$ by $w_s(y)$. Given some initial configuration, $w_0(\cdot)$, each event $(t,u,x,r) \in \Pi_S$ induces the following changes:

(i) define $B_r(x)$ to be a ball of radius r with centre x,

(ii) choose a parental point uniformly at random, $z \in B_r(x)$,

(iii) choose a type from that position: type a_1 is chosen with probability $w_{t-}(z)$,

(iv) for all $y \in B_r(x)$, set $w_t(y) = (1-u)w_{t-}(y) + u\,\mathbb{1}_{a_1\text{-parent}}$.

As in the non-spatial analogue, we may not use arbitrary measures in the Poisson point processes; Barton *et al.* [2013] show that

$$\int_0^\infty r^d \int_0^1 u\nu_r(du)\mu(dr) < \infty$$

is a sufficient condition to guarantee the existence of the process (although slightly weaker conditions have been established in Etheridge and Kurtz [2014]).

In order to visualize the process on a non-abstract metric space, consider the unit sphere. Take $E = S_2 := \{(y_1,y_2,y_3) \in \mathbb{R}^3 : y_1^2 + y_2^2 + y_3^2 = 1\}$. Set $\mu(d\tilde{r}) = \delta_r(d\tilde{r})$ and $\nu_r(d\tilde{u}) = \delta_u(d\tilde{u})$ so that we can think of r and u as parameters themselves. Since S_2 is a compact space, we can view the centre of each event as being a draw from $\mathcal{U}(S_2)$, the uniform distribution on the sphere. The region of the event is defined through

$$\tilde{B}_r(x) = \{y \in \mathbb{R}^3 : ||y-x||_2 \leq r\} \subset \mathbb{R}^3;$$

put $B_r(x) = \tilde{B}_r(x) \cap S_2$. With these definitions in place, we can simulate this model using a discrete version of the sphere. As we do not have enough dimensions, we must use colour in order to represent w_t: black corresponds to zero, white to one, and the shades of grey interpolate between these two extremes.

Two simulations, spawned from two different initial conditions, have been depicted in Figures 3.1 and 3.2. In the former, the initial condition is the 'hemisphere' landscape, where $w_0(y) = \mathbb{1}_{\{y_1 \leq 0\}}$, $y \in$

Figure 3.1: A simulation of the spherical Λ-Fleming–Viot process with the 'hemisphere' initial condition and fixed parameters $r = 0.05$ and $u = 1$. We have shown the process after 0 (left), 200,000 (centre) and 400,000 (right) events

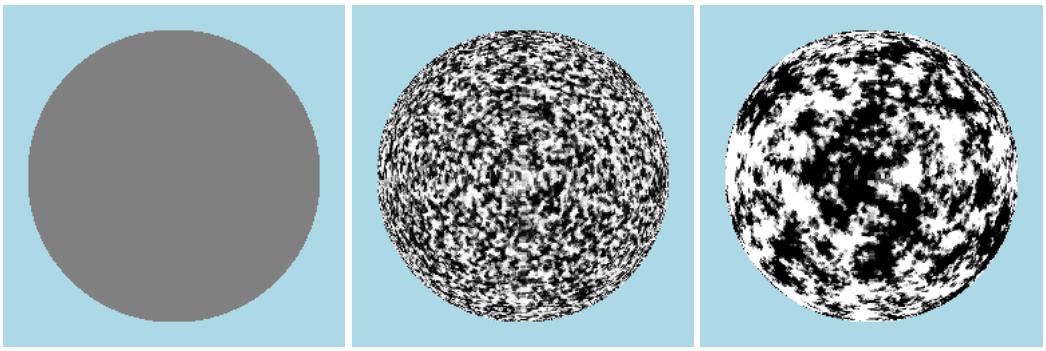


Figure 3.2: A simulation of the spherical Λ-Fleming–Viot process with the 'homogeneous' initial condition and fixed parameters $r = 0.01$ and $u = 0.5$. We have shown the process after 0 (left), 200,000 (centre) and 10^7 (right) events

S_2, and in the latter, the 'homogeneous' landscape, $w_0(y) = 0.5,\ y \in S_2$. Note that *all* events are shown, even if they do not make an overall difference to the state of the process.

In Figure 3.1, we have a small radius of size $r = 0.05$ and impact parameter $u = 1$, which ensures that every locality will always be homogeneous, i.e. completely white or black. Evidently, it may be difficult to characterize the behaviour of the process, as some complicated boundary behaviour unfolds. We can see from the images that after a large number of events, we obtain a strange geography including many peninsulas and enclaves that have migrated far beyond the obvious borderline.

In Figure 3.2, we observe what happens when every point has the same half-half configuration. The parameters are $r = 0.01$ and $u = 0.5$. After 200,000 events, the process appears to still be quite homogeneous, but in a different way: small patches of different coloured regions are distributed almost

uniformly. After a very large number of events, however, we begin to observe extensive regions that are apparently either white or black in colour. Due to the initial condition, $w_t(y) \in (0,1)$ for all time and so these regions cannot be truly homozygous. However, it seems that the process finds more stability in very light or very dark shades of grey.

Remark 3.3. An excellent depiction of the Λ-Fleming–Viot dynamics on the real line ($E = \mathbb{R}$) is presented in Figure 1 of Saadi [2014].

3.4 Duality: Λ-Coalescent

In this section we shall define the genealogy of the Λ-Fleming–Viot process through a duality argument. This will require some more understanding of the Λ-coalescent, which was discovered simultaneously by Donnelly and Kurtz [1999], Pitman [1999] and Sagitov [1999]. First, we need some notation.

A *partition* of a set A is a set of sets (or 'blocks') contained in A whose pairwise complement is empty, and whose union is A. Let $\mathscr{P}_A$ be the set of partitions of the set A, and $\mathscr{P}_{A,b}$ the set of partitions of A with exactly $b \in \mathbb{N}$ blocks. If $\pi_1, \pi_2 \in \mathscr{P}_{[n]}$, then we say that π_2 *covers* π_1 if each block in π_1 is contained in a block of π_2. For example, the *complete* partition $\{[n]\}$ covers the *singletons* $\Delta_n := \{\{1\}, \{2\}, \ldots, \{n\}\}$. (Note that, in fact, each partition $\pi \in \mathscr{P}_{[n]}$ is covered by $\{[n]\}$. Also, each π covers itself, trivially, and Δ_n.) On the other hand, if we define $\pi = \{[n-3], \{n-2, n-1\}, \{n\}\}$ and $\pi' = \{[n-3], \{n-2\}, \{n-1, n\}\}$ for $n > 3$, then π does not cover π', nor vice versa. Finally, define $|\pi|$ to be the number of blocks in the partition.

Definition 3.4. Let $(\Pi^n(t))_{t\geq 0}$ be a time-homogeneous Markov process on the space $\mathscr{P}_{[n]}$, $n \in \mathbb{N}$, with initial condition Δ_n. Suppose that $\Pi^n(t)$ has b blocks and π is a partition which covers $\Pi^n(t)$ such that $b - |\pi| + 1 = k$. If the process jumps to π at rate

$$\lambda_{b,k} = \int_0^1 u^k (1-u)^{b-k} \frac{\Lambda(du)}{u^2}, \qquad 2 \leq k \leq b, \tag{3.5}$$

where Λ is a finite measure on $(0,1]$, then $\Pi^n(t)$ is said to be a Λ-*coalescent*. Such a jump is referred to as a k-*merger*, indicating that k of the b blocks merged into one.

So far, we have said nothing about the *order* in which the blocks are arranged. An important property of coalescents is that they possess the *exchangeability* property, meaning that any permutation of the blocks is as likely as any other. For this reason, the order of the blocks is unimportant, and by default we list them by their minimum element.

Suppose we take $m \leq n$ and restrict the Λ-coalescent (of size n) to the set $[m]$, then this restricted process has the same law as that of a Λ-coalescent (of size m). This is called the *consistency* property. It turns out that any consistent coalescent must satisfy (3.5). From a biological point of view, this is a desirable property because any sample that you take is unaffected by adding more points to your sample (and in particular old samples can be combined with new ones). What we have really defined so far is the n-Λ-coalescent, but due to consistency, this is embedded within the $(n+k)$-Λ-coalescent for each $k \in \mathbb{N}$. This allows us to define, as a projective limit, the full Λ-coalescent whose sample space is $\mathscr{P}_{\mathbb{N}}$, the partitions of $\mathbb{N}$.

The Λ-coalescent on $\mathscr{P}_{\mathbb{N}}$ has infinitely many blocks at time zero. However, depending on the measure Λ, it is possible that coalescence events happen so fast that it has finitely many blocks for any positive time. If such a phenomenon occurs, the coalescent is said to *come down from infinity*. In Section 3.5.3, we shall study the manner in which a certain range of Beta coalescents (see Example 3.5) comes down from infinity.

We remark that $\lambda_{b,k}$ is finite $\forall b \in \mathbb{N}, k \in \{2, \dots, b\}$ due to the restriction placed on Λ. The process halts as soon as it reaches the complete partition, which it does so in finite time *a.s.* for any finite n. Informally, we can think of the Λ-coalescent in the following way: whenever there is a u-event, perform b independent coin tosses, one for each block, with a coin which shows heads with probability u. If the block received heads, then it will take place in the merger.

The process at the partition level contains fine-grain information; in what follows we will be interested in something more coarse. Define $C(t) := |\Pi^n(t)|$ to be the *block-counting process* of Π^n with initial value $C(0) = n$. Since there are $\binom{b}{k}$ subsets of size k, any one of them merging together will

cause the number of blocks to decrease by $k-1$, and so we may define the transition rates of $C(t)$ to be

$$q(b, b-k+1) := \binom{b}{k}\lambda_{b,k}, \qquad 2 \le k \le b. \tag{3.6}$$

In the literature, it is conventional to define

$$\lambda_b := \sum_{k=2}^{b}\binom{b}{k}\lambda_{b,k} = -q(b,b). \tag{3.7}$$

What the Kingman coalescent is to the Wright–Fisher diffusion, the Λ-coalescent is to the Λ-Fleming–Viot process. This may be observed using Proposition 2.6 and through a simple moment duality calculation. Applying the moment duality function $f_n(x) = x^n,\ n \ge 2$, to (3.3), we get

$$\begin{aligned}
\mathscr{L}_\Lambda f_n(x) &= \int_0^1 \left[\sum_{k=0}^{n}\binom{n}{k}u^k(1-u)^{n-k}(x^{n-k+1}-x^n) + x^n(1-x)(1-u)^n\right]\frac{\Lambda(du)}{u^2} \\
&= \int_0^1 \left[\sum_{k=2}^{n}\binom{n}{k}u^k(1-u)^{n-k}(x^{n-k+1}-x^n)\right]\frac{\Lambda(du)}{u^2} \\
&= \sum_{k=2}^{n}\binom{n}{k}(x^{n-k+1}-x^n)\int_0^1 u^{k-2}(1-u)^{n-k}\Lambda(du) \\
&= \sum_{k=2}^{n}\binom{n}{k}\lambda_{n,k}\,(f_{n-k+1}(x)-f_n(x)).
\end{aligned} \tag{3.8}$$

In the first equality, we made use of some binomial expansions, including the identity

$$\sum_{k=0}^{n}\binom{n}{k}u^k(1-u)^{n-k} = 1.$$

In the second, we recognized that shedding the terms corresponding to $k = 0, 1$ in the summation would lead to cancellations. In the third, we simply swapped the sum and the integral; and finally we made use of definition (3.5) in the last.

Equation (3.8) is a special case of the generator, (2.3), of a pure-jump Markov process in its second (subscript) coordinate. In fact, it is precisely the generator of the block-counting process, $C(t)$, of the Λ-coalescent started with n blocks. Using Proposition 2.6, we conclude that

$$\mathbb{E}_x^X\left[X_t^n\right] = \mathbb{E}_n^C[x^{C(t)}], \tag{3.9}$$

$\forall x \in [0,1], n \in \mathbb{N}, t \geq 0$, where X_t is the two-type Λ-Fleming–Viot process. Similarly to our remark in Section 2.3.2, questions about X_t can now be cast into those concerning the block-counting process of the Λ-coalescent.

3.4.1 Tree Generation

There is a simple recursive algorithm that may be used to generate random trees described by a Λ-coalescent law. This algorithm, whilst not at all novel, may be helpful to understand the process in more detail. It also describes the method by which the images in Figures 3.3 and 3.4 were generated.

Algorithm. *Firstly, fix a finite measure* Λ *so that* $\lambda_{b,k}$ *are specified. Initially, set* $t \leftarrow 0$ *and* $b \leftarrow n \in \mathbb{N}$*, and label each lineage uniquely with a number in* $[n]$*. Repeat the following steps while* $b > 1$*.*

- *Generate* $T \overset{d}{=} Exp(\lambda_b)$*, the random time that has elapsed since the previous event, and set* $t \leftarrow t + T$*.*
- *For* $2 \leq k \leq b$*, the probability that the event is a* k*-merger is* $\frac{q(b,b-k+1)}{\lambda_b}$*. Generate a uniform random variate* $U \overset{d}{=} \mathcal{U}(0,1)$ *to decide which value of* k *is picked. More precisely, defining* $S_k := \sum_{i=2}^{k} \frac{q(b,b-i+1)}{\lambda_b}$*, a* k*-merger happens if* $U \in [S_{k-1}, S_k)$*.*
- *Given the value of* k*, choose* k *of the* b *lineages independently and* without *replacement. Relabel each of these lineages with the smallest label from those selected, and declare this to be a single block from now on.*
- *Set* $b \leftarrow b - k + 1$*.*

3.4.2 Examples

Example 3.5 (Beta coalescent)**.** One well-known and well-studied family of Λ-coalescents is that of the Beta coalescent. Whilst one could legitimately take any Beta measure with two positive parameters as the driving measure, there is a one-parameter family which arises naturally. For $0 < \alpha < 2$, the

Λ-coalescent with driving measure

$$\Lambda^{\alpha}(du) := \frac{1}{K_{\alpha}} \left(\frac{u}{1-u} \right)^{1-\alpha} du \tag{3.10}$$

where $K_{\alpha} := \Gamma(\alpha)\Gamma(2-\alpha)$ and $\Gamma(z)$ is the Gamma function, is referred to as the *Beta coalescent*. It can be viewed as a time change of the genealogy of an α-stable continuous-state branching process, as outlined in Birkner *et al.* [2005]. Moreover, in the parameter range $\alpha \in (1, 2)$, this coalescent arises as the limit of the model defined by Schweinsberg [2003].

Remark 3.6. It is possible to define the *generalized* Beta coalescent, allowing a two-parameter measure, as

$$\Lambda_{a,b}(dx) := B(a,b)^{-1} x^{a-1}(1-x)^{b-1} dx, \quad a, b > 0, \tag{3.11}$$

where $B(a, b)$ is the Beta function. It contains the one-parameter family given in (3.10) by setting $a = 2 - \alpha$ and $b = \alpha$.

This family contains other important coalescents. For instance, when $\alpha = 1$ and Λ is simply the Lebesgue measure on $(0, 1]$, we have the Bolthausen-Sznitman coalescent. A realization of this process with 50 leaves is depicted in Figure 3.3, along with two higher values of α. One may observe informally from these images that, as $\alpha \uparrow 2$, pairwise mergers occur more frequently. This is a corollary of the fact that the Beta coalescent tends in distribution to the Kingman coalescent under this limit (Theorem 3.4, Berestycki [2009]). Note that the normalization constant K_{α} is important for this convergence scheme: if it were to be omitted, a time-change would be necessary. At the other extreme, as $\alpha \downarrow 0$, the Beta coalescent converges to the star-shaped coalescent, in which all n lineages merge together at some random, exponentially distributed, time.

Example 3.7 (Eldon-Wakeley coalescent)**.** Another one-parameter family arises as a limit of a variant of a discrete-time Moran model in which the parent of a reproduction event either replaces one individual from the population (with high probability) or a large fraction, u, of the population with vanishingly small probability. This type of reproductive behaviour, coined 'reproduction sweepstakes' by Hedgecock [1994], has been observed in some marine species such as Pacific oysters, the organism studied by Eldon

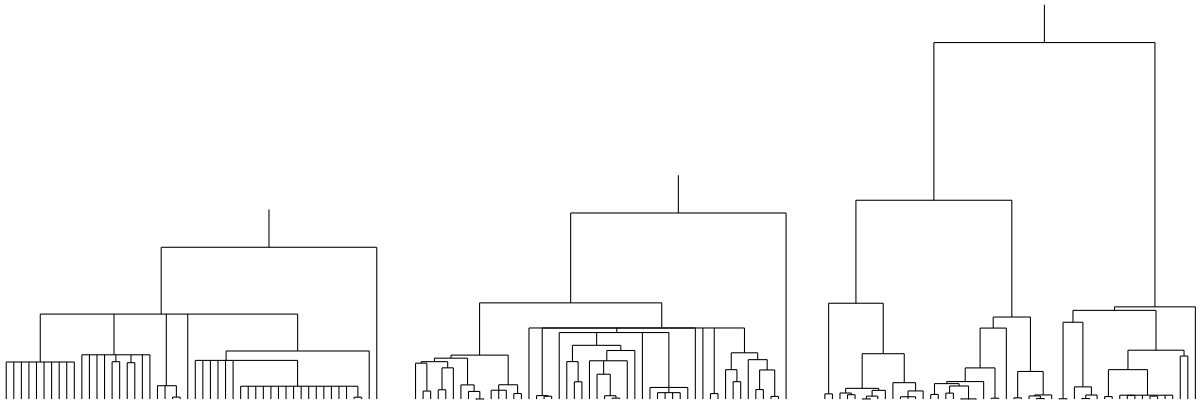


Figure 3.3: Three realizations of the Beta coalescent with different parameter values. Left: $\alpha = 1.0$ (Bolthausen-Sznitman); centre: $\alpha = 1.5$; right: $\alpha = 1.8$. They each have the same time scale on the vertical axis

and Wakeley [2006]. The genealogy is described by the driving measure $\Lambda(d\tilde{u}) = \delta_u(d\tilde{u}),\ u \in (0, 1]$. This family has similar limiting behaviour to that of the Beta coalescent in that, as $u \downarrow 0$, the Eldon-Wakeley coalescent converges weakly to the Kingman coalescent; when $u \uparrow 1$, it converges to the star-shaped coalescent. This property may be observed informally in Figure 3.4 in which three realizations of the process have been depicted for various values of u.

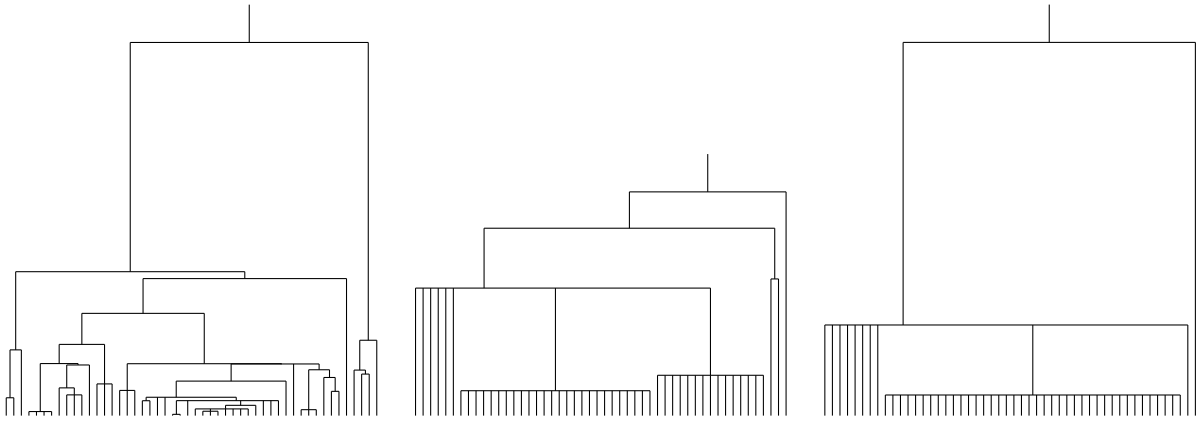


Figure 3.4: Three realizations of the Eldon-Wakeley coalescent with different parameter values. Left: $u = 0.0775$; centre: $u = 0.5$; right: $u = 0.75$. They each have the same time scale on the vertical axis

Eldon and Wakeley [2006] were the first to reject the null hypothesis that Kingman's coalescent is appropriate for the genealogy of the sample of Pacific oysters, and proposed $\hat{u} \approx 0.08$ as the sweepstake parameter. Using a different statistical test, Steinrücken *et al.* [2013] again reject the Kingman coalescent

for that dataset. They chose to model the genealogy differently, this time using a Beta coalescent, and concluded that $\hat{\alpha} = 1.15$. Their study also extended to several samples of Atlantic cod from various geographical locations, some of which could reject the Kingman hypothesis, with an average parameter value of $\hat{\alpha} \approx 1.4$.

3.5 Block Size Spectra

We shall study the so-called *block size spectrum* which records the number of blocks of every possible size in the coalescent $\Pi^n(t)$. In order to say anything meaningful, we have to consider special cases – firstly, we shall study Kingman's n-coalescent in Section 3.5.1 and derive a new formula for the falling factorial moments for its block size spectrum. Then, in Section 3.5.2, we shall discuss a rescaling of this spectrum whose limit coincides with the solution of the Smoluchowski coagulation equation [Smoluchowski, 1916] with an appropriate kernel. We shall refer to this as the hydrodynamic limit of the spectrum. There is a striking similarity between the expected value of the number of blocks of size j in the finite spectrum and the corresponding solution of the Smoluchowski equation.

In Section 3.5.3, we perform a similar exercise but for the generalized Beta coalescent, driven by $\Lambda_{a,b}$ defined in (3.11), deriving a system of ordinary differential equations which the deterministic hydrodynamic limit of the block size spectrum satisfies. Naturally, the time-rescaling that is required to obtain the limiting process depends on the parameters of the underlying coalescent, and is different to that of the Kingman coalescent. The solution of the system, Theorem 3.18, has a rather succinct form that is a complete Bell polynomial. It generalizes the solution of the Smoluchowski coagulation equations to the setting of the Beta coalescent.

All the notation that we use in this section will be defined formally, but it may be useful to understand the conventions used. The letter 'c' will be used to denote the fact that we are dealing with counting processes. A capital will be used to refer to random processes (which are, ordinarily, pre-limiting versions), whilst a lower-case letter will be used to denote the corresponding deterministic limits. A process without an index refers to the block-counting process, but a subscript indicates that the process counts

blocks of that size. The superscript will be used to specify which coalescent we are working with.

This is joint work with Helmut Pitters. To the best of our knowledge, the results in Sections 3.5.1 and 3.5.3 are novel.

3.5.1 Kingman's n-Coalescent

In his seminal paper, Kingman [1982] showed that the jump chain (i.e. the sequence of partitions) of the n-coalescent and its block-counting process, $C^K(t)$, are independent. (The block-counting process was defined on p.30, but was denoted K_t there to be consistent with other notation.) Here, we will show that certain moments of the block size spectrum depend exclusively on $C^K(t)$. Before doing so, we need some notation.

Definition 3.8. For any partition $\pi \in \mathscr{P}_{[n]}$, let the operator $\mathfrak{c}_i\pi := |\{B \in \pi : |B| = i\}|$ count the number of blocks of size i in π. The *block size spectrum* of π is the vector $\mathfrak{c}\pi = (\mathfrak{c}_1\pi, \ldots, \mathfrak{c}_n\pi)$. Since the Kingman coalescent $\Pi_K^n(t)$ is a partition-valued process, we may further define $C_i^K(t) := \mathfrak{c}_i\Pi_K^n(t)$ along with $\underline{\boldsymbol{C}}^K(t) := (C_1^K(t), \ldots, C_n^K(t))$, the block size spectrum of the n-coalescent.

Remark 3.9. Whilst $\underline{\boldsymbol{C}}^K(t)$ summarizes the details of $\Pi_K^n(t)$, it certainly contains more information than the block-counting process $C^K(t)$. We underline the vector $\underline{\boldsymbol{C}}^K(t)$ to clearly distinguish it from $C^K(t)$.

Theorem 3.10. *For any vector $\boldsymbol{k} = (k_1, \ldots, k_n) \in \mathbb{N}_0^n$, the joint falling factorial moments of $\mathfrak{c}\Pi_K^n(t)$ are given by*

$$\mathbb{E}\left[C_1^K(t)^{\underline{k_1}} \cdots C_n^K(t)^{\underline{k_n}}\right] = \frac{\mathbb{E}\left[C^K(t)^{\underline{|\boldsymbol{k}|}} \left(C^K(t)-1\right)^{\underline{|\boldsymbol{k}|}} \left(n - C^K(t)\right)^{\underline{|\boldsymbol{k}|_w - |\boldsymbol{k}|}}\right]}{(n-1)^{\underline{|\boldsymbol{k}|_w}}}, \tag{3.12}$$

where $|\boldsymbol{k}|$ is the L^1-norm defined in Appendix A, $|\boldsymbol{k}|_w := \sum_{i=1}^n ik_i$ is the weighted norm of $\boldsymbol{k}$ and $z^{\underline{k}}$ is the kth falling factorial power of z.

Proof. Firstly, consider the restricted expectation

$$\begin{aligned}
&\mathbb{E}[C_1^K(t)^{\underline{k_1}} \cdots C_n^K(t)^{\underline{k_n}} \mathbb{1}_{\{|\Pi_K^n(t)|=b\}}] \\
&\quad = \sum_{\pi:|\pi|=b} (\mathfrak{c}_1\pi)^{\underline{k_1}} \cdots (\mathfrak{c}_n\pi)^{\underline{k_n}} \, \mathbb{P}\left(\Pi_K^n(t) = \pi \middle| |\pi| = b\right) \\
&\quad = \sum_{\pi:|\pi|=b} (\mathfrak{c}_1\pi)^{\underline{k_1}} \cdots (\mathfrak{c}_n\pi)^{\underline{k_n}} \, \mathbb{P}(C^K(t) = b) \frac{(n-b)!b!(b-1)!}{n!(n-1)!} \prod_{i=1}^{n} (i!)^{\mathfrak{c}_i\pi} .
\end{aligned}$$

The final equality essentially makes use of the identities (2.3) and (2.5) in Kingman [1982], but reformulated in terms of the spectrum notation. We see that the summands depend only upon the block size spectrum, $\mathfrak{c}\pi$. There are of course many partitions with such a spectrum. Referring to Appendix B, the number of partitions π satisfying $\mathfrak{c}_i\pi = m_i,\ i \in [n]$, with $|\boldsymbol{m}| = b$ and $|\boldsymbol{m}|_w = n$, is

$$\frac{n!}{\prod_{i=1}^{n}(i!)^{m_i} m_i!}. \tag{3.13}$$

Recalling the definition of the Lah number (Appendix B),

$$\begin{aligned}
&\mathbb{E}[C_1^K(t)^{\underline{k_1}} \cdots C_n^K(t)^{\underline{k_n}} \mathbb{1}_{\{|\Pi_K^n(t)|=b\}}] \\
&\quad = \frac{\mathbb{P}(C^K(t) = b)}{L_{n,b}} \sum_{\substack{m_1 \geq 0, \ldots, m_n \geq 0; \\ |\boldsymbol{m}|=b, |\boldsymbol{m}|_w = n}} \frac{n!}{\prod_{i=1}^{n}(i!)^{m_i} m_i!} \prod_{i=1}^{n} (m_i)^{\underline{k_i}} (i!)^{m_i} \\
&\quad = \frac{n! \, \mathbb{P}(C^K(t) = b)}{L_{n,b}} \sum_{\substack{m_1 \geq k_1, \ldots, m_n \geq k_n; \\ |\boldsymbol{m}|=b, |\boldsymbol{m}|_w = n}} \prod_{i=1}^{n} \frac{1}{(m_i - k_i)!} \\
&\quad = \frac{n! \, \mathbb{P}(C^K(t) = b)}{L_{n,b}} \sum_{\substack{\tilde{m}_1 \geq 0, \ldots, \tilde{m}_n \geq 0; \\ |\tilde{\boldsymbol{m}}|=b-|\boldsymbol{k}|, |\tilde{\boldsymbol{m}}|_w = n-|\boldsymbol{k}|_w}} \prod_{i=1}^{n} \frac{1}{\tilde{m}_i!} \\
&\quad = \frac{(n)^{\underline{|\boldsymbol{k}|_w}} \, \mathbb{P}(C^K(t) = b)}{L_{n,b}} \sum_{\substack{\tilde{m}_1 \geq 0, \ldots, \tilde{m}_n \geq 0; \\ |\tilde{\boldsymbol{m}}|=b-|\boldsymbol{k}|, |\tilde{\boldsymbol{m}}|_w = n-|\boldsymbol{k}|_w}} \frac{(n-|\boldsymbol{k}|_w)!}{\prod_{i=1}^{n}(i!)^{\tilde{m}_i} \tilde{m}_i!} \prod_{i=1}^{n} (i!)^{\tilde{m}_i}.
\end{aligned}$$

In the second equality, we recognized some cancellations that could be made and in the third we shifted the indices of the summation so that $\tilde{m}_i = m_i - k_i$. In the final equality, we split $n!$ in such a way

that we can make use of Equation (3.13) once more. The prefactor to the product inside the sum is the number of partitions $\tilde{\pi}$ in $\mathscr{P}_{[n-|\boldsymbol{k}|_w]}$ with $b-|\boldsymbol{k}|$ blocks in which there are $\tilde{m}_i$ of size $i \in [n]$. In fact, the summation is the partial Bell polynomial $B_{n-|\boldsymbol{k}|_w,b-|\boldsymbol{k}|}(\bullet!)$. (Bell polynomials are reviewed in Appendix B.) Such a polynomial is a special Lah number, $L_{n-|\boldsymbol{k}|_w,b-|\boldsymbol{k}|}$. This allows us to perform some drastic cancellations involving factorials and falling factorials, which overall yield

$$\begin{aligned}\mathbb{E}[C_1^K(t)^{\underline{k_1}} \cdots C_n^K(t)^{\underline{k_n}} \mathbb{1}_{\{|\Pi_K^n(t)|=b\}}] &= (n)^{\underline{|\boldsymbol{k}|_w}}\ \mathbb{P}(C^K(t)=b)\frac{L_{n-|\boldsymbol{k}|_w,b-|\boldsymbol{k}|}}{L_{n,b}} \\ &= \frac{\mathbb{P}(C^K(t)=b)}{(n-1)^{\underline{|\boldsymbol{k}|_w}}}\,(b)^{\underline{|\boldsymbol{k}|}}\,(b-1)^{\underline{|\boldsymbol{k}|}}\,(n-b)^{\underline{|\boldsymbol{k}|_w-|\boldsymbol{k}|}}.\end{aligned}$$

It only remains to partition over b to obtain the final result. □

Griffiths [2006] derives the formula for the falling factorial moments for $C^K(t)$. This means, for example, that $\mathbb{E}[C_1^K(t)] = (n-1)^{-1}\,\mathbb{E}[C^K(t)^{\underline{2}}]$ has an immediate explicit form. In the sequel, we will show that (3.12) can be written in terms of falling factorial moments of $C^K(t)$ for all $\boldsymbol{k} \in \mathbb{N}_0^n$, but before doing so, we shall derive an explicit result for the first moments of $C_j^K(t),\ j \in [n]$.

Corollary 3.11. *For $j \in [n]$,*

$$\mathbb{E}[C_j^K(t)] = \sum_{k=0}^{j-1}(-1)^k \binom{j-1}{k} \frac{\mathbb{E}[C^K(t)^{\underline{k+2}}]}{(n-1)^{\underline{k+1}}}. \tag{3.14}$$

Proof. By fixing j and putting $\boldsymbol{k} = \boldsymbol{e}_j$, we obtain

$$\begin{aligned}\mathbb{E}[C_j^K(t)] &= \left((n-1)^{\underline{j}}\right)^{-1} \mathbb{E}\left[C^K(t)(C^K(t)-1)\left(n-C^K(t)\right)^{\underline{j-1}}\right] \\ &= \left((n-1)^{\underline{j}}\right)^{-1} \mathbb{E}\left[C^K(t)^{\underline{2}}\left(n-C^K(t)\right)^{\underline{j-1}}\right].\end{aligned} \tag{3.15}$$

Combining (3.15) with an application of Lemma C.3.1 with $a=2$, $b=j-1$ and $\zeta = C^K(t)$ yields

$$\mathbb{E}[C_j^K(t)] = \frac{1}{(n-1)^{\underline{j}}} \sum_{k=0}^{j-1}(-1)^k \binom{j-1}{k} \frac{(n-(k+2))!}{(n-(j+1))!}\mathbb{E}[C^K(t)^{\underline{k+2}}]. \tag{3.16}$$

It is then just a matter of manipulating the factorials and falling factorial terms. □

This result applies to the case $|\boldsymbol{k}| = 1$. It is possible to go one step further and make Theorem 3.10 explicit in the falling factorial moments of $C^K(t)$ when $|\boldsymbol{k}| > 1$.

Lemma 3.12. *For $\boldsymbol{k} \in \mathbb{N}_0^n$ satisfying $|\boldsymbol{k}| > 1$,*

$$\mathbb{E}\left[C^K(t)^{\underline{|\boldsymbol{k}|}}\left(C^K(t)-1\right)^{\underline{|\boldsymbol{k}|}}\left(n-C^K(t)\right)^{\underline{|\boldsymbol{k}|_w-|\boldsymbol{k}|}}\right] = \sum_{l=0}^{|\boldsymbol{k}|-1}\binom{|\boldsymbol{k}|}{l}(|\boldsymbol{k}|-1)^{\underline{l}}\sum_{m=0}^{|\boldsymbol{k}|_w-|\boldsymbol{k}|}(-1)^m\binom{|\boldsymbol{k}|_w-|\boldsymbol{k}|}{m}\frac{(n-(m+2|\boldsymbol{k}|-l))!}{(n-(|\boldsymbol{k}|_w+|\boldsymbol{k}|-l))!}\mathbb{E}[C^K(t)^{\underline{m+2|\boldsymbol{k}|-l}}]. \tag{3.17}$$

Remark 3.13. This result can be combined with Theorem 3.10 in order to simplify the expression for the falling factorial moments of $\underline{\boldsymbol{C}}^K(t)$. However, note that the prefactor $\left((n-1)^{\underline{|\boldsymbol{k}|_w}}\right)^{-1}$ is not included in Lemma 3.12.

Proof. We shall make use of Lemma A.2 in Etheridge *et al.* [2010], which states that

$$\zeta^{\underline{a}}(\zeta+c)^{\underline{b}} = \sum_{l=0}^{b}\zeta^{\underline{a+b-l}}(b)^{\underline{l}}\binom{a+c}{l}.$$

According to that result, we may only use $c \in \mathbb{Z}_+$, but the result also works if we choose $a = |\boldsymbol{k}|$ and $c = -1$, provided we stipulate that $|\boldsymbol{k}| > 1$. This allows us to expand the binomial coefficient in the sequel. It follows that

$$\zeta^{\underline{|\boldsymbol{k}|}}(\zeta-1)^{\underline{|\boldsymbol{k}|}} = \sum_{l=0}^{|\boldsymbol{k}|}\binom{|\boldsymbol{k}|}{l}(|\boldsymbol{k}|-1)^{\underline{l}}\zeta^{\underline{2|\boldsymbol{k}|-l}}.$$

Note that the $l = |\boldsymbol{k}|$ term in this sum vanishes due to the first falling factorial in $|\boldsymbol{k}|-1$. For compactness, define $|\boldsymbol{k}|_* := |\boldsymbol{k}|_w - |\boldsymbol{k}|$. Using this alongside Lemma C.3.1 with $a = 2|\boldsymbol{k}| - l$ and $b = |\boldsymbol{k}|_*$ yields the expression,

$$\zeta^{\underline{|\boldsymbol{k}|}}(\zeta-1)^{\underline{|\boldsymbol{k}|}}(n-\zeta)^{\underline{|\boldsymbol{k}|_*}} = \sum_{l=0}^{|\boldsymbol{k}|-1}\binom{|\boldsymbol{k}|}{l}(|\boldsymbol{k}|-1)^{\underline{l}}\sum_{m=0}^{|\boldsymbol{k}|_*}(-1)^m\binom{|\boldsymbol{k}|_*}{m}\frac{(n-(m+2|\boldsymbol{k}|-l))!}{(n-(|\boldsymbol{k}|_w+|\boldsymbol{k}|-l))!}\zeta^{\underline{m+2|\boldsymbol{k}|-l}},$$

from which the result easily follows. □

3.5.2 Rescaling the n-Coalescent

There is a rescaling of Kingman's coalescent which will allow us to approximate the number of blocks as a deterministic function as $n \to \infty$. Define

$$c^K(t) := \lim_{n\to\infty} \frac{C^K(n^{-1}t)}{n}, \tag{3.18}$$

which converges in the uniform topology. The intuition behind the rescaling is that we look at the *relative* number of blocks as we *slow* time down by a factor of n. The limit describes the manner in which the block-counting process behaves as the Kingman coalescent comes down from infinity.

Remark 3.14. There is a more general rescaling. For any $\gamma > 0$, we can divide by n^γ and correspondingly slow time down by a factor n^γ. This yields the same limit $c^K(t)$, which is independent of γ. Equation (3.18) is the special case $\gamma = 1$.

It can be shown (in the review paper of Aldous [1999] for example) that this hydrodynamic limit satisfies the ordinary differential equation (ODE)

$$\frac{d}{dt}c^K(t) + \frac{1}{2}c^K(t)^2 = 0, \quad c^K(0) = 1, \tag{3.19}$$

whose unique solution is

$$c^K(t) = \frac{2}{2+t}. \tag{3.20}$$

Correspondingly, define

$$c_j^K(t) := \lim_{n\to\infty} \frac{C_j^K(n^{-1}t)}{n}.$$

These quantities behave according to the system of ODEs

$$\frac{d}{dt}c_j^K(t) + c_j^K(t)\sum_{i=1}^{\infty} c_i^K(t) = \frac{1}{2}\sum_{i=1}^{j-1} c_i^K(t)c_{j-i}^K(t), \quad c_j^K(0) = \mathbb{1}\{j=1\}, \quad j \in \mathbb{N}. \tag{3.21}$$

One can easily verify that the solution to this system is

$$c_j^K(t) = c^K(t)^2(1-c^K(t))^{j-1} = \sum_{k=0}^{j-1}(-1)^k \binom{j-1}{k} c^K(t)^{k+2}. \tag{3.22}$$

This solution can be found in *Theorie der Koagulation* of Smoluchowski [1916], along with a hand-

drawn image that very much resembles Figure 3.5. It is worthwhile comparing the similarities between the summation form of (3.22) and its pre-limiting analogue (3.14). The ODE system (3.21) is a special case of what is now known as the Smoluchowski coagulation equation – see Aldous [1999] for an excellent review of this topic.

One property of these curves that is immediately clear from (3.22) is that they are decreasing in j for all $t > 0$. That is, $c^K(t) > c_1^K(t) > c_2^K(t) > c_3^K(t) > \dots, \forall t > 0$. This stems from the fact that $c^K(t) \in (0,1)$ and the factor $(1-c^K(t))^{j-1}$ in the solution. One other question that we may address is that concerning the maxima of the curves. Using basic differentiation, it is not difficult to show that, for each $j \in \mathbb{N}$,

$$\arg\max_{t\geq 0} c_j^K(t) = j-1, \qquad \max_{t\geq 0} c_j^K(t) = 4\frac{(j-1)^{j-1}}{(j+1)^{j+1}},$$

where we use the convention $0^0 = 1$ for the case $j = 1$. These properties are shown pictorially for small values of j in Figure 3.5. One final property that we shall mention here is that the fraction of blocks of size *at least* j is given by

$$\sum_{i=j}^{\infty} c_i^K(t) = c^K(t)\left(1 - c^K(t)\sum_{l=0}^{j-2}(1-c^K(t))^l\right) = c^K(t)\left(1-c^K(t)\right)^{j-1} = \frac{c_j^K(t)}{c^K(t)}.$$

One method of finding the solution to (3.21) is the following. Define the generating function

$$G^K(x,t) := \sum_{i=1}^{\infty} c_i^K(t)x^i \tag{3.23}$$

which converges at least for $|x| \leq 1$ since $0 < c_i^K(t) < 1$ for all positive times. One can use this to derive the differential equation

$$\frac{\partial}{\partial t}G^K(x,t) = \frac{1}{2}G^K(x,t)^2 - c^K(t)G^K(x,t), \quad G^K(x,0) = x, \tag{3.24}$$

using the fact that $G^K(1,t) = c^K(t)$ for all $t \geq 0$. Whilst at first sight this appears to be a *partial* differential equation, there is no derivative in the x-variable, allowing it to be treated as an ordinary differential equation for each x. In view of this, (3.24) is simply of Ricatti type, and has the solution

$$G^K(x,t) = \frac{xc^K(t)^2}{1-x(1-c^K(t))}. \tag{3.25}$$

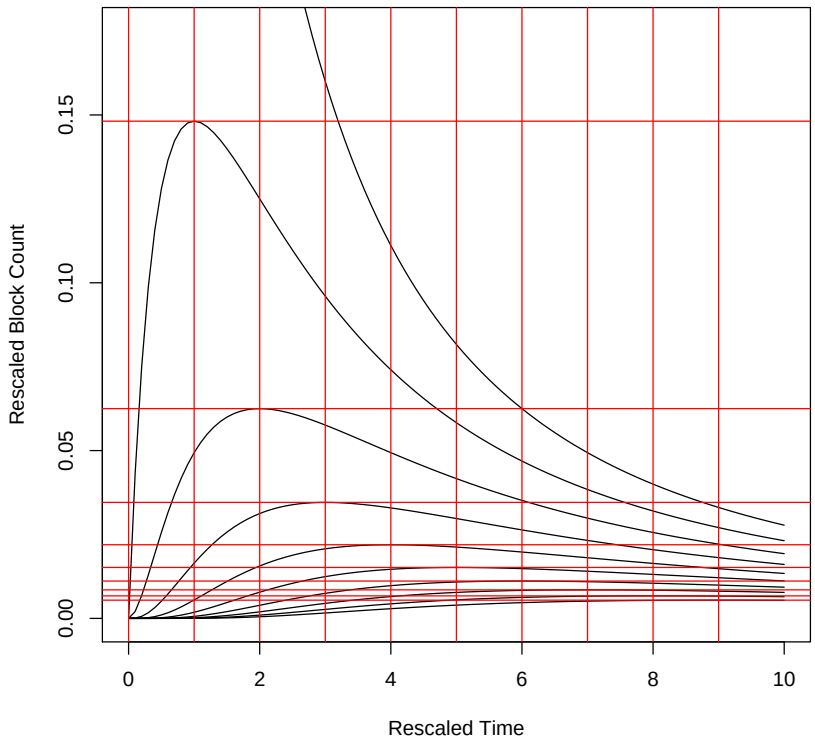


Figure 3.5: The rescaled curves $c_j^K(t)$ for $j \in [10]$ ($c_1^K(t)$ is shown only partially). Each curve may be identified by using the monotonicity property. The vertical red lines represent the location of a maximum, whilst the horizontal ones represent the corresponding value (to find which value of j they correspond to, simply count from left to right, or from top to bottom, noting that the horizontal line for $j = 1$ is not depicted)

One can then obtain $c_i^K(t)$ by differentiating $G^K(x,t)$ i times in x, dividing by $i!$, and then setting $x = 0$. Whilst this is a nice application of the generating function methodology, one might ask why this was necessary if the answer was already known. In fact, this turns out to be just the right way of thinking to solve the harder problem in Section 3.5.3.

What is even more remarkable about the equation for the generating function G^K is that, if one makes the simple transformation $H^K(x,t) = c^K(t) - G^K(x,t)$, then we obtain almost the same equation as (3.19), the only differences being the x-dependence and the in initial condition. In the next section, we shall observe the same phenomenon that, with a simple transformation, the equation for the generating function of the spectral elements mimics that of the block-counting process.

3.5.3 Beta Coalescent Spectra

In this section we shall consider the generalized Beta coalescent, and study its hydrodynamic limit as the number of leaves tend to infinity in a similar fashion to the case of the Kingman coalescent. Throughout, we shall take Λ to be the two-parameter Beta probability measure, $\Lambda_{a,b}(du)$, defined in (3.11). Define its block-counting process to be $C^{a,b}(t)$, which implicitly depends on the number of leaves n. It follows from (3.5) and (3.11) that

$$\lambda_{n,k} := \int_0^1 x^{k-2}(1-x)^{n-k}\Lambda_{a,b}(dx) = \frac{B(k-2+a, n-k+b)}{B(a,b)} = \frac{a^{\overline{k-2}}b^{\overline{n-k}}}{(a+b)^{\overline{n-2}}}. \tag{3.26}$$

We shall denote the rescaled block-counting process $c^n(t) = n^{-1}C^{a,b}(\tau_n t)$, where $\tau_n = n^{a-1}$, a process supported on $\{\frac{1}{n}, \frac{2}{n}, \ldots, 1\}$ for all $n \in \mathbb{N}$. The quantity τ_n is the amount by which we have to slow time down in order to study the way that Beta coalescent comes down from infinity – notice that it is not as slow as for the Kingman coalescent, in which $a = 0$. This will only work for the set of Beta coalescents which do in fact come down from infinity, i.e. those which satisfy $a < 1$. We shall enforce this condition for the remainder of this section.

It follows from (3.6) that $c^n(t)$ makes the transition

$$c \mapsto c - \frac{l-1}{n}, \quad 2 \le l \le nc, \tag{3.27}$$

at rate $\tau_n \binom{nc}{l} \lambda_{nc,l}$. We would like to find a diffusion approximation for this process by rescaling the generators. Let $\mathscr{L}^n_{BCP}$ be the generator of $c^n(t)$. As the following theorem states, the limiting generator can be seen as that corresponding to a stochastic differential equation – it is of the form (2.2) – with no noise term.

Theorem 3.15. *Let $a \in (0,1)$, $\tau_n = n^{a-1}$ and $f \in \mathscr{D}(\mathscr{L}^n_{BCP})$ with continuous second derivative. Then, as $n \to \infty$, $\mathscr{L}^n_{BCP} f(c) \to \mathscr{L}_{BCP} f(c)$, where*

$$\mathscr{L}_{BCP} f(c) = -K c^{2-a} f'(c), \tag{3.28}$$

and

$$K := \frac{\Gamma(a+b)}{(a-1)(a-2)\Gamma(b)}. \tag{3.29}$$

Proof. The generator of $c^n(t)$ is given by

$$\begin{aligned}\mathscr{L}^n_{BCP} f(c) &= \tau_n \sum_{l=2}^{nc} \left(f\left(c - \frac{l-1}{n}\right) - f(c) \right) \binom{nc}{l} \lambda_{nc,l} \\ &= \tau_n \sum_{l=2}^{nc} \left(-\frac{l-1}{n} f'(c) + R_2(x_{n,l}) \right) \binom{nc}{l} \lambda_{nc,l},\end{aligned}$$

where we used Taylor's approximation in the second equality. Taylor's approximation ensures the existence of a value $x_{n,l} \in (c-(l-1)/n, c)$ such that the remainder term $R_2(x_{n,l})$ is given, for instance, by its Lagrange form

$$R_2(x_{n,l}) = \frac{1}{2}\left(\frac{l-1}{n}\right)^2 f''(x_{n,l}). \tag{3.30}$$

For $k \in \mathbb{N}$ define

$$\gamma_n^{(k)} := \sum_{l=2}^{n} \binom{n}{l} (l-1)^k \lambda_{n,l}. \tag{3.31}$$

First notice that by Lemma C.3.3

$$-\frac{\tau_n}{n} \sum_{l=2}^{nc} \binom{nc}{l} (l-1) \lambda_{nc,l} = -\frac{\tau_n}{n} \gamma_{nc}^{(1)} \to -Kc^{2-a}$$

as $n \to \infty$, since τ_n is chosen to be of order n^{a-1}, and $a < 1$ by assumption. (Lemma C.3.3 fails for $a \geq 1$, so this assumption is necessary.) Since f is twice continuously differentiable on $[0,1]$, f'' attains its supremum $||f''||_\infty := \sup_{x\in[0,1]} |f''(x)|$. Consequently, as $n \to \infty$ we obtain

$$\begin{aligned}\tau_n \sum_{l=2}^{nc} \binom{nc}{l} \lambda_{nc,l} R_2(x_{n,l}) &\leq \frac{\tau_n}{2n^2} ||f''||_\infty \sum_{l=2}^{nc} \binom{nc}{l} \lambda_{nc,l} (l-1)^2 \\ &= \frac{\tau_n}{2n^2} ||f''||_\infty \gamma_{nc}^{(2)} \sim \frac{1}{2} \tau_n ||f''||_\infty \to 0,\end{aligned}$$

where we used Lemma C.3.4 in the third step. Note that $f \sim g$ as $n \to \infty$ if and only if $\lim_{n\to\infty} \frac{f(n)}{g(n)} = 1$, as defined in Appendix A. □

Consequently, for $a < 1$, the diffusion limit is in fact deterministic and the limiting process $c^{a,b}(t)$ solves the ODE (see Example 2.2)

$$\frac{d}{dt}c^{a,b}(t) = -Kc^{a,b}(t)^{2-a} \tag{3.32}$$

with boundary condition $c^{a,b}(0) = 1$. This ODE is separable and of Bernoulli type, and consequently has solution

$$c^{a,b}(t) = \left(1 - \frac{\Gamma(a+b)}{(a-2)\Gamma(b)}t\right)^{-\frac{1}{1-a}} = \left(\frac{(2-a)\Gamma(b)}{(2-a)\Gamma(b) + \Gamma(a+b)t}\right)^{\frac{1}{1-a}}. \tag{3.33}$$

Remark 3.16. To recover the solution for the ordinary Beta coalescent, we just set $a = 2 - \alpha$ and $b = \alpha$ for $\alpha \in (1, 2)$, for which (3.33) becomes

$$c^{2-\alpha,\alpha}(t) = \left(1 + \frac{t}{\Gamma(\alpha+1)}\right)^{-\frac{1}{\alpha-1}}.$$

From here, we can take the limit $\alpha \uparrow 2$ in order to recover $c^K(t)$ from the Kingman case.

Having established this result, we would now like to understand the behaviour of the block size spectrum, denoted $\underline{\boldsymbol{C}}^{a,b}(t) := \mathfrak{c}\Pi^n_{a,b}(t) := (C^{a,b}_1(t), C^{a,b}_2(t), \ldots, C^{a,b}_n(t))$, where $\Pi^n_{a,b}(t)$ is the partition-valued generalized Beta coalescent. To do so, choose some fixed positive $d \leq n$ and, for $i \in [d]$, define

$$c^n_i(t) := n^{-1}C^{a,b}_i(\tau_n t), \tag{3.34}$$

where $\tau_n = n^{a-1}$ as before. In addition, we define an extra dimension,

$$c^n_{d+1}(t) := n^{-1}\sum_{i=d+1}^{n} C^{a,b}_i(\tau_n t), \tag{3.35}$$

which accounts for all the blocks which have size greater than d. Finally, if we define the $(d+1)$-dimensional vector $\underline{\boldsymbol{c}}^n(t) = (c^n_1(t), c^n_2(t), \ldots, c^n_d(t), c^n_{d+1}(t))$, then the relationship $n|\underline{\boldsymbol{c}}^n(t)| = nc^n(t) = C^{a,b}(\tau_n t)$, naturally follows for all $t \geq 0$.

For $\boldsymbol{l} \in \mathbb{N}^{d+1}_0$ with $|\boldsymbol{l}| > 1$, we say that an $\boldsymbol{l}$-merger occurs in $\Pi^n_{a,b}$ if among the merging blocks there are l_1 singletons, l_2 blocks of size 2, ..., l_d blocks of size d and l_{d+1} blocks of size *at least* $d+1$. The

process $\underline{\boldsymbol{c}}^n(t)$ makes the transition

$$\boldsymbol{c} \mapsto \boldsymbol{c} - \frac{\boldsymbol{l} - \boldsymbol{e}_{|\boldsymbol{l}|_w \wedge (d+1)}}{n} \qquad \text{at rate} \qquad \lambda_{n|\boldsymbol{c}|,|\boldsymbol{l}|} \prod_{i=1}^{d+1} \binom{nc_i}{l_i},$$

where $|\boldsymbol{l}|_w$ is the weighted norm defined in Appendix A. This transition occurs because the size of the new block resulting from the merger is the weighted norm of $\boldsymbol{l}$ since l_i blocks of size i coalesce – if this total is greater than d, then the new block just goes into the extra dimension $d+1$. The division by n follows from the rescaling of space made in (3.34). To understand the rate, note that $\binom{nc_i}{l_i}$ is the number of ways that l_i blocks of size i could be chosen, and the product is the number of ways of choosing $\boldsymbol{l}$ from $n\boldsymbol{c}$ – each of these occur at rate $\lambda_{n|\boldsymbol{c}|,|\boldsymbol{l}|}$.

Now, let $\mathscr{L}^n_{BSS}$ denote the generator of $\underline{\boldsymbol{c}}^n$, acting on $(d+1)$-dimensional functions, and let $\partial_i = \frac{\partial}{\partial x_i}$ denote the ith partial derivative.

Proposition 3.17. *Let $a \in (0,1), \tau_n = n^{a-1}$ and $f \in \mathscr{D}(\mathscr{L}^n_{BSS})$ with continuous second derivative. Then, as $n \to \infty$, $\mathscr{L}^n_{BSS} f(\boldsymbol{c}) \to \mathscr{L}_{BSS} f(\boldsymbol{c})$, where*

$$\begin{aligned}\mathscr{L}_{BSS} f(\boldsymbol{c}) = \frac{\Gamma(a+b)}{\Gamma(b)} \sum_{i=1}^{d} \left(\frac{c_i |\boldsymbol{c}|^{1-a}}{a-1} + \sum_{m=2}^{i} a^{\overline{m-2}} |\boldsymbol{c}|^{2-a-m} \sum_{\substack{\boldsymbol{l} \in \mathbb{N}_0^{d+1} \\ |\boldsymbol{l}|=m, |\boldsymbol{l}|_w = i}} \prod_{k=1}^{d+1} \frac{c_k^{l_k}}{l_k!} \right) \partial_i f(\boldsymbol{c}) \\ - \frac{\Gamma(a+b)}{\Gamma(b)} \left(\frac{|\boldsymbol{c}|^{2-a}}{a-2} - \frac{c_{d+1}|\boldsymbol{c}|^{1-a}}{a-1} + \sum_{r=1}^{d} \sum_{m=2}^{r} a^{\overline{m-2}} |\boldsymbol{c}|^{2-a-m} \sum_{\substack{\boldsymbol{l} \in \mathbb{N}_0^{d+1} \\ |\boldsymbol{l}|=m, |\boldsymbol{l}|_w = r}} \prod_{k=1}^{d+1} \frac{c_k^{l_k}}{l_k!} \right) \partial_{d+1} f(\boldsymbol{c}).\end{aligned} \tag{3.36}$$

Proof. For a function $f : \mathbb{R}^{d+1} \to \mathbb{R}$ and vector $\boldsymbol{\kappa} \in \mathbb{N}_0^{d+1}$, let $D^{\boldsymbol{\kappa}} f(x) := \partial_1^{\kappa_1} \cdots \partial_{d+1}^{\kappa_{d+1}} f(x)$. Moreover, let $\boldsymbol{\kappa}! := \prod_{i=1}^{d+1} \kappa_i!$ and for any vector $\boldsymbol{y} \in \mathbb{R}^{d+1}$, define $\boldsymbol{y}^{\boldsymbol{\kappa}} := \prod_{i=1}^{d+1} y_i^{\kappa_i}$. For two vectors $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{N}_0^{d+1}$, the inequality $\boldsymbol{x} \leq \boldsymbol{y}$ is satisfied if and only if $x_i \leq y_i$ for all $i \in [d+1]$. A Taylor expansion yields

$$\begin{aligned}\mathscr{L}^n_{BSS}f(\boldsymbol{c}) &= \tau_n \sum_{\substack{\boldsymbol{l}\in\mathbb{N}_0^{d+1}\\ |\boldsymbol{l}|>1,\, \boldsymbol{l}\le n\boldsymbol{c}}} \left(f\left(\boldsymbol{c}-\frac{\boldsymbol{l}-\boldsymbol{e}_{|\boldsymbol{l}|_w\wedge(d+1)}}{n}\right)-f(\boldsymbol{c})\right)\lambda_{n|\boldsymbol{c}|,|\boldsymbol{l}|}\prod_{i=1}^{d+1}\binom{nc_i}{l_i}\\ &= \tau_n \sum_{\substack{\boldsymbol{l}\in\mathbb{N}_0^{d+1}\\ |\boldsymbol{l}|>1}}\Bigg(-\sum_{\substack{\boldsymbol{\kappa}\in\mathbb{N}_0^{d+1}\\ |\boldsymbol{\kappa}|=1}}\frac{(\boldsymbol{l}-\boldsymbol{e}_{|\boldsymbol{l}|_w\wedge(d+1)})^{\boldsymbol{\kappa}}}{n}D^{\boldsymbol{\kappa}}f(\boldsymbol{c}) \qquad (3.37)\\ &\quad+\sum_{\substack{\boldsymbol{\kappa}\in\mathbb{N}_0^{d+1}\\ |\boldsymbol{\kappa}|=2}}\frac{(\boldsymbol{l}-\boldsymbol{e}_{|\boldsymbol{l}|_w\wedge(d+1)})^{\boldsymbol{\kappa}}}{n^2\boldsymbol{\kappa}!}D^{\boldsymbol{\kappa}}f\left(\boldsymbol{c}-\vartheta_{n,\boldsymbol{l}}\frac{\boldsymbol{l}-\boldsymbol{e}_{|\boldsymbol{l}|_w\wedge(d+1)}}{n}\right)\Bigg)\lambda_{n|\boldsymbol{c}|,|\boldsymbol{l}|}\prod_{i=1}^{d+1}\binom{nc_i}{l_i},\end{aligned}$$

where we have used the Lagrange form of the remainder for some $\vartheta_{n,\boldsymbol{l}}\in[0,1]$.

Part 1. Let us first consider summands corresponding to $|\boldsymbol{\kappa}|=2$ by focusing on

$$T^{(2)}(n) := \frac{\tau_n}{n^2}\sum_{\substack{\boldsymbol{l}\in\mathbb{N}_0^{d+1}\\ |\boldsymbol{l}|>1}}\sum_{\substack{\boldsymbol{\kappa}\in\mathbb{N}_0^{d+1}\\ |\boldsymbol{\kappa}|=2}}\frac{(\boldsymbol{l}-\boldsymbol{e}_{|\boldsymbol{l}|_w\wedge(d+1)})^{\boldsymbol{\kappa}}}{\boldsymbol{\kappa}!}D^{\boldsymbol{\kappa}}f\left(\boldsymbol{c}-\vartheta_{n,\boldsymbol{l}}\frac{\boldsymbol{l}-\boldsymbol{e}_{|\boldsymbol{l}|_w\wedge(d+1)}}{n}\right)\lambda_{n|\boldsymbol{c}|,|\boldsymbol{l}|}\prod_{i=1}^{d+1}\binom{nc_i}{l_i}, \qquad (3.38)$$

Evidently, in this case, there exist (possibly equal) $i,j\in[d+1]$ with $\boldsymbol{\kappa}=\boldsymbol{e}_i+\boldsymbol{e}_j$. Notice that

$$(\boldsymbol{l}-\boldsymbol{e}_{|\boldsymbol{l}|_w\wedge(d+1)})^{\boldsymbol{\kappa}}\le\boldsymbol{l}^{\boldsymbol{\kappa}}=\begin{cases} l_i^2 & \text{if } \boldsymbol{\kappa}=2\boldsymbol{e}_i, i\in[d+1],\\ l_il_j & \text{if } \boldsymbol{\kappa}=\boldsymbol{e}_i+\boldsymbol{e}_j, i,j\in[d+1], i\neq j.\end{cases}$$

Hence, for fixed $i\in[d+1]$ and $\boldsymbol{\kappa}=2\boldsymbol{e}_i$,

$$\begin{aligned}&\tau_n\sum_{\substack{\boldsymbol{l}\in\mathbb{N}_0^{d+1}\\ |\boldsymbol{l}|>1}}\frac{1}{2n^2}(\boldsymbol{l}-\boldsymbol{e}_{|\boldsymbol{l}|_w\wedge(d+1)})^{2\boldsymbol{e}_i}D^{2\boldsymbol{e}_i}f\left(\boldsymbol{c}-\vartheta_{n,\boldsymbol{l}}\frac{\boldsymbol{l}-\boldsymbol{e}_{|\boldsymbol{l}|_w\wedge(d+1)}}{n}\right)\lambda_{n|\boldsymbol{c}|,|\boldsymbol{l}|}\prod_{k=1}^{d+1}\binom{nc_k}{l_k}\\ &\quad\le||D^{2\boldsymbol{e}_i}f||_\infty\frac{\tau_n}{2n^2}\sum_{\substack{\boldsymbol{l}\in\mathbb{N}_0^{d+1}\\ |\boldsymbol{l}|>1}}l_i^2\lambda_{n|\boldsymbol{c}|,|\boldsymbol{l}|}\prod_{k=1}^{d+1}\binom{nc_k}{l_k}\\ &\quad=\frac{1}{2}\frac{||D^{2\boldsymbol{e}_i}f||_\infty}{(a-1)(a-2)}\frac{\tau_n}{(a+b)^{\overline{n|\boldsymbol{c}|-2}}n^2}\sum_{\substack{\boldsymbol{l}\in\mathbb{N}_0^{d+1}\\ |\boldsymbol{l}|>1}}l_i^2(a-2)^{\overline{|\boldsymbol{l}|}}b^{\overline{n|\boldsymbol{c}|-|\boldsymbol{l}|}}\prod_{k=1}^{d+1}\binom{nc_k}{l_k}.\end{aligned}$$

In addition to the identity $l_i^2 = (l_i)^{\underline{2}} + l_i$, we shall use Lemmas C.3.2 and C.3.5 (with $\boldsymbol{k} = 2\boldsymbol{e}_i$ in the first case and $\boldsymbol{k} = \boldsymbol{e}_i$ in the second), in order to deduce that, for each $i \in [d+1]$,

$$\begin{aligned}
&\frac{\tau_n}{(a+b)^{\overline{n|\boldsymbol{c}|-2}}n^2} \sum_{\substack{\boldsymbol{l}\in\mathbb{N}_0^{d+1}\\ |\boldsymbol{l}|>1}} (l_i)^{\underline{2}} (a-2)^{\overline{|\boldsymbol{l}|}} b^{\overline{n|\boldsymbol{c}|-|\boldsymbol{l}|}} \prod_{k=1}^{d+1} \binom{nc_k}{l_k} \\
&= \frac{\tau_n}{(a+b)^{\overline{n|\boldsymbol{c}|-2}}n^2} (a-2)(a-1)nc_i(nc_i-1)(a+b)^{\overline{n|\boldsymbol{c}|-2}} \\
&\sim (a-2)(a-1)c_i^2 n^{a-1} \to 0,
\end{aligned}$$

and

$$\begin{aligned}
&\frac{\tau_n}{(a+b)^{\overline{n|\boldsymbol{c}|-2}}n^2} \sum_{\substack{\boldsymbol{l}\in\mathbb{N}_0^{d+1}\\ |\boldsymbol{l}|>1}} l_i (a-2)^{\overline{|\boldsymbol{l}|}} b^{\overline{n|\boldsymbol{c}|-|\boldsymbol{l}|}} \prod_{k=1}^{d+1} \binom{nc_k}{l_k} \\
&= (a-2)\frac{\tau_n}{(a+b)^{\overline{n|\boldsymbol{c}|-2}}n^2} \left(nc_i(a+b-1)^{\overline{n|\boldsymbol{c}|-1}} - nc_i b^{\overline{n|\boldsymbol{c}|-1}} \right) \\
&\sim (a-2)\Gamma(a+b)c_i \left(\frac{n^{a-2}}{\Gamma(a+b-1)} - \frac{|\boldsymbol{c}|^{1-a}n^{-1}}{\Gamma(b)} \right) \to 0,
\end{aligned}$$

as $n \to \infty$, since $a \in (0,1)$.

For fixed $i, j \in [d+1]$ such that $i \neq j$ and $\boldsymbol{\kappa} = \boldsymbol{e}_i + \boldsymbol{e}_j$ we obtain

$$\begin{aligned}
&\frac{\tau_n}{n^2} \sum_{\substack{\boldsymbol{l}\in\mathbb{N}_0^{d+1}\\ |\boldsymbol{l}|>1}} l_i l_j D^{\boldsymbol{e}_i+\boldsymbol{e}_j} f\left(\boldsymbol{c} - \vartheta_{n,\boldsymbol{l}} \frac{\boldsymbol{l} - \boldsymbol{e}_{|\boldsymbol{l}|_w \wedge (d+1)}}{n} \right) \lambda_{n|\boldsymbol{c}|,|\boldsymbol{l}|} \prod_{k=1}^{d+1} \binom{nc_k}{l_k} \\
&\leq \frac{||D^{\boldsymbol{e}_i+\boldsymbol{e}_j} f||_\infty}{(a-1)(a-2)} \frac{\tau_n}{n^2(a+b)^{\overline{n|\boldsymbol{c}|-2}}} \sum_{\substack{\boldsymbol{l}\in\mathbb{N}_0^{d+1}\\ |\boldsymbol{l}|>1}} l_i l_j (a-2)^{\overline{|\boldsymbol{l}|}} b^{\overline{n|\boldsymbol{c}|-|\boldsymbol{l}|}} \prod_{k=1}^{d+1} \binom{nc_k}{l_k} \\
&= \frac{||D^{\boldsymbol{e}_i+\boldsymbol{e}_j} f||_\infty}{(a-1)(a-2)} \frac{\tau_n}{n^2(a+b)^{\overline{n|\boldsymbol{c}|-2}}} (a-2)(a-1)c_i c_j n^2 (a+b)^{\overline{n|\boldsymbol{c}|-2}} \\
&= ||f^{(\boldsymbol{e}_i+\boldsymbol{e}_j)}||_\infty c_i c_j n^{a-1} \to 0,
\end{aligned}$$

as $n \to \infty$, where we applied Lemma C.3.5 (with $\boldsymbol{k} = \boldsymbol{e}_i + \boldsymbol{e}_j$). This proves that $T^{(2)}(n)$ is bounded by something which tends to zero as $n \to \infty$, and hence vanishes in the limit.

Part 2. We now focus on $|\boldsymbol{\kappa}| = 1$. Let

$$T^{(1)}(n) := -\frac{\tau_n}{n} \sum_{\substack{\boldsymbol{l}\in\mathbb{N}_0^{d+1}\\ |\boldsymbol{l}|>1}} \sum_{\substack{\boldsymbol{\kappa}\in\mathbb{N}_0^{d+1}\\ |\boldsymbol{\kappa}|=1}} (\boldsymbol{l}-\boldsymbol{e}_{|\boldsymbol{l}|_w\wedge(d+1)})^{\boldsymbol{\kappa}} D^{\boldsymbol{\kappa}} f(\boldsymbol{c}) \lambda_{n|\boldsymbol{c}|,|\boldsymbol{l}|} \prod_{k=1}^{d+1} \binom{nc_k}{l_k}$$
$$= -\frac{\tau_n}{n(a+b)^{\overline{n|\boldsymbol{c}|-2}}} \sum_{i=1}^{d+1} \partial_i f(\boldsymbol{c}) \sum_{\substack{\boldsymbol{l}\in\mathbb{N}_0^{d+1}\\ |\boldsymbol{l}|>1}} (l_i - \mathbb{1}_{\{i=|\boldsymbol{l}|_w\wedge(d+1)\}}) a^{\overline{|\boldsymbol{l}|-2}} b^{\overline{n|\boldsymbol{c}|-|\boldsymbol{l}|}} \prod_{k=1}^{d+1} \binom{nc_k}{l_k}.$$

Now consider the summands in $T^{(1)}(n)$ corresponding to a fixed $i \in [d+1]$. We partition these summands and analyse their asymptotic behaviour separately. Firstly, by Lemma C.3.5 (setting $\boldsymbol{k} = \boldsymbol{e}_i$), we have that

$$O_i(n) := \frac{\tau_n}{n(a+b)^{\overline{n|\boldsymbol{c}|-2}}} \sum_{\substack{\boldsymbol{l}\in\mathbb{N}_0^{d+1}\\ |\boldsymbol{l}|>1}} l_i a^{\overline{|\boldsymbol{l}|-2}} b^{\overline{n|\boldsymbol{c}|-|\boldsymbol{l}|}} \prod_{k=1}^{d+1} \binom{nc_k}{l_k}$$
$$= \frac{1}{(a-1)(a-2)} \frac{\tau_n}{n(a+b)^{\overline{n|\boldsymbol{c}|-2}}} \sum_{\substack{\boldsymbol{l}\in\mathbb{N}_0^{d+1}\\ |\boldsymbol{l}|>1}} l_i (a-2)^{\overline{|\boldsymbol{l}|}} b^{\overline{n|\boldsymbol{c}|-|\boldsymbol{l}|}} \prod_{k=1}^{d+1} \binom{nc_k}{l_k}$$
$$= \frac{c_i}{a-1} \frac{\tau_n}{(a+b)^{\overline{n|\boldsymbol{c}|-2}}} \left((a+b-1)^{\overline{n|\boldsymbol{c}|-1}} - b^{\overline{n|\boldsymbol{c}|-1}} \right)$$
$$\sim \frac{\Gamma(a+b)}{a-1} c_i n^{a-1} \left(\frac{1}{\Gamma(a+b-1)} - \frac{(|\boldsymbol{c}|n)^{1-a}}{\Gamma(b)} \right) \sim -\frac{\Gamma(a+b)}{(a-1)\Gamma(b)} c_i |\boldsymbol{c}|^{1-a},$$

as $n \to \infty$. Secondly, for any fixed $i \in [d]$, the summand corresponding to the indicator $\mathbb{1}_{\{i=|\boldsymbol{l}|_w\wedge(d+1)\}}$ has asymptotic behaviour

$$I_i(n) := -\frac{\tau_n}{n(a+b)^{\overline{n|\boldsymbol{c}|-2}}} \sum_{\substack{\boldsymbol{l}\in\mathbb{N}_0^{d+1}\\ |\boldsymbol{l}|>1}} \mathbb{1}_{\{i=|\boldsymbol{l}|_w\wedge(d+1)\}} a^{\overline{|\boldsymbol{l}|-2}} b^{\overline{n|\boldsymbol{c}|-|\boldsymbol{l}|}} \prod_{k=1}^{d+1} \binom{nc_k}{l_k}$$
$$= -\frac{\tau_n}{n(a+b)^{\overline{n|\boldsymbol{c}|-2}}} \sum_{\substack{\boldsymbol{l}\in\mathbb{N}_0^{d+1}\\ |\boldsymbol{l}|>1, |\boldsymbol{l}|_w=i}} a^{\overline{|\boldsymbol{l}|-2}} b^{\overline{n|\boldsymbol{c}|-|\boldsymbol{l}|}} \prod_{k=1}^{d+1} \binom{nc_k}{l_k}$$
$$= -\frac{\tau_n}{n(a+b)^{\overline{n|\boldsymbol{c}|-2}}} \sum_{m=2}^{n|\boldsymbol{c}|} a^{\overline{m-2}} b^{\overline{n|\boldsymbol{c}|-m}} \sum_{\substack{\boldsymbol{l}\in\mathbb{N}_0^{d+1}\\ |\boldsymbol{l}|=m, |\boldsymbol{l}|_w=i}} \prod_{k=1}^{d+1} \binom{nc_k}{l_k}$$

$$\sim -\frac{\Gamma(a+b)}{\Gamma(b)}\sum_{m=2}^{i} a^{\overline{m-2}}|\boldsymbol{c}|^{2-a-m}\sum_{\substack{\boldsymbol{l}\in\mathbb{N}_0^{d+1}\\ |\boldsymbol{l}|=m,|\boldsymbol{l}|_w=i}}\prod_{k=1}^{d+1}\frac{c_k^{l_k}}{l_k!}.$$

However, $I_{d+1}(n)$ must be treated as a special case. Using the set equality

$$\{\boldsymbol{l}\in\mathbb{N}_0^{d+1}:|\boldsymbol{l}|>1,|\boldsymbol{l}|_w\geq d+1\}=\mathbb{N}_0^{d+1}\setminus\big(\{\boldsymbol{l}:|\boldsymbol{l}|\geq 2,|\boldsymbol{l}|_w\in[d]\}\cup\{\boldsymbol{l}:0\leq|\boldsymbol{l}|\leq 1\}\big),$$

it follows that

$$\begin{aligned}
I_{d+1}(n) &:= -\frac{1}{(a-1)(a-2)}\frac{\tau_n}{n\,(a+b)^{\overline{n|\boldsymbol{c}|-2}}}\Bigg(\sum_{\boldsymbol{l}\in\mathbb{N}_0^{d+1}}(a-2)^{\overline{|\boldsymbol{l}|}}\,b^{\overline{n|\boldsymbol{c}|-|\boldsymbol{l}|}}\prod_{k=1}^{d+1}\binom{nc_k}{l_k}\\
&\quad-\sum_{r=1}^{d}\sum_{m=2}^{r}(a-2)^{\overline{m}}\,b^{\overline{n|\boldsymbol{c}|-m}}\sum_{\substack{\boldsymbol{l}\in\mathbb{N}_0^{d+1}\\ |\boldsymbol{l}|=m,|\boldsymbol{l}|_w=r}}\prod_{k=1}^{d+1}\binom{nc_k}{l_k}\\
&\quad-\sum_{\substack{\boldsymbol{l}\in\mathbb{N}_0^{d+1}\\ |\boldsymbol{l}|=1}}(a-2)^{\overline{|\boldsymbol{l}|}}\,b^{\overline{n|\boldsymbol{c}|-|\boldsymbol{l}|}}\prod_{k=1}^{d+1}\binom{nc_k}{l_k}-b^{\overline{n|\boldsymbol{c}|}}\Bigg)\\
&=: -\Big(I_{d+1}^{(1)}(n)-I_{d+1}^{(2)}(n)-I_{d+1}^{(3)}(n)\Big),
\end{aligned}$$

where $I_{d+1}^{(i)}(n)$ is defined by the line on which it lies in the above equality. Using Lemma C.3.5 (in the special case $\boldsymbol{k}=\boldsymbol{0}$),

$$I_{d+1}^{(1)}(n)=\frac{1}{(a-1)(a-2)}\frac{\tau_n}{n\,(a+b)^{\overline{n|\boldsymbol{c}|-2}}}(a+b-2)^{\overline{n|\boldsymbol{c}|}}\sim\frac{1}{(a-1)(a-2)}\frac{\Gamma(a+b)}{\Gamma(a+b-2)}n^{a-2}$$

vanishes as $n\to\infty$, since $a<1$. The other two sums, as we shall see, are non-trivial. Firstly, using Lemma C.3.2,

$$I_{d+1}^{(2)}(n)\sim\frac{\Gamma(a+b)}{\Gamma(b)}\sum_{r=1}^{d}\sum_{m=2}^{r}a^{\overline{m-2}}|\boldsymbol{c}|^{2-a-m}\sum_{\substack{\boldsymbol{l}\in\mathbb{N}_0^{d+1}\\ |\boldsymbol{l}|=m,|\boldsymbol{l}|_w=r}}\prod_{k=1}^{d+1}\frac{c_k^{l_k}}{l_k!}.$$

Secondly, making use of the equation

$$\sum_{\substack{\boldsymbol{l}\in\mathbb{N}_0^{d+1}\\ |\boldsymbol{l}|=1}}(a-2)^{\overline{|\boldsymbol{l}|}}\,b^{\overline{n|\boldsymbol{c}|-|\boldsymbol{l}|}}\prod_{k=1}^{d+1}\binom{nc_k}{l_k}=\sum_{m=1}^{d+1}(a-2)b^{\overline{n|\boldsymbol{c}|-1}}nc_m=(a-2)b^{\overline{n|\boldsymbol{c}|-1}}n|\boldsymbol{c}|,$$

the final term satisfies

$$I^{(3)}_{d+1}(n) = \frac{1}{(a-1)(a-2)} \frac{\tau_n}{n\,(a+b)^{\overline{n|\boldsymbol{c}|-2}}} \left(\sum_{\substack{\boldsymbol{l}\in\mathbb{N}_0^{d+1} \\ |\boldsymbol{l}|=1}} (a-2)^{\overline{|\boldsymbol{l}|}}\, b^{\overline{n|\boldsymbol{c}|-|\boldsymbol{l}|}} \prod_{k=1}^{d+1} \binom{nc_k}{l_k} + b^{\overline{n|\boldsymbol{c}|}} \right)$$
$$\sim \frac{\Gamma(a+b)}{(a-2)\Gamma(b)} |\boldsymbol{c}|^{2-a}.$$

Since, by definition, $T^{(1)}(n) = -\sum_{i=1}^{d+1}(O_i(n) + I_i(n))\partial_i f(\boldsymbol{c})$, and $T^{(2)}(n) \to 0$ as $n \to \infty$ (see Part 1), the result follows. □

Once again, we see that the limit is deterministic. It follows that, for each $i \in [d]$, the limiting fraction of blocks of size i satisfies the ODE

$$\frac{d}{dt} c_i^{a,b}(t) = K \sum_{m=1}^{i} (a-2)^{\overline{m}}\, c^{a,b}(t)^{2-a-m} \sum_{\substack{\boldsymbol{l}\in\mathbb{N}_0^{d+1} \\ |\boldsymbol{l}|=m,\ |\boldsymbol{l}|_w=i}} \prod_{k=1}^{i} \frac{c_k^{a,b}(t)^{l_k}}{l_k!}, \tag{3.39}$$

with initial condition $c_i^{a,b}(0) = \mathbb{1}\{i=1\}$, using $|\boldsymbol{c}^{a,b}(t)| = c^{a,b}(t)$. This equation holds for all $i \in \mathbb{N}$ since we can make d arbitrarily large. It is a recursion in the sense that, for each $i \in \mathbb{N}$, the solution for $c_i^{a,b}(t)$ can be found by knowing only $c_j^{a,b}(t), j \in [i-1]$, and $c^{a,b}(t)$. However, $c_i^{a,b}$ depends on its earlier counterparts in rather complicated ways, meaning that it is difficult to solve the ODE in a straightforward manner. One can however make progress with the help of generating functions; the following theorem shows that the solution of this system can be rather simply expressed in terms of a complete Bell polynomial. (For a review of Bell polynomials, see Appendix B.)

Theorem 3.18. *Assume that $a \in (0,1)$ and $b > 0$. For each $i \in \mathbb{N}$, the solution of differential equation (3.39) is given by*

$$c_i^{a,b}(t) = \frac{c^{a,b}(t)^{2-a}}{i!} B_i\left(\left(\frac{1}{1-a}\right)^{\overline{\bullet}} (-c^{a,b}(t)^{1-a})^{\bullet-1}, (1-a)^{\overline{\bullet}} \right) \tag{3.40}$$

where $B_i(v_\bullet, w_\bullet)$ is the ith complete Bell polynomial and $c^{a,b}(t)$ is defined in (3.33).

Proof. We shall proceed in three steps: firstly, we shall rewrite equation (3.39) into one involving a Bell polynomial; secondly, we shall define a generating function for the spectral elements and find the

corresponding differential equation; thirdly, we solve this equation and simplify the solution using Faà di Bruno's formula. The second step in this scheme closely mimics that of the generating function approach described at the end of Section 3.5.2 in the Kingman case. Observe that

$$\begin{aligned}\frac{d}{dt}c_i^{a,b}(t) &= \frac{K}{i!}\sum_{m=1}^{i}(a-2)^{\overline{m}}\,c^{a,b}(t)^{2-a-m}\sum_{\substack{\boldsymbol{l}\in\mathbb{N}_0^d\\ |\boldsymbol{l}|=m,\ |\boldsymbol{l}|_w=i}}\frac{i!}{\prod_{k=1}^{i}l_k!(k!)^{l_k}}\prod_{k=1}^{i}(k!\,c_k^{a,b}(t))^{l_k}\\ &= \frac{K}{i!}\sum_{m=1}^{i}(a-2)^{\overline{m}}\,c^{a,b}(t)^{2-a-m}B_{i,m}\left(w_\bullet\right)\\ &= \frac{K}{i!}c^{a,b}(t)^{2-a}B_i\left(v_\bullet,w_\bullet\right),\end{aligned} \tag{3.41}$$

where

$$v_k := (a-2)^{\overline{k}}\,c^{a,b}(t)^{-k},\quad w_k := k!\,c_k^{a,b}(t).$$

In order to proceed we shall define the three following generating functions;

$$G^{a,b}(x,t) := \sum_{k=1}^{\infty}x^k c_k^{a,b}(t),\quad v(\theta,t) := \sum_{k=1}^{\infty}\frac{\theta^k}{k!}v_k,\quad w(\eta,t) := \sum_{k=1}^{\infty}\frac{\eta^k}{k!}w_k = G^{a,b}(\eta,t). \tag{3.42}$$

We can calculate the generating function of the sequence v explicitly using the generalized binomial series,

$$v(\theta,t) = \sum_{k=1}^{\infty}\binom{2-a}{k}\left(\frac{-\theta}{c^{a,b}(t)}\right)^k = \left(1-\frac{\theta}{c^{a,b}(t)}\right)^{2-a}-1.$$

At this point, we are in a position to fully exploit the properties of Bell polynomials, and in particular the formula for the composition of their generating functions (see Appendix B) to derive a differential equation for $G^{a,b}$. To do so, we make use of (3.41), observing that

$$\begin{aligned}\frac{\partial}{\partial t}G^{a,b}(x,t) &= \sum_{k=1}^{\infty}x^k\frac{d}{dt}c_k^{a,b}(t) = Kc^{a,b}(t)^{2-a}\sum_{k=1}^{\infty}\frac{x^k}{k!}B_k\left(v_\bullet,w_\bullet\right) = Kc^{a,b}(t)^{2-a}v(w(x,t),t)\\ &= K\left(\left(c^{a,b}(t)-G^{a,b}(x,t)\right)^{2-a}-c^{a,b}(t)^{2-a}\right),\end{aligned}$$

with initial condition $G^{a,b}(x,0) = x$. Since there is no derivative in the x-variable, this differential equation may be regarded as an ODE for each x in the radius of convergence. By making the simple

transformation $H^{a,b}(x,t) = c^{a,b}(t) - G^{a,b}(x,t)$ and using the relation $(c^{a,b}(t))' = -Kc^{a,b}(t)^{2-a}$, one may show that $H^{a,b}$ solves

$$\frac{\partial}{\partial t}H^{a,b}(x,t) = -KH^{a,b}(x,t)^{2-a}, \qquad H^{a,b}(x,0) = 1-x. \tag{3.43}$$

What is remarkable is that, besides the x-dependence and the initial condition, this is precisely the same equation as (3.32) for $c^{a,b}$, the limiting function of the block-counting process. In particular, it is again separable and of Bernoulli type, and may be solved using standard methods to obtain

$$H^{a,b}(x,t) = \left((1-x)^{a-1} - (a-1)Kt\right)^{\frac{1}{a-1}}. \tag{3.44}$$

The final step in the proof is to find an explicit formula for $c_i^{a,b}(t)$ from the generating function. First observe that, for $i \geq 1$,

$$c_i^{a,b}(t) = -\frac{1}{i!}\frac{\partial^i}{\partial x^i}H^{a,b}(x,t)\bigg|_{x=0}.$$

We shall define the the auxiliary functions and their derivatives

$$\begin{aligned} f(y) &:= (y-(a-1)Kt)^{\frac{1}{a-1}} \quad &\Longrightarrow \quad f^{(k)}(y) &= \left(\frac{1}{a-1}\right)^{\underline{k}}(y-(a-1)Kt)^{\frac{1}{a-1}-k}, \\ g(y) &:= (1-y)^{a-1} \quad &\Longrightarrow \quad g^{(k)}(y) &= (-1)^k\,(a-1)^{\underline{k}}(1-y)^{a-1-k}. \end{aligned}$$

Finally, using (3.33) and the Faà di Bruno formula for the derivatives of compositions (see Appendix B), and substituting in the above formulas, we deduce that

$$\begin{aligned} c_i^{a,b}(t) &= -\frac{1}{i!}\frac{\partial^i}{\partial x^i}f(g(x))\bigg|_{x=0} = -\frac{1}{i!}\sum_{\pi\in\mathscr{P}_{[i]}} f^{(|\pi|)}(g(x))\prod_{B\in\pi} g^{(|B|)}(x)\bigg|_{x=0} \\ &= -\frac{1}{i!}\sum_{\pi\in\mathscr{P}_{[i]}}\left(\frac{1}{a-1}\right)^{\underline{|\pi|}}(1-(a-1)Kt)^{\frac{1}{a-1}-|\pi|}\prod_{B\in\pi}(-1)^{|B|}\,(a-1)^{\underline{|B|}} \\ &= -\frac{c^{a,b}(t)}{i!}\sum_{m=1}^{i}(-1)^m\left(\frac{1}{1-a}\right)^{\overline{m}} c^{a,b}(t)^{m(1-a)}B_{i,m}\left((1-a)^{\overline{\bullet}}\right) \\ &= \frac{c^{a,b}(t)^{2-a}}{i!}B_i\left(\left(\frac{1}{1-a}\right)^{\overline{\bullet}}(-c^{a,b}(t)^{1-a})^{\bullet-1},\ (1-a)^{\overline{\bullet}}\right). \end{aligned}$$

□

CHAPTER 4

Strong Mutation

In the Wright–Fisher diffusion with weak mutation, the reproduction and mutation mechanisms act independently by construction. In the generator of this process, (2.29), the mutation term is added separately in the form of a first order differential operator. Since this form of mutation appears quite organically in the classical model, it is quite natural to assume it when studying new models, as many authors have (see, for example, Eldon and Wakeley [2006]).

This chapter will introduce a new mechanism which intrinsically links mutation to reproduction in the Λ-Fleming–Viot model (introduced in Chapter 3). We can think of such a process as modelling populations in which a small number of parents produce a large number of offspring, through the 'sweepstake' mechanism for example, such that a large fraction of the brood mutate to the other type. Correspond-

ingly, throughout this exposition, we shall use the notational convention of letting 'M' denote strong mutation, and 'm' for weak mutation. For example, the generator of the Λ-Fleming–Viot process with strong mutation will be denoted $\mathscr{L}_{\Lambda M}$. Such a model may be suitable for modelling marine organisms which experience mutation during meiosis.

The layout of this chapter is as follows. In Section 4.1, we outline the definition of Λ-Fleming–Viot process with strong mutation (ΛFVM), and give a condition for the existence of this process. An example of an individual-based model, which can be approximated by the ΛFVM process, is discussed in Section 4.2. In Section 4.3 we derive some properties of the process and realize that, by making a certain transformation, its generator can be decomposed into the sum of the generator corresponding to the ordinary Λ-Fleming–Viot process and a non-generator term that accounts for mutation (Lemma 4.3). This alternative representation of the process will be crucial for deriving the stationary distribution in the special case of the Beta coalescent in Section 4.4, using an adaptation of the method of Handa [2014]. We are able to characterize the density of this stationary distribution in terms of (a variant of) its generalized Stieltjes transform, but unfortunately are not able to invert this explicitly. However, we are able to use this transformed version of the density to obtain an explicit recursion for the moments (Section 4.4.2). In Section 4.5, we use the duality arguments of Chapter 2 to study the weak genealogy of the ΛFVM process, which is fully characterized in Theorem 4.7. Finally, in Section 4.6, we discuss some possible generalizations of the process that could be studied in the future.

4.1 The Λ-Fleming–Viot Process with Strong Mutation

We shall consider a version of the ordinary two-type Λ-Fleming–Viot (hereafter, ΛFV) process, but with mutation between the two types. In particular, a certain fraction of the offspring, rather than receiving the genetic type of their parent (as is usual), will instead mutate to the other type.

We shall say that there are two genetic types, a_1 and a_2, and the process of interest, $(X_t)_{t\geq 0}$, tracks the proportion of individuals of the former type. Mutation is controlled by two parameters, $v_1, v_2 \in (0,1)$. For convenience, we shall enforce the condition $v_2 < 1 - v_1$ in order to streamline some of the

arguments in Section 4.3. There is no loss of generality as $v_2 > 1 - v_1$ would also lead to the same results, but the case $v_2 = 1 - v_1$ would not.

The model can be described easily through its generator, which is the linear operator

$$\mathscr{L}_{\Lambda M} f(x) = \int_0^1 \left[x f((1-u)x + u(1-v_1)) + (1-x) f((1-u)x + uv_2) - f(x) \right] \nu(du), \qquad (4.1)$$

acting on bounded, measurable functions f, where $\nu(du) = u^{-2}\Lambda(du)$ and Λ is a measure on $(0, 1]$. As in Chapter 3, we cannot take any measure Λ, and the restrictions upon it can be found in Section 4.1.1. This jump process *essentially* lives on a subset of $(0, 1)$, which we call the *core*, the region which the process will eventually be supported on – see Section 4.3 for more details. The heuristic interpretation of this model is very similar to that given in Section 3.2, except that we now allow the newborns to mutate during the reproduction event.

The measure ν is the driving measure that determines how the impact parameter u is picked – it plays the same role as ν in the ordinary ΛFV process, except that it must satisfy a stronger condition, namely that $\frac{\Lambda(du)}{u}$ be a finite measure – see Section 4.1.1. Given u, a parent is chosen and is of type a_1 with probability x, and of type a_2 with the complimentary probability $1 - x$. In either case, a fraction u of the population is uniformly killed (regardless of type). In order to preserve a 'constant population size', we need to reallocate this missing fraction: if the parental type is $a_i, i \in \{1, 2\}$, then the new offspring will also be of that type, except that a fraction v_i mutate to the other type. These dynamics comprise the high-density version of those of the individual-based model described in Section 4.2. If we set $v_1 = v_2 = 0$, we get back the usual ΛFV process.

Note that the capitalized 'M' used in the subscript of the generator stands for this form of strong mutation which is linked to the reproduction mechanism. This is in contrast to the Wright–Fisher diffusion with weak mutation, where the mutation was driven by an independent Poisson point process (see Sections 2.3.3 and 2.3.4), and the subscript 'm' was used instead.

4.1.1 Conditions on Λ

It is customary to write the measure ν in (4.1) in terms of another measure, Λ. This drives the rates of coalescence in the usual Λ-coalescent, the dual process of the ordinary ΛFV process (see Section 3.4). The relationship is written as $\nu(du) = \frac{\Lambda(du)}{u^2}$, where Λ is a *finite* measure on the interval $(0, 1]$. The finiteness ensures that the ΛFV process exists and the rates of coalescence in the Λ-coalescent are finite. This topic was discussed in Chapter 3.

Our model deviates from the ΛFV process in that it includes a mutation mechanism. This, in turn, forces an extra condition on Λ, meaning that a smaller class of driving measures can be used by this model. If we take the Taylor expansion of a smooth function f about x in (4.1), we obtain the expression

$$\begin{aligned}\mathscr{L}_{\Lambda M} f(x) &= \int_0^1 \left[(v_2(1-x) - v_1 x)uf'(x) + (x(1-v_1-x)^2 + (1-x)(v_2-x)^2)u^2 f''(x) + \mathcal{O}(u^3)\right] \frac{\Lambda(du)}{u^2} \\ &= c_{-1}(x)f'(x)\int_0^1 \frac{\Lambda(du)}{u} + c_0(x)f''(x)\int_0^1 \Lambda(du) + \sum_{i=1}^{\infty} c_i(x) f^{(i+2)}(x) \int_0^1 u^i \Lambda(du).\end{aligned}$$

To ensure that the process is well-defined, we will enforce the condition

$$\int_0^1 \frac{\Lambda(du)}{u} < \infty, \tag{4.2}$$

which is stronger than the condition $\int_0^1 \Lambda(du) < \infty$ normally associated to the ΛFV process. We will see in Section 4.5.1 that (4.2) has an interpretation in the genealogical context. Through personal communication with Tom Kurtz, we believe that this condition is actually necessary to guarantee the existence of the of the process – we shall discuss this further in Section 4.6.

Remark 4.1. Upon setting $v_1 = v_2 = 0$, which eradicates the mutation mechanism, we get back to the usual ΛFV process. In addition, we see that $c_{-1}(x) = (v_2(1-x) - v_1 x) \equiv 0$ in this case. Hence, condition (4.2) becomes unnecessary, and Λ need only be a finite measure.

4.2 Individual-Based Model

We shall now consider a Moran-type model that provides the underlying individual-based dynamics for the Λ-Fleming–Viot process with strong mutation defined in Section 4.1. The model that we shall de-

scribe here could also serve as an alternative pre-limiting model to the Λ-Fleming–Viot process *without* mutation by setting the mutation parameters to zero (an alternative to that of Section 3.2).

Let $N \in \mathbb{N}$ be the fixed population size. Let Z_t^N be the number of individuals of type a_1 at time t (the number of type-a_2 individuals is $N - Z_t^N$). We shall define $X_t^N = Z_t^N/N$ to be the corresponding proportion at time t. Suppose that, at rate λ^N, reproduction events occur in which the genetic make-up of the population is modified. During that event (which takes place at time t), the following actions take place instantaneously:

(1) A parent is chosen uniformly at random from the entire population – the parent is of type a_1 with probability X_{t-}^N.

(2) The number of individuals that will be killed, K, is chosen according to some distribution $\nu^N(k) = \mathbb{P}(K = k)$, supported on $[N]$. Then, K individuals are selected independently and uniformly without replacement to be removed from the population. Given K, the number of type-a_1 individuals that will die, K_1, has a hypergeometric distribution, $\text{Hyp}(N, NX_{t-}^N, K)$.

(3) There are K offspring which replace the deceased that receive the genetic type of their parent, *unless* they mutate. This happens with probability v_i^N independently for each newborn if the parent is of type a_i. If the parent is of type a_1, the number of offspring that do *not* mutate, given K, has a binomial distribution, $M_1 \overset{d}{=} \text{Bin}(K, 1 - v_1^N)$; if the parent is of type a_2, the number of newborns that *do* mutate, given K, also has a binomial distribution, $M_2 \overset{d}{=} \text{Bin}(K, v_2^N)$.

The novelty in this model is that the mutation mechanism is incorporated into reproduction, so that mutations occur at birth. This is in contrast to other models of this nature in which mutation behaves as an independent Poisson process, allowing mutations to occur between births. Another difference to be noted is that the extinction part of this model, (2), differs to that of its counterpart in Section 3.2. In the latter, we specify an impact parameter $u \in (0, 1]$ according to a measure ν and declare that each individual dies with that probability. In this model, we select the number of deaths in advance according to the distribution ν^N, and then select the K individuals that will die. These aspects behave differently in the pre-limiting models, but they yield the same kind of behaviour in the limit.

When the frequency of type-a_1 individuals in the population is x, the rate at which we see events in which the parent is of type a_1, there are k deaths of which k_1 are of type a_1 and m_1 of the offspring do not mutate is given by

$$q_1\left(x,k,k_1,m_1\right) := x\lambda^N\nu^N(k)\mathbb{P}(K_1=k_1|K=k)\mathbb{P}(M_1=m_1|K=k),$$

in which the transition

$$x \mapsto x - \frac{k_1-m_1}{N}$$

is made. Similarly, the rate at which events occur in which the parent is of type a_2, there are k deaths of which k_1 are of type a_1 and m_2 of the offspring do mutate is given by

$$q_2\left(x,k,k_1,m_2\right) := (1-x)\lambda^N\nu^N(k)\mathbb{P}(K_1=k_1|K=k)\mathbb{P}(M_2=m_2|K=k),$$

in which the transition

$$x \mapsto x - \frac{k_1-m_2}{N}$$

is made. With these definitions, the generator of the jump process X_t^N can be written

$$\begin{aligned}
\tilde{\mathscr{L}}^N_{\Lambda M} f(x) &= \sum_{k=1}^{N}\sum_{k_1=0}^{k}\sum_{m_1=0}^{k} q_1(x,k,k_1,m_1)\left(f\left(x-\tfrac{k_1-m_1}{N}\right)-f(x)\right)\\
&\quad+\sum_{k=1}^{N}\sum_{k_1=0}^{k}\sum_{m_2=0}^{k} q_2(x,k,k_1,m_2)\left(f\left(x-\tfrac{k_1-m_2}{N}\right)-f(x)\right)\\
&= x\lambda^N\sum_{k=1}^{N}\nu^N(k)\mathbb{E}\left[f\left(x-\tfrac{K_1-M_1}{N}\right)-f(x)|K=k\right]\\
&\quad+(1-x)\lambda^N\sum_{k=1}^{N}\nu^N(k)\mathbb{E}\left[f\left(x-\tfrac{K_1-M_2}{N}\right)-f(x)|K=k\right].
\end{aligned}$$

It is possible to approximate the expectations with their means provided a certain condition is met. Consider

$$\begin{aligned}
E^N(k) &:= \mathbb{E}\left[x\left(f\left(x-\tfrac{K_1-M_1}{N}\right)-f(x)\right)+(1-x)\left(f\left(x-\tfrac{K_1-M_2}{N}\right)-f(x)\right)|K=k\right]\\
&= \mathbb{E}\big[x\left(f\left(x-\tfrac{K_1-M_1}{N}\right)-f\left(x-\tfrac{k}{N}(x-(1-v_1^N))\right)+f\left(x-\tfrac{k}{N}(x-(1-v_1^N))\right)-f(x)\right)
\end{aligned}$$

$$+ (1-x)\left(f\left(x - \tfrac{K_1-M_2}{N}\right) - f\left(x - \tfrac{k}{N}(x - v_2^N)\right) + f\left(x - \tfrac{k}{N}(x - v_2^N)\right) - f(x)\right)|K = k\big]$$
$$= G_k^N f(x) + r^N(k),$$

where

$$G_k^N f(x) := xf\left(x - \tfrac{k}{N}(x - (1 - v_1^N))\right) + (1-x)f\left(x - \tfrac{k}{N}(x - v_2^N)\right) - f(x)$$

is the part that will eventually form the limiting generator, and

$$r^N(k) := \mathbb{E}\big[x\left(f\left(x - \tfrac{K_1-M_1}{N}\right) - f\left(x - \tfrac{k}{N}(x - (1 - v_1^N))\right)\right)$$
$$+ (1-x)\left(f\left(x - \tfrac{K_1-M_2}{N}\right) - f\left(x - \tfrac{k}{N}(x - v_2^N)\right)\right)|K = k\big]$$

is the remainder term that we hope will vanish as $N \to \infty$. Making use of Taylor expansions inside the expectation we find that, for some $\xi, \xi', \eta, \eta' \in [0,1]$,

$$\begin{aligned} r^N(k) = {} & xf'(x)\left(\tfrac{k}{N}(x - (1 - v_1^N)) - \mathbb{E}[\tfrac{K_1-M_1}{N}|K = k]\right) \\ & + (1-x)f'(x)\left(\tfrac{k}{N}(x - v_2^N) - \mathbb{E}[\tfrac{K_1-M_2}{N}|K = k]\right) \\ & + \tfrac{x}{2N^2}\left(\mathbb{E}[(K_1 - M_1)^2 f''(\xi)|K = k] - (k(x - (1 - v_1^N)))^2 f''(\xi')\right) \\ & + \tfrac{1-x}{2N^2}\left(\mathbb{E}[(K_1 - M_2)^2 f''(\eta)|K = k] - (k(x - v_2^N))^2 f''(\eta')\right). \end{aligned}$$

The first two lines cancel to zero using the properties $\mathbb{E}[K_1|K = k] = kx$, $\mathbb{E}[M_1|K = k] = k(1 - v_1^N)$ and $\mathbb{E}[M_2|K = k] = kv_2^N$ (but this was of course the point of subtracting and adding the mean term). Now, the remainder term may be bounded by

$$\begin{aligned} r^N(k) &\leq \frac{||f''||_\infty}{2N^2}\Big(x\left(\text{Var}(K_1|K = k) + \text{Var}(M_1|K = k)\right) \\ &\qquad + (1-x)\left(\text{Var}(K_1|K = k) + \text{Var}(M_2|K = k)\right)\Big) \\ &= \frac{||f''||_\infty}{2N^2}\left(kx(1-x)\left(1 - \tfrac{k-1}{N-1}\right) + kxv_1^N(1 - v_1^N) + k(1-x)v_2^N(1 - v_2^N)\right) \\ &\leq \frac{||f''||_\infty}{2N^2}k, \end{aligned}$$

uniformly in $x, v_1^N, v_2^N \in (0,1)$, where $||\cdot||_\infty$ is the supremum norm defined in Appendix A. Hence,

defining $\mathscr{L}^N_{\Lambda M} f(x) := \lambda^N \sum_{k=1}^N G^N_k f(x) \nu^N(k)$ and $R^N := \lambda^N \sum_{k=1}^N r^N(k) \nu^N(k)$, it follows that

$$\tilde{\mathscr{L}}^N_{\Lambda M} f(x) = \mathscr{L}^N_{\Lambda M} f(x) + R^N \overset{N\to\infty}{\longrightarrow} \mathscr{L}_{\Lambda M} f(x)$$

if we can find sequences λ^N, $\nu^N(k)$, v_1^N and v_2^N such that

$$\begin{aligned}
&\mathscr{L}^N_{\Lambda M} f(x) = \\
&\quad \lambda^N \sum_{k=1}^N \left[x f\left((1-\tfrac{k}{N})x + \tfrac{k}{N}(1-v_1^N) \right) + (1-x) f\left((1-\tfrac{k}{N})x + \tfrac{k}{N} v_2^N \right) - f(x) \right] \nu^N(k) \\
&\longrightarrow \int_0^1 \left[x f((1-u)x + u(1-v_1)) + (1-x) f((1-u)x + u v_2) - f(x) \right] \nu(du)
\end{aligned}$$

and

$$\frac{\lambda^N}{N^2} \sum_{k=1}^N k \nu^N(k) \longrightarrow 0$$

as $N \to \infty$ for all twice differentiable functions f on $(0,1)$ satisfying $f(x) = \mathcal{O}(x^2)$ as $x \downarrow 0$. If this is the case, then $X_t^N \to X_t$ weakly as $N \to \infty$. From this point onwards, we shall no longer think about this individual-based model, but rather the limiting process which approximates it.

4.3 Generator Decomposition

With a view to looking for the stationary distribution of the limiting process, we shall take a closer look at some of its properties. We will show that, whilst X_0 may take any value in the closed unit interval, the state space can be reduced, and the process will always eventually be trapped inside a more fundamental interval, $[v_2, 1 - v_1]$, which we will call the *core*. Recalling that we chose $v_2 < 1 - v_1$, the *width* of the core,

$$w := 1 - v_1 - v_2, \tag{4.3}$$

is positive. Define the functions

$$f_\uparrow(x) = (1-u)x + u(1-v_1) \quad \text{and} \quad f_\downarrow(x) = (1-u)x + u v_2,$$

which have respective fixed points $1 - v_1$ and v_2. These functions appear in the generator (4.1), correspond to the a_1-parent and a_2-parent events respectively, and determine the boundaries of the core. The case analysis below describes how the process behaves when it is in (i) $[0, v_2)$, the left exterior region, (ii) $(1 - v_1, 1]$, the right exterior region, and (iii) $[v_2, 1 - v_1]$, the core.

(i) Suppose $x < v_2$. Then,

(a) $f_\downarrow(x) = x - ux + uv_2 > x$.

(b) $f_\uparrow(x) = x - ux + u(1 - v_1) > x + uw$.

This means that, no matter which parent is picked, the process, when situated in the left exterior region, will always make a jump toward the core, since u is almost surely positive and $w > 0$. In fact, we can see that

$$\lim_{x \uparrow v_2} (f_\downarrow(x) - x) = 0 \quad \text{and} \quad \lim_{x \uparrow v_2} (f_\uparrow(x) - x) = uw.$$

That is, if an unusual run of events meant that an a_2-parent was chosen again and again, X_t would tend to v_2 from below without breaching it. However, with probability 1, an a_1-parent must eventually be picked, which would allow X_t to jump *into* the core.

(ii) Suppose $x > 1 - v_1$. A symmetric argument shows that (a) $f_\uparrow(x) < x$ and (b) $f_\downarrow(x) < x - uw$, so that the process always decreases, making progress towards the core. Similarly, $\lim_{x \downarrow 1 - v_1} (f_\uparrow(x) - x) = 0$ and $\lim_{x \downarrow 1 - v_1} (f_\downarrow(x) - x) = -uw < 0 \ \ a.s.$. In words, X_t started from the right exterior region will always take steps toward the core, and eventually jump into it.

(iii) Suppose $v_2 \leq x \leq 1 - v_1$. Then, (a) $f_\uparrow(x) \in [v_2 + uw, 1 - v_1]$ and (b) $f_\downarrow(x) \in [v_2, (1 - v_1) - uw]$, which proves that, once X_t is in the core, it cannot exit. This follows because w exactly coincides with the width of the core, and since $u \in (0, 1)$ almost surely, $0 < uw < w$.

Remark 4.2. Letting $v_2 > 1 - v_1$ does not fundamentally change the behaviour just described – the roles of $f_\uparrow$ and $f_\downarrow$ simply swap, and the core is defined as $[1 - v_1, v_2]$ instead. However, if $v_2 = 1 - v_1$, then the core degenerates to a single point (and the stationary distribution will be a point mass in this case).

We are now in a position to make the alternative representation of (4.1). Because X_t is pulled toward the irreducible core, we should expect that, if a stationary distribution exists, it will be non-trivial there, and zero on the exterior. To reflect this fact, we shall make the linear change of variables

$$y = y(x) := \frac{x - v_2}{w}, \tag{4.4}$$

with $y(v_2) = 0$ and $y(1 - v_1) = 1$, which focuses on the core. We write the inverse as

$$x = v_2 + wy =: g(y). \tag{4.5}$$

Proposition 4.3 (Generator decomposition). *For all $f \in \mathscr{D}(\mathscr{L}_{\Lambda M})$, and defining $h(y) = (f \circ g)(y)$, we can write the decomposition*

$$\mathscr{L}_{\Lambda M} h(y) = \mathscr{L}_{\Lambda} h(y) + \mathscr{L}_{M} h(y), \tag{4.6}$$

where

$$\mathscr{L}_{M} h(y) = (v_2 - (v_1 + v_2)y) \int_0^1 (h((1-u)y + u) - h((1-u)y))\nu(du)$$

and $\mathscr{L}_{\Lambda}$ is the generator of the ordinary ΛFV process (see (3.3) of Section 3.1).

Proof. Observe that

$$\begin{aligned} f_{\uparrow}(x) &= f_{\uparrow}(g(y)) = w((1-u)y + u) + v_2 = g((1-u)y + u) \\ f_{\downarrow}(x) &= f_{\downarrow}(g(y)) = (1-u)wy + v_2 = g((1-u)y) \end{aligned}$$

Hence, setting $h(y) = (f \circ g)(y)$, Equation (4.1) may be rewritten as

$$\begin{aligned} \mathscr{L}_{\Lambda M} f(g(y)) &= \int_0^1 [g(y) f(f_{\uparrow}(g(y))) + (1 - g(y)) f(f_{\downarrow}(g(y))) - f(g(y))]\, \nu(du) \\ &= \int_0^1 [(y + v_2 - (v_1 + v_2)y) f(g((1-u)y + u)) \\ &\qquad + (1 - y - (v_2 - (v_1 + v_2)y)) f(g((1-u)y)) - f(g(y))]\, \nu(du) \\ &= \mathscr{L}_{\Lambda} h(y) + \mathscr{L}_{M} h(y). \end{aligned}$$

□

Remark 4.4. Upon removing mutation by setting $v_1 = v_2 = 0$, we return to the ordinary case of the Λ-Fleming–Viot process, since the decomposition is trivial. That is, $\mathscr{L}_M \equiv 0$ and $h(y) = f(y)$.

4.4 The Stationary Distribution

In order to find the stationary distribution of a Markov process with generator $\mathscr{L}$, with domain $\mathscr{D}$, we need to find a non-trivial measure, ψ, which solves

$$\int_S \mathscr{L} f(x)\psi(dx) = 0, \qquad \forall f \in \mathscr{D}(\mathscr{L}),$$

where S is the support of ψ (Ethier and Kurtz [1986], Section 4.9). In our case, assuming that the measure ψ is absolutely continuous, we have

$$\int_{v_2}^{1-v_1} \mathscr{L}_{\Lambda M} f(x)\tilde{\psi}(x)dx = 0 \quad \text{or} \quad \int_0^1 \mathscr{L}_{\Lambda M} h(y)\psi(y)dy = 0 \tag{4.7}$$

where $\psi(y) = w\tilde{\psi}(g(y))$, g is defined in (4.5) and w in (4.3). The second form is equivalent to the first and follows after performing the change of variables (4.4).

4.4.1 Beta Coalescent

It would be ideal if we could solve this problem for the general measure ν, but that would be very challenging. Instead, we shall specialize to the (restricted) class of Beta coalescents, introduced in Section 3.4.2. That is, for $\alpha \in (0, 1)$, set

$$\nu(du) = \frac{\Lambda^\alpha(du)}{u^2}, \tag{4.8}$$

where Λ^α is the measure corresponding to the ordinary Beta coalescent, defined in (3.10). To study the stationary distribution for this special but important class of measures, we shall adapt the method of Handa [2014]. The mutation mechanism that he considers in his model is rather different from ours, but using (a modification of) his transform method allows significant progress to be made here too.

Handa points out that, to prove the identity (4.7), it is enough to check that it holds true for the set of

functions $h_n(y) := y^n$ for $n \in \mathbb{N}$ (this is a result of Ethier and Kurtz [1986]). Further, he argues that it is possible then to simply check (4.7) for the functions $h_t(y) = (1+ty)^{-1}$ for $t \in [0, 1/R)$ where $R > 0$ is to be determined. Observe that

$$\int_0^1 \mathscr{L}_{\Lambda M} h_t(y)\psi(y)dy = \int_0^1 \mathscr{L}_{\Lambda M}\left(\sum_{n=0}^{\infty}(-ty)^n\right)\psi(y)dy = \sum_{n=1}^{\infty}(-t)^n\int_0^1 \mathscr{L}_{\Lambda M}(y^n)\psi(y)dy$$

is valid by Fubini's theorem if we can, for example, estimate $|\mathscr{L}_{\Lambda M}(y^n)| \leq C \cdot R^n$ for each $n \in \mathbb{N}$ and for some $C > 0$. This would imply

$$\int_0^1 \sum_{n=0}^{\infty}|(-t)^n\mathscr{L}_{\Lambda M}(y^n)|\psi(y)dy = \int_0^1 \sum_{n=0}^{\infty} t^n|\mathscr{L}_{\Lambda M}(y^n)|\psi(y)dy \leq C\int_0^1 \sum_{n=1}^{\infty}(Rt)^n\psi(y)dy,$$

which converges as long as $t < 1/R$. Indeed, we show that such an estimate can be obtained in Lemma C.4.1 with $R = 2$, and hence we can proceed with the method. The next step is to make use of the generator decomposition (Lemma 4.3) to calculate

$$\mathscr{L}_{\Lambda} h_t(y) = \frac{t^2y(1-y)}{1+ty}\int_0^1 \frac{\Lambda^{\alpha}(du)}{(au+b)(a'u+b)}, \tag{4.9}$$

$$\mathscr{L}_{M} h_t(y) = -t(v_2-(v_1+v_2)y)\int_0^1 \frac{\Lambda^{\alpha}(du)}{u(au+b)(a'u+b)}, \tag{4.10}$$

where $a = t(1-y), a' = -ty$ and $b = 1+ty$. We shall simplify (4.9) and (4.10) by using partial fractions and then applying Lemmas C.4.2 and C.4.4 (which is possible since $|a|, |a'| < |b|$).

Remark 4.5. Note that, for Lemma C.4.2, the requirement that $\alpha \in (0,1)$ precisely coincides with (4.2) in this special case – if $\alpha \in [1,2)$, this condition fails.

After performing these steps, we obtain

$$\mathscr{L}_{\Lambda} h_t(y) = -ty(1-y)\frac{(1+t)^{\alpha-1}-1}{(1-\alpha)(1+ty)^{1+\alpha}}, \tag{4.11}$$

$$\mathscr{L}_{M} h_t(y) = -t(v_2-(v_1+v_2)y)\frac{(1-y)(1+t)^{\alpha-1}+y}{(1-\alpha)(1+ty)^{1+\alpha}}. \tag{4.12}$$

It follows from (4.7), (4.11) and (4.12) that, in order to find the stationary distribution, we must find ψ

such that

$$t\int_0^1 \frac{\psi(y)dy}{(1+ty)^{1+\alpha}}\left[y(1-y)((1+t)^{\alpha-1}-1)+(v_2-(v_1+v_2)y)((1-y)(1+t)^{\alpha-1}+y)\right]=0. \tag{4.13}$$

Equation (4.13) is conducive to an integral transform. In our problem, the appropriate transform is

$$S(t) := \int_0^1 (1+ty)^{1-\alpha}\psi(y)dy = \int_0^1 \frac{(1+ty)^2}{(1+ty)^{1+\alpha}}\psi(y)dy, \tag{4.14}$$

where S implicitly depends upon α. This is slightly different to the generalized Stieltjes transform used by Handa [2014]. Differentiating in t, we obtain

$$S'(t) = (1-\alpha)\int_0^1 \frac{y(1+ty)}{(1+ty)^{1+\alpha}}\psi(y)dy \quad \text{and} \quad S''(t) = -\alpha(1-\alpha)\int_0^1 \frac{y^2}{(1+ty)^{1+\alpha}}\psi(y)dy$$

Recall that $z^{\underline{k}}$ is the kth falling factorial power of z, defined in Appendix A. The relations

$$\begin{aligned}
\int_0^1 \frac{y}{(1+ty)^{1+\alpha}}\psi(y)dy &= -t\frac{S''(t)}{(1-\alpha)^{\underline{2}}} + \frac{S'(t)}{1-\alpha},\\
\int_0^1 \frac{y(1-y)}{(1+ty)^{1+\alpha}}\psi(y)dy &= -(1+t)\frac{S''(t)}{(1-\alpha)^{\underline{2}}} + \frac{S'(t)}{1-\alpha},\\
\int_0^1 \frac{1-y}{(1+ty)^{1+\alpha}}\psi(y)dy &= t(1+t)\frac{S''(t)}{(1-\alpha)^{\underline{2}}} - (1+2t)\frac{S'(t)}{1-\alpha} + S(t),
\end{aligned}$$

transform (4.13) into the second order differential equation

$$q(t)\frac{S''(t)}{(1-\alpha)^{\underline{2}}} + \tilde{q}(t)\frac{S'(t)}{1-\alpha} + v_2(1+t)^{\alpha-1}S(t) = 0, \tag{4.15}$$

where

$$q(t) = t + (w-v_2t)(1-(1+t)^{\alpha}) \quad \text{and} \quad \tilde{q}(t) = v_2 - 1 + (w - v_2(1+2t))(1+t)^{\alpha-1}. \tag{4.16}$$

The first initial condition is $S(0) = 1$ since ψ is a stationary distribution. The second condition takes a little more work. Using a result of Ethier and Kurtz [1986], ψ is a stationary distribution associated with the generator $\mathscr{L}_{\Lambda M}$ if and only if

$$\int_0^1 \mathscr{L}_{\Lambda M}(y^n)\psi(y)dy = 0, \quad \forall n \in \mathbb{N}. \tag{4.17}$$

Therefore, taking $n = 1$, we get the second condition

$$\frac{S'(0)}{1-\alpha} = \frac{v_2}{v_1 + v_2}.$$

Using computer algebra software (Maple), we find that the solution is given by

$$S(t) = 1 + (\alpha - 1)e^{(1-\alpha)I(t)} \int_0^t \frac{1 - p(z)}{q(z)} e^{(\alpha-1)I(z)} dz \tag{4.18}$$

where

$$p(z) = 1 - v_2(1 - (1+z)^\alpha) \quad \text{and} \quad I(z) = \int_0^z \frac{p(\tilde{z})}{q(\tilde{z})} d\tilde{z}, \tag{4.19}$$

and q is defined above in (4.16). This solution is not final, as we have yet to invert the transform to obtain ψ, but it does encode some information, such as the moments.

4.4.2 Moments

Suppose that Y is a random variable with density ψ. We can view $S(t)$ as an alternative to the moment generating function. Observe that

$$\begin{aligned} S(t) &= \int_0^1 (1+ty)^{1-\alpha}\psi(y)dy = \int_0^1 \psi(y)dy \sum_{k=0}^{\infty} \binom{1-\alpha}{k} (ty)^k \\ &= \sum_{k=0}^{\infty} \binom{1-\alpha}{k} t^k \int_0^1 y^k \psi(y)dy = \sum_{k=0}^{\infty} \frac{t^k}{k!} (1-\alpha)^{\underline{k}} \mathbb{E}^\psi[Y^k], \end{aligned}$$

where $z^{\underline{k}}$ is the falling factorial defined in Appendix A. By Fubini's theorem, the interchange of the integral and the sum is valid since the integrand is non-negative. Hence, the general form for the moments is controlled by the derivatives of S:

$$\mathbb{E}^\psi[Y^k] = \frac{S^{(k)}(0)}{(1-\alpha)^{\underline{k}}}, \quad k \in \mathbb{N}. \tag{4.20}$$

We do in fact get the first moment for free – it is the initial condition of Equation (4.15),

$$\mathbb{E}^\psi[Y] = \frac{v_2}{v_1 + v_2}.$$

To calculate the second moment, we need the first two derivatives of S which are

$$\begin{aligned} S'(t) &= \frac{(1-\alpha)(p(t)S(t)-1)}{q(t)}, \\ S''(t) &= \frac{(1-\alpha)(p'(t)S(t)+p(t)S'(t))-q'(t)S'(t)}{q(t)} =: \frac{n_2(t)}{q(t)}. \end{aligned} \tag{4.21}$$

It turns out that $n_2(0) = 0 = q(0)$, meaning that an application of l'Hôpital's rule is necessary. For this, we need

$$n_2'(t) = (1-\alpha)(p''(t)S(t)+2p'(t)S'(t)+p(t)S''(t)) - (q''(t)S'(t)+q'(t)S''(t)).$$

In the sequel, we use the algebra of limits in the following way: whenever we use one of the rules, we postulate that the limits on the right-hand side exist, and derive a contradiction later if they do not. In particular, we shall assert that $S''(0)$ exists, and see that no contradictions surface. So,

$$\begin{aligned} S''(0) &= \lim_{t\downarrow 0} \frac{n_2(t)}{q(t)} \overset{\text{l'H}}{=} \lim_{t\downarrow 0} \frac{n_2'(t)}{q'(t)} \\ &= \lim_{t\downarrow 0} \frac{(1-\alpha)(p''(t)S(t)+2p'(t)S'(t))-q''(t)S'(t)}{q'(t)} \\ &\quad + (1-\alpha)S''(0)\lim_{t\downarrow 0}\frac{p(t)}{q'(t)} - S''(0)\lim_{t\downarrow 0}\frac{q'(t)}{q'(t)}. \end{aligned}$$

By rearranging for $S''(0)$ and using the definitions of $p(t), q(t), S(t)$ and their derivatives, we obtain

$$S''(0) = \frac{1}{2-\frac{1-\alpha}{1-\alpha w}} \frac{(1-\alpha)\left(p''(0)S(0)+2p'(0)S'(0)\right)-q''(0)S'(0)}{q'(0)}, \tag{4.22}$$

and we can use this to find an expression for the second moment

$$\mathbb{E}^{\psi}[Y^2] = \frac{S''(0)}{(1-\alpha)^2} = \frac{v_2}{v_1+v_2}\frac{1-\alpha+2\alpha v_2}{1+\alpha-2\alpha w},$$

and the variance

$$\text{Var}^{\psi}(Y) = \mathbb{E}^{\psi}[Y^2] - \mathbb{E}^{\psi}[Y]^2 = \frac{1-\alpha}{1+\alpha-2\alpha w}\frac{v_1 v_2}{(v_1+v_2)^2}.$$

Since Y is not in the original space – we made the linear transformation (4.4) in order to reach the alternative representation – we need to use the inverse (4.5) to return to the original space. For example,

if X is a random variable with density $\tilde{\psi}$, then

$$\mathbb{E}^{\tilde{\psi}}[X] = v_2 + w\mathbb{E}^{\psi}[Y] = \frac{v_2}{v_1 + v_2} = \mathbb{E}^{\psi}[Y] \quad \text{and} \quad \mathrm{Var}^{\tilde{\psi}}(X) = w^2 \mathrm{Var}^{\psi}(Y). \tag{4.23}$$

It is possible to derive explicit expressions for $\mathbb{E}^{\psi}[Y^k]$, $k \in \mathbb{N}$. The procedure is very similar to the above, just in more generality. Firstly, form a recursion for the derivatives of $S(t)$ (Lemma C.4.5) evaluated at zero (Lemma C.4.6). Whilst those equations are rather cumbersome, the reader should bear in mind that the proofs simply generalize the argument above for finding $S''(0)$. Having found a recursion for $S^{(k)}(0)$, Equation (C.3) can be manipulated further to arrive at an expression for the moments of Y:

$$\mathbb{E}^{\psi}[Y^k] = \frac{\sum_{i=0}^{k-1} \left((1-\alpha)\binom{k}{i} p^{(k-i)}(0) - \binom{k}{i-1} q^{(k-i+1)}(0) \right) S^{(i)}(0)}{(1 + \alpha - 2\alpha w)\,(1-\alpha)^{\underline{k}}}, \tag{4.24}$$

using the convention $\binom{k}{-1} = 0$. With the help of Lemma C.4.7, the moments of Y are explicit, albeit cumbersome, up to a recursion. Finally, $\mathbb{E}^{\tilde{\psi}}[X^n]$, the nth moment of X, can be found by expressing it as a linear combination of $\mathbb{E}^{\psi}[Y^k]$, $k \in [n]_0$.

4.4.3 Inversion

We have yet to obtain an expression for ψ. In order to do this, we have to invert the transform $S(t)$. Recall from (4.14) that our transformation is given by

$$S(t) = \int_0^1 (1+ty)^{1-\alpha} \psi(y) dy, \quad t \in (0, \tfrac{1}{2}),\ \alpha \in (0,1).$$

A simple application of integration by parts reveals that we can rewrite $S(t)$ in terms of the Stieltjes transform.

$$\begin{aligned} S(t) &= \left[(1+ty)^{1-\alpha} \int_0^y \psi(\tilde{y}) d\tilde{y} \right]_{y=0}^{y=1} - (1-\alpha)t \int_0^1 (1+ty)^{-\alpha} \left(\int^y \psi(\tilde{y}) d\tilde{y} \right) dy \\ &= (1+t)^{1-\alpha} \int_0^1 \psi(\tilde{y}) d\tilde{y} - \int_0^0 \psi(\tilde{y}) d\tilde{y} - (1-\alpha)t \int_0^1 \frac{\Psi(y)}{(1+ty)^{\alpha}} dy \\ &= (1+t)^{1-\alpha} - (1-\alpha)t\tilde{S}(t), \end{aligned}$$

where Ψ is the *distribution function* of ψ, and $\tilde{S}(t)$ is its generalized Stieltjes transform of order α. As S is known explicitly, it follows that we need to apply the inversion formula to

$$\tilde{S}(t) = \frac{1}{(1-\alpha)t}\left((1+t)^{1-\alpha} - 1 + (1-\alpha)e^{(1-\alpha)I(t)}J(t)\right),$$

where

$$J(t) := \int_0^t \frac{1-p(z)}{q(z)} e^{(\alpha-1)I(z)} dz.$$

Yano [2006] defines the inversion formula through an auxiliary function

$$\phi(\sigma) := \lim_{\eta \downarrow 0} \frac{1}{2\pi i}\left(\tilde{S}(-\sigma - i\eta) - \tilde{S}(\sigma + i\eta)\right), \tag{4.25}$$

from which the desired function is obtained via the fractional integral

$$\Psi(y) = \int_{0+}^{y-} (y-\xi)^{\alpha-1}\phi(\xi)d\xi. \tag{4.26}$$

In practice, finding ϕ is going to be very difficult, given the complexity of $\tilde{S}$. We can make some progress, however. In the case of the Beta coalescent, ϕ reduces to

$$\begin{aligned}\phi(\sigma) &= \lim_{\eta\downarrow 0} \frac{(\sigma^2+\eta^2)^{-1}}{2\pi i}\left((-\sigma+i\eta)e^{(1-\alpha)I(-\sigma-i\eta)}J(-\sigma-i\eta) + (\sigma+i\eta)e^{(1-\alpha)I(-\sigma+i\eta)}J(-\sigma+i\eta)\right)\\ &= \lim_{\eta\downarrow 0}\frac{1}{2\pi i}\int_0^1 Q_-(t)e^{(1-\alpha)I^-_{t,1}} - Q_+(t)e^{(1-\alpha)I^+_{t,1}}dt,\end{aligned}$$

where

$$I^{\pm}_{t,1} := \int_{(-\sigma\pm i\eta)t}^{-\sigma\pm i\eta} \frac{p(z)}{q(z)}dz \quad \text{and} \quad Q_{\pm}(t) := \left.\frac{1-p(z)}{q(z)}\right|_{z=(-\sigma\pm i\eta)t}$$

and the functions p and q are defined in (4.19) and (4.16) respectively.

4.5 The Genealogy

In this section we aim to examine the genealogical process associated with (4.1), which will turn out to be a variant of the Λ-coalescent. Ultimately, we would like to find the genealogy when we condition on knowing the types in the sample; before doing so we shall briefly motivate this method by contrasting it

to the unconditioned version. Both of these were discussed in Section 2.3.4 in the context of duality.

Suppose you have a sample of size $k \in \mathbb{N}$. To obtain a genealogical relationship between the lineages in this sample, perform the following steps.

(i) Generate a Λ-coalescent with k leaves (using the algorithm in Section 3.4.1, for example).

(ii) Assign a type to the root node by sampling from the stationary distribution associated with (4.1). (More generally, one could sample from the non-stationary distribution of types at some time in the past.)

(iii) Starting at the root node, run the mutation process down the tree: for each branching event, and for each lineage in that event, if the parent is of type a_i, the newborn will be of the same type with probability $1 - v_i$ or mutate to the other type with probability v_i.

Upon completion, this will yield a fully typed realization of the genealogy for the sample. The trouble with this method is that, although simple to simulate, it can be rather wasteful. Typically, one already knows the types in their sample, so if the pattern of types in the generated genealogy does not match that of the data (which is quite likely to happen), then that tree must be thrown away. For large sample sizes, this form of rejection sampling could be very expensive indeed. In addition, it may be difficult to perform step (ii) as, although we studied the stationary distribution in Section 4.4, we didn't characterize it enough to be able to draw a sample from it. The conditioned genealogy, on the other hand, depends only on the stationary distribution through its moments, which were calculated in Section 4.4.2 when Λ was the Beta measure.

4.5.1 Conditioned Genealogy: the Typed Λ-Coalescent

Another approach to obtain a genealogy is to look for a weak duality for the generator (4.1). In particular, we look for duality which reflects the fact that we have knowledge of the types in our sample – we assume that the initial configuration of the sample in question is $\boldsymbol{k} = (k_1, k_2)$, where k_i is the number of lineages with type a_i. We impose that $\boldsymbol{k} > \mathbf{0}$, by which we mean that both k_1 and k_2 are non-negative and that $|\boldsymbol{k}| > 0$.

This method was outlined originally in Barbour *et al.* [2000], subsequently applied in later papers such as Etheridge *et al.* [2010] and described in full generality in Section 2.2.1. In this section, we will apply it to the Λ-Fleming–Viot process with strong mutation. Given the vector $\boldsymbol{k}$, we seek a duality function $f_{\boldsymbol{k}}(x)$ such that, when substituted into (4.1), we obtain the same expression as if we applied the generator of a suitable continuous-time discrete state space jump process to the function with x fixed and $\boldsymbol{k}$ varying. Duality will then follow through Proposition 2.6. We begin by considering the unnormalized duality of (2.34),

$$f_{\boldsymbol{k}}(x) = x^{k_1}(1-x)^{k_2}, \tag{4.27}$$

which, when applied to the generator $\mathscr{L}_{\Lambda M}$, can be written in the form of (2.11). We will use the methods of Section 2.2.1 to derive the appropriate (normalized) duality function: one has to correct (4.27) by dividing by the appropriate moments of the stationary distribution of the forward process (which were discussed in Section 4.4 for a special case). More precisely, the normalized duality function is

$$g_{\boldsymbol{k}}(x) = \frac{f_{\boldsymbol{k}}(x)}{m_{\boldsymbol{k}}} \tag{4.28}$$

where

$$m_{\boldsymbol{k}} = \mathbb{E}[f_{\boldsymbol{k}}(X)] \tag{4.29}$$

and X is governed by the stationary distribution of the ΛFVM process. For example, when we specialize to the measure $\Lambda^{\alpha}(du)$, discussed in Section 4.4.1, the stationary distribution has density $\tilde{\psi}$ and we can express identities (4.23) as

$$m_{(1,0)} = \mathbb{E}^{\tilde{\psi}}[X] = \frac{v_2}{v_1+v_2} \quad \text{and} \quad m_{(0,1)} = \mathbb{E}^{\tilde{\psi}}[1-X] = \frac{v_1}{v_1+v_2}. \tag{4.30}$$

In many cases, the moments $m_{\boldsymbol{k}}$ may be unobtainable or difficult to calculate, but the method of duality holds nonetheless, provided the stationary distribution exists and has finite moments. The general result will be given in Theorem 4.7. In what follows, we shall demonstrate how the method works by example, concentrating on the case when Λ is the Beta measure Λ^{α} of (3.10) and using the moments that were calculated in Section 4.4.2. The example is intended to provide intuition of how the genealogy

behaves.

For the first example, we consider the most simple configuration we can think of: $\boldsymbol{k} = (1, 0)$. Then, we observe that

$$\begin{aligned}\mathscr{L}_{\Lambda M} g_{(1,0)}(x) &= \frac{1}{m_{(1,0)}} \int_0^1 \left[x((1-u)x + u(1-v_1)) + (1-x)((1-u)x + uv_2) - x\right] \frac{\Lambda^\alpha(du)}{u^2} \\ &= v_2 \frac{m_{(0,1)}}{m_{(1,0)}} \int_0^1 u \frac{\Lambda^\alpha(du)}{u^2} \left(g_{(0,1)}(x) - g_{(1,0)}(x)\right), \end{aligned} \tag{4.31}$$

where we have used the relationship

$$\frac{m_{(0,1)}}{m_{(1,0)}} = \frac{v_1}{v_2}, \tag{4.32}$$

which follows from (4.30).

We have shown that, from the configuration $(1, 0)$, the genealogical process will jump to the state $(0, 1)$ at rate

$$q[(1,0),(0,1)] = v_2 \frac{m_{(0,1)}}{m_{(1,0)}} \int_0^1 u \frac{\Lambda^\alpha(du)}{u^2} = \frac{v_1}{1-\alpha}, \tag{4.33}$$

where $q(\boldsymbol{k}, \boldsymbol{k}')$ denotes the rate at which the process jumps from state $\boldsymbol{k}$ to $\boldsymbol{k}'$, $\boldsymbol{k} \neq \boldsymbol{k}'$. This notation coincides with that used in Section 2.2.1. (Square brackets will be used instead of round ones when referring to specific vectors in order to aid visual bracket-balancing.)

Whilst $q[(1, 0), (0, 1)]$ in (4.33) has a very slick expression, the clunkier version will help us to *interpret* the rate probabilistically, and help us to identify more general rates for this process. In understanding the rates of changes in state of the backward process, we must consider the configurations and events that, going forwards, might have led to what we see now. For this reason, we appeal to the forward process, and the notation thereof, to derive the backward process.

The factor $\int_0^1 \frac{\Lambda(du)}{u}$ in (4.33) is simply the probability that, in a coalescence event, our single lineage is picked to be part of the 'merger' (we use the term in a loose sense since the merger, from the point of view of the sample, is trivial). Notice that once again we use condition (4.2), which now has an interpretation from a coalescence point of view: in the usual Λ-coalescent, mergers involving one block are not allowed, but in this model such mergers are permitted since they can result in a change of state. For this reason we need the more stringent condition.

The other factor in (4.33) is the mutation probability. That is, in order to see a change in state going backwards, we need that, in addition to the fact that our lineage was an offspring of the reproduction event that gave rise to it, that it mutated from the parent's type (a_2) to its current type (a_1). Really, it is an application of Bayes' Theorem: suppose that L_t is the random variable marking the type of our one lineage, where t is genealogical time (ergo $L_0 = a_1$). Then,

$$\mathbb{P}(L_{t+\delta t} = a_2 | L_t = a_1) = \mathbb{P}(L_t = a_1 | L_{t+\delta t} = a_2) \frac{\mathbb{P}(L_{t+\delta t} = a_2)}{\mathbb{P}(L_t = a_1)} = v_2 \frac{m_{(0,1)}}{m_{(1,0)}} \delta t + o(\delta t).$$

That $\mathbb{P}(L_t = a_1 | L_{t+\delta t} = a_2) = v_2 \delta t + o(\delta t)$ follows from the fact that reverse genealogical time is in fact the direction of time in the forward process, and v_2 is the probability that the a_2 individual mutates to type a_1 forwards in time.

Suppose that we want to calculate the probability of sampling a particular configuration (of a given size) from the stationary distribution X. Let $\boldsymbol{K}$ be a random vector representing this sample. Then, *given* X, the probability that the first lineage sampled is of type a_1 is just X, whilst the probability that it is of type a_2 is $1 - X$. Since the population size is regarded as infinite, the sampling of each lineage is independent, implying that $\mathbb{P}(\boldsymbol{K} = (k_1, k_2) | X) = X^{k_1}(1 - X)^{k_2}$. Partitioning over X,

$$\mathbb{P}(\boldsymbol{K} = (k_1, k_2)) = \mathbb{E}[X^{k_1}(1 - X)^{k_2}] = m_{\boldsymbol{k}}.$$

In particular, sampling just one lineage, $\mathbb{P}(\boldsymbol{K} = (1, 0)) = \mathbb{P}(L_t = a_1) = m_{(1,0)}$ and $\mathbb{P}(\boldsymbol{K} = (0, 1)) = \mathbb{P}(L_{t+\delta t} = a_2) = m_{(0,1)}$, using the above notation.

In general, because we are thinking about the rates of events going backwards in terms of forward probabilities, these Bayes factors will always be a multiplicative factor in the rate $q(\boldsymbol{k}, \boldsymbol{k}')$ of the form

$$\frac{\mathbb{P}(\text{new configuration})}{\mathbb{P}(\text{current configuration})} = \frac{\mathbb{P}(\boldsymbol{K} = \boldsymbol{k}')}{\mathbb{P}(\boldsymbol{K} = \boldsymbol{k})} = \frac{m_{\boldsymbol{k}'}}{m_{\boldsymbol{k}}}.$$

A similar calculation, analogous to (4.31), for the initial configuration $\boldsymbol{k} = (0, 1)$ leads to the generator

$$\mathscr{L}_{\Lambda M} g_{(0,1)}(x) = v_1 \frac{m_{(1,0)}}{m_{(0,1)}} \int_0^1 u \frac{\Lambda^\alpha(du)}{u^2} \left(g_{(1,0)}(x) - g_{(0,1)}(x) \right), \tag{4.34}$$

which means that, from state $(0,1)$, the only possible transition is to $(1,0)$, and this occurs at rate

$$q[(0,1),(1,0)] = v_1 \frac{m_{(1,0)}}{m_{(0,1)}} \int_0^1 u \frac{\Lambda^\alpha(du)}{u^2} = \frac{v_2}{1-\alpha},$$

which is symmetrical to (4.33).

We shall now explore the cases for which we have two lineages in the initial configuration, arguing probabilistically for general Λ, starting with $\boldsymbol{k} = (2,0)$. There are three possible states to which this process might jump. Firstly, this configuration could have arisen through a mutation on one of the lineages from the state $(1,1)$. Since mutation is intrinsically linked to reproduction, the mutating line had to be involved in a reproduction event for this to have happened. More precisely, one of the lineages had to be in the event (and mutate) and the other one must not have been. For the same reason, the transition $(2,0) \to (0,2)$ is *not* possible in a single sampling event – the two lineages cannot mutate independently as there can only be one coalescence at a time (this is a Λ-, not a Ξ-coalescent). In fact, if they both mutated, they would have to come from the same mother. The rate we seek is

$$q[(2,0),(1,1)] = 2v_2 \frac{m_{(1,1)}}{m_{(2,0)}} \int_0^1 u(1-u) \frac{\Lambda(du)}{u^2}. \tag{4.35}$$

The factor $\int_0^1 u(1-u)\frac{\Lambda(du)}{u^2}$ represents the probability of selecting one lineage but not the other. The Bayes factor is adjusted to take into account the transition; the factor v_2 means that the selected lineage must have made the mutation $a_2 \to a_1$ going forwards; the factor of 2 is purely combinatorial and accounts for the fact that the mutation could have happened to either lineage.

We have covered the case when there is a trivial coalescence. There are two other ways that the state $(2,0)$ could have arisen – in both cases the lineages came from the same mother in that reproduction event. Firstly, if the mother had type a_1 then neither child mutated at birth; if the mother had type a_2, then both mutated at birth. The corresponding rates are

$$q[(2,0),(1,0)] = (1-v_1)^2 \frac{m_{(1,0)}}{m_{(2,0)}} \int_0^1 u^2 \frac{\Lambda(du)}{u^2}, \tag{4.36}$$

$$q[(2,0),(0,1)] = v_2^2 \frac{m_{(0,1)}}{m_{(2,0)}} \int_0^1 u^2 \frac{\Lambda(du)}{u^2}. \tag{4.37}$$

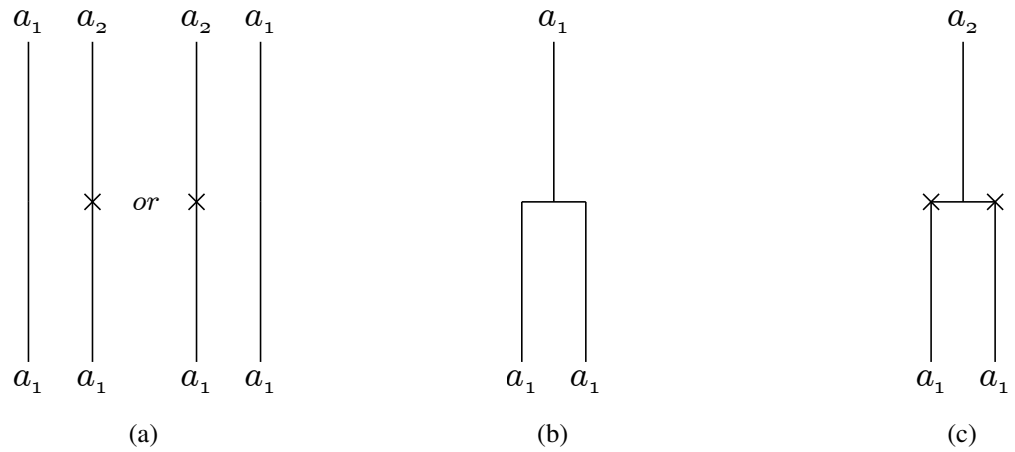


Figure 4.1: The possible transitions from the state $(2,0)$ to $(1,1)$ ((a), at rate (4.35)), $(1,0)$ ((b), at rate (4.36)) and $(0,1)$ ((c), at rate (4.37)). No other transitions are possible. Crosses signify that there has been a mutation event from one allelic type to the other

The possible transitions are presented pictorially in Figure 4.1.

For the special case when Λ is taken to be the Beta coalescent, we obtain the following lemma.

Lemma 4.6. *For each $\alpha \in (0,1)$, in the special case when Λ is taken to be the Beta measure Λ^{α}, the only possible transitions from the initial configuration $\boldsymbol{k} = (2,0)$ are to the states $(1,1)$, $(1,0)$ or $(0,1)$ which occur at rates given by (4.35), (4.36) and (4.37) respectively. This is summarized by the generator equation*

$$\mathscr{L}_{\Lambda M} g_{(2,0)}(x) = q[(2,0),(1,1)](g_{(1,1)}(x) - g_{(2,0)}(x)) + q[(2,0),(1,0)](g_{(1,0)}(x) - g_{(2,0)}(x)) + q[(2,0),(0,1)](g_{(0,1)}(x) - g_{(2,0)}(x)).$$

This will also be true for all other Λ for which condition (4.2) holds. The proof of Theorem 4.7, which holds in more generality, renders a proof of this result unnecessary. However, one can check it explicitly, and it is reassuring to see that the moments derived in Section 4.4.2, and the general methodology, are in good working order.

The initial configuration $\boldsymbol{k} = (0,2)$ is purely symmetrical to the case we have just studied; it is just

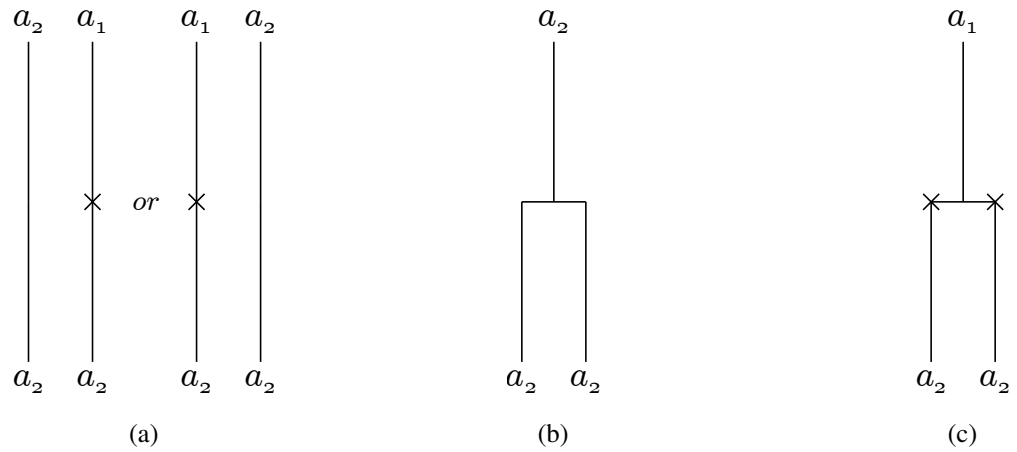


Figure 4.2: The possible transitions from the state $(0,2)$ to $(1,1)$ ((a), at rate (4.38)), $(0,1)$ ((b), at rate (4.39)) and $(1,0)$ ((c), at rate (4.40)). No other transitions are possible. Crosses signify that there has been a mutation event from one allelic type to the other. This figure demonstrates the symmetry between $\boldsymbol{k} = (2,0)$ and $(0,2)$ – to get from Figure 4.1 to this one, just switch the labels so that a_1 becomes a_2 and a_2 becomes a_1

an exercise in swapping labels. The likeness can be seen from the rates of transitions

$$q[(0,2),(1,1)] = 2v_1 \frac{m_{(1,1)}}{m_{(0,2)}} \int_0^1 u(1-u) \frac{\Lambda(du)}{u^2}, \tag{4.38}$$

$$q[(0,2),(0,1)] = (1-v_2)^2 \frac{m_{(0,1)}}{m_{(0,2)}} \int_0^1 u^2 \frac{\Lambda(du)}{u^2}, \tag{4.39}$$

$$q[(0,2),(1,0)] = v_1^2 \frac{m_{(1,0)}}{m_{(0,2)}} \int_0^1 u^2 \frac{\Lambda(du)}{u^2}, \tag{4.40}$$

and pictorially from Figure 4.2.

The case of the configuration $\boldsymbol{k} = (1,1)$ is slightly different since a single mutation will result in a different state depending on which lineage it falls upon. Therefore, four possible states are available. By now, the best way to convey the process from this state is through Figure 4.3 and equations (4.41)-(4.44).

$$q[(1,1),(2,0)] = v_1 \frac{m_{(2,0)}}{m_{(1,1)}} \int_0^1 u(1-u) \frac{\Lambda(du)}{u^2}, \tag{4.41}$$

$$q[(1,1),(0,2)] = v_2 \frac{m_{(0,2)}}{m_{(1,1)}} \int_0^1 u(1-u) \frac{\Lambda(du)}{u^2}, \tag{4.42}$$

$$q[(1,1),(1,0)] = v_1(1-v_1) \frac{m_{(1,0)}}{m_{(1,1)}} \int_0^1 u^2 \frac{\Lambda(du)}{u^2}, \tag{4.43}$$

$$q[(1,1),(0,1)] = v_2(1-v_2) \frac{m_{(0,1)}}{m_{(1,1)}} \int_0^1 u^2 \frac{\Lambda(du)}{u^2}. \tag{4.44}$$

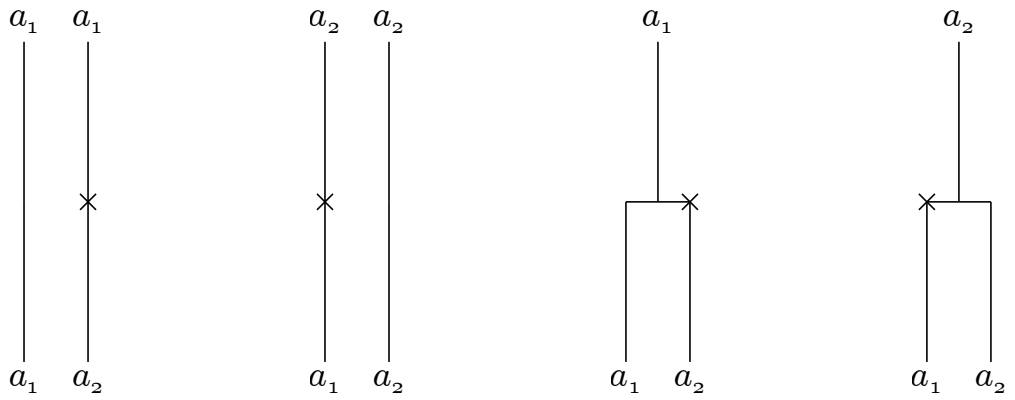


Figure 4.3: The possible transitions from the state $(1,1)$ to $(2,0)$ ((a), at rate (4.41)), $(0,2)$ ((b), at rate (4.42)), $(1,0)$ ((c), at rate (4.43)) and $(0,1)$ ((d), at rate (4.44)). No other transitions are possible. Crosses signify that there has been a mutation event from one allelic type to the other

Given the rates of transitions (4.38)-(4.44) for the backwards process, we can state and prove results analogous to Lemma 4.6 for the initial configurations $\boldsymbol{k} = (0,2)$ and $(1,1)$. However, as this will reveal no new insight, and they are characterized by Figures 4.2 and 4.3, we choose to omit them. Moreover, they are special cases of Theorem 4.7.

For $|\boldsymbol{k}| \leq 2$, we have now given a complete description of the genealogy using probabilistic arguments. We could continue in this manner ad nauseam, but the complexity of each individual case grows with $|\boldsymbol{k}|$. Instead, thinking more generally, we can write down the process for any configuration $\boldsymbol{k} > \boldsymbol{0}$. Before doing so, we need some notation.

For vectors $\boldsymbol{k} > \boldsymbol{0}$, define the set $L_{\boldsymbol{k}} := \{\boldsymbol{l} \in \mathbb{Z}^2 : \boldsymbol{0} < \boldsymbol{l} \leq \boldsymbol{k}\}$, where $\boldsymbol{l} \leq \boldsymbol{k}$ means that $l_i \leq k_i, i \in \{1,2\}$, with no further restrictions, and $\boldsymbol{l} > \boldsymbol{0}$ retains its usual meaning: that $l_i \geq 0$ and $|\boldsymbol{l}| > 0$.

Theorem 4.7. *Fix $\boldsymbol{k} > \boldsymbol{0}$ and let Λ be a measure satisfying (4.2). For each $\boldsymbol{l} \in L_{\boldsymbol{k}}$, $i \in \{1,2\}$, the transition $\boldsymbol{k} \to \boldsymbol{k} - \boldsymbol{l} + \boldsymbol{e}_i$ in the dual of the Λ-Fleming–Viot process with strong mutation occurs at rate*

$$q(\boldsymbol{k}, \boldsymbol{k} - \boldsymbol{l} + \boldsymbol{e}_i) = \begin{cases} \binom{k_1}{l_1}\binom{k_2}{l_2} v_i^{|\boldsymbol{l}|-l_i}(1-v_i)^{l_i} \dfrac{m_{\boldsymbol{k}-\boldsymbol{l}+\boldsymbol{e}_i}}{m_{\boldsymbol{k}}} \displaystyle\int_0^1 u^{|\boldsymbol{l}|-2}(1-u)^{|\boldsymbol{k}|-|\boldsymbol{l}|}\Lambda(du), & |\boldsymbol{l}| \geq 2, \\ k_j v_i \dfrac{m_{\boldsymbol{k}-\boldsymbol{e}_j+\boldsymbol{e}_i}}{m_{\boldsymbol{k}}} \mathbb{1}_{\{i \neq j\}} \displaystyle\int_0^1 u^{-1}(1-u)^{|\boldsymbol{k}|-1}\Lambda(du), & \boldsymbol{l} = \boldsymbol{e}_j. \end{cases}$$

Hence, we have the generator equation

$$\mathscr{L}_{\Lambda M} g_{\boldsymbol{k}}(x) = \sum_{\boldsymbol{l} \in L_{\boldsymbol{k}}} \sum_{i=1,2} q(\boldsymbol{k}, \boldsymbol{k} - \boldsymbol{l} + \boldsymbol{e}_i) \left[g_{\boldsymbol{k}-\boldsymbol{l}+\boldsymbol{e}_i}(x) - g_{\boldsymbol{k}}(x) \right].$$

This theorem can either be understood in terms of the pictorial arguments we gave above, or in terms of equations. We postpone the rigorous proof until we have given a more intuitive explanation.

The vector $\boldsymbol{l}$ represents the set of lineages eligible to take part in the merger; it can consist of lineages of both types, reflecting the fact that, going forward, the mother of a reproduction event may give birth to mutant and non-mutant offspring. The vector $\boldsymbol{l} = \boldsymbol{e}_j$ represents the 'trivial' merger, where just one of the lineages mutates to the other type (it is important that there is indeed a mutation – hence the condition $\boldsymbol{e}_j \neq \boldsymbol{e}_i$ – otherwise we would not see a jump in state). At the other extreme, $\boldsymbol{l} = \boldsymbol{k}$ represents the event that *all* of the lineages in the current configuration came from the same mother. Anything between these two extremes is possible due to the multiple-merger nature of the Λ-coalescent.

Given $\boldsymbol{l}$, the sub-configuration of lineages that could merge, jumps to only two states are possible: $\boldsymbol{k} - \boldsymbol{l} + \boldsymbol{e}_i, i \in \{1, 2\}$. That is, the $|\boldsymbol{l}|$ lineages will collapse into just one, which will have the type of the mother. More precisely, for the new state, we subtract l_i from k_i, but there is a bonus one for the mother's type, represented by $\boldsymbol{e}_i$.

Now, given both $\boldsymbol{l} \in L_{\boldsymbol{k}}$ and $i \in \{1, 2\}$, we must calculate the probability that, going forwards, the configuration $\boldsymbol{k} - \boldsymbol{l} + \boldsymbol{e}_i$ gave rise to the current configuration of types. Firstly, the combinatorial factor $\binom{k_1}{l_1}\binom{k_2}{l_2}$ counts the number of ways that, from the configuration $\boldsymbol{k}$, we can form subsets of size $|\boldsymbol{l}|$ such that l_1 are of type a_1 and l_2 are of type a_2. Then, from a mother of type a_i, it must have been that l_i of the $|\boldsymbol{l}|$ offspring that we have sampled did *not* mutate (which, individually, has probability $1 - v_i$), whilst the remaining $|\boldsymbol{l}| - l_i$ did mutate to the other type (with probability v_i). Since each mutation event is independent, we must multiply them all together, resulting in the mutation factor $v_i^{|\boldsymbol{l}|-l_i}(1 - v_i)^{l_i}$.

As described earlier in this section, the Bayes factor $\frac{m_{\boldsymbol{k}-\boldsymbol{l}+\boldsymbol{e}_i}}{m_{\boldsymbol{k}}}$ is an artefact of how we describe the genealogy in terms of the forward process, and must always take this form. Finally, the integral factor emerges from the general theory of Λ-coalescents and represents the probability that, of our $|\boldsymbol{k}|$ lineages, $|\boldsymbol{l}|$ are selected to be in the merger and the remaining $|\boldsymbol{k}| - |\boldsymbol{l}|$ are not.

Proof. The strategy is to apply the methods of Section 2.2.1, and in particular to rewrite the generator of the forwards-in-time model in the form of (2.11) when acting on functions $f_{\boldsymbol{k}}(x) = x^{k_1}(1-x)^{k_2}$. That is,

$$\begin{aligned}
\mathscr{L}_{\Lambda M} f_{\boldsymbol{k}}(x) = \int_0^1 \big[&x((1-u)x + u(1-v_1))^{k_1}((1-u)(1-x) + uv_1)^{k_2} \\
&+ (1-x)((1-u)x + uv_2)^{k_1}((1-u)(1-x) + u(1-v_2))^{k_2} - x^{k_1}(1-x)^{k_2}\big]\nu(du) \\
= \int_0^1 \Bigg[&\sum_{i=0}^{k_1}\binom{k_1}{i}(1-v_1)^i u^i (1-u)^{k_1-i}x^{k_1-i+1}\sum_{j=0}^{k_2}\binom{k_2}{j}v_1^j u^j (1-u)^{k_2-j}(1-x)^{k_2-j} \\
&+ \sum_{i=0}^{k_1}\binom{k_1}{i}v_2^i u^i (1-u)^{k_1-i}x^{k_1-i}\sum_{j=0}^{k_2}\binom{k_2}{j}(1-v_2)^j u^j (1-u)^{k_2-j}(1-x)^{k_2-j+1} \\
&- x^{k_1}(1-x)^{k_2}\Bigg]\nu(du)
\end{aligned}$$

where, in the first line we simply plugged $f_{\boldsymbol{k}}(x)$ into the definition of generator (4.1), and in the second we performed binomial expansions of the first two terms. What we are left with in the first two terms are products of two sums. To continue, we shall peel off the zeroth term in all sums and multiply out to produce four terms per product;

$$\begin{aligned}
\mathscr{L}_{\Lambda M} f_{\boldsymbol{k}}(x) = \int_0^1 \big[&(1-u)^{|\boldsymbol{k}|}x^{k_1+1}(1-x)^{k_2} + \sum_{i=1}^{k_1}\binom{k_1}{i}(1-v_1)^i u^i (1-u)^{|\boldsymbol{k}|-i}x^{k_1-i+1}(1-x)^{k_2} \\
&+ \sum_{j=1}^{k_2}\binom{k_2}{j}v_1^j u^j (1-u)^{|\boldsymbol{k}|-j}x^{k_1+1}(1-x)^{k_2-j} \\
&+ \sum_{i=1}^{k_1}\sum_{j=1}^{k_2}\binom{k_1}{i}\binom{k_2}{j}v_1^j (1-v_1)^i u^{i+j}(1-u)^{|\boldsymbol{k}|-(i+j)}x^{k_1-i+1}(1-x)^{k_2-j} \\
&+ (1-u)^{|\boldsymbol{k}|}x^{k_1}(1-x)^{k_2+1} + \sum_{i=1}^{k_1}\binom{k_1}{i}v_2^i u^i (1-u)^{|\boldsymbol{k}|-i}x^{k_1-i}(1-x)^{k_2+1} \\
&+ \sum_{j=1}^{k_2}\binom{k_2}{j}(1-v_2)^j u^j (1-u)^{|\boldsymbol{k}|-j}x^{k_1}(1-x)^{k_2-j+1} \\
&+ \sum_{i=1}^{k_1}\sum_{j=1}^{k_2}\binom{k_1}{i}\binom{k_2}{j}v_2^i (1-v_2)^j u^{i+j}(1-u)^{|\boldsymbol{k}|-(i+j)}x^{k_1-i}(1-x)^{k_2-j+1} \\
&- x^{k_1}(1-x)^{k_2}\big]\nu(du).
\end{aligned}$$

The final objective is to separate out all those terms which are a multiple of $f_{\boldsymbol{k}}(x)$, in order to identify those which actually differ from it. There are precisely five of these: the two with prefactor $(1-u)^{|\boldsymbol{k}|}$, the final term, and the first summands in the second and seventh terms. Hence,

$$\begin{aligned}\mathscr{L}_{\Lambda M} f_{\boldsymbol{k}}(x) = \int_0^1 \Bigg[&\sum_{i=2}^{k_1} \binom{k_1}{i} (1-v_1)^i u^i (1-u)^{|\boldsymbol{k}|-i} x^{k_1-i+1}(1-x)^{k_2} \\ &+ \sum_{j=1}^{k_2} \binom{k_2}{j} v_1^j u^j (1-u)^{|\boldsymbol{k}|-j} x^{k_1+1}(1-x)^{k_2-j} \\ &+ \sum_{i=1}^{k_1}\sum_{j=1}^{k_2} \binom{k_1}{i}\binom{k_2}{j} v_1^j (1-v_1)^i u^{i+j} (1-u)^{|\boldsymbol{k}|-(i+j)} x^{k_1-i+1}(1-x)^{k_2-j} \\ &+ \sum_{i=1}^{k_1} \binom{k_1}{i} v_2^i u^i (1-u)^{|\boldsymbol{k}|-i} x^{k_1-i}(1-x)^{k_2+1} \\ &+ \sum_{j=2}^{k_2} \binom{k_2}{j} (1-v_2)^j u^j (1-u)^{|\boldsymbol{k}|-j} x^{k_1}(1-x)^{k_2-j+1} \\ &+ \sum_{i=1}^{k_1}\sum_{j=1}^{k_2} \binom{k_1}{i}\binom{k_2}{j} v_2^i (1-v_2)^j u^{i+j} (1-u)^{|\boldsymbol{k}|-(i+j)} x^{k_1-i}(1-x)^{k_2-j+1} \\ &- \Big(1-(1-u)^{|\boldsymbol{k}|} - (k_1(1-v_1)+k_2(1-v_2))u(1-u)^{|\boldsymbol{k}|-1}\Big) x^{k_1}(1-x)^{k_2} \Bigg] \nu(du)\end{aligned} \tag{4.45}$$

We have now written this equation in the form of (2.11), which essentially allows us to read off the jump rates. The positive constants $c(\boldsymbol{k}, \boldsymbol{k}')$ are implicitly defined in this equation, and converting these into the rates $q(\boldsymbol{k}, \boldsymbol{k}')$ introduces the Bayes factors $\frac{m_{\boldsymbol{k}'}}{m_{\boldsymbol{k}}}$. It takes a little work to show that $c(\boldsymbol{k}, \boldsymbol{k})$ is indeed positive, but it is nonetheless true.

In (4.45), the single sums are linked to mergers in which all the children are of the *same* type – either nobody mutated or everybody did. That the index starts from 2, not 1, in two of these summations reflects the fact that a single lineage may not 'mutate to its own type'. The double sums represent all mergers involving both types. By associating index i with l_1 and j with l_2, the result follows. □

Remark 4.8. There is *not* a one-to-correspondence between the cardinality of $L_{\boldsymbol{k}}$ and the number of possible states reachable from configuration $\boldsymbol{k}$: if $\boldsymbol{l}, \boldsymbol{l}' \in L_{\boldsymbol{k}}$ it is possible that the process would jump to the same state from $\boldsymbol{l}$ and $\boldsymbol{l}'$, albeit at different rates.

4.6 Generalizations

Condition (4.2), discussed in Section 4.1.1, is an unfortunate one to have to impose. This is because it restricts the range of possible driving measures Λ. For example, if we would like to use a Beta coalescent for this model, then the parameter range is only $\alpha \in (0,1)$, rather than the full $(0,2)$ that was possible in the model without mutation. Recall that $\Lambda(du) = u^2\nu(du)$ and that X_t denotes the ΛFVM process.

Here, we shall ask whether this condition is really necessary. The answer, found in the following calculation received through personal communication with Tom Kurtz, is yes. We shall use this as motivation to propose more general models which may be studied in the future.

In the following, we shall investigate the stochastic equation of X_t constructed via a Poisson point process. We also allow the mutation mechanism to be generalized, so that it depends randomly on the fraction of the population replaced: let $v_i(u,z)$ be the fraction of individuals that mutate from type a_i to the other type during a reproduction event (z will be replaced by a $\mathcal{U}[0,1]$ random variable).

Let ξ be a Poisson random measure on $[0,\infty) \times (0,1]^3$ with mean measure $ds\nu(du)dz_1dz_2$. In addition, define $\tilde{\xi}$ to be $\xi - ds\nu(du)dz_1dz_2$, the compensated version. Then, the stochastic integral equation for X_t can be written

$$X_t = X_0 + \int_0^t \int_{(0,1]^3} \Big[\mathbb{1}_{\{z_1 \in [0, X_{s-})\}}(z_1) u \left(-X_{s-} + 1 - v_1(u,z_2)\right) + \mathbb{1}_{z_1 \in \{(X_{s-}, 1]\}}(z_1) u \left(-X_{s-} + v_2(u,z_2)\right)\Big] \xi(ds, du, dz_1, dz_2).$$

In order for this stochastic integral to exist, we require that $\int_0^1 u\nu(du) < \infty$, which is equivalent to (4.2). By compensating ξ, we hope to weaken this condition:

$$\begin{aligned} X_t &= X_0 + \int_0^t \int_{(0,1]^3} \Big[\mathbb{1}_{\{z_1 \in [0, X_{s-})\}}(z_1) u \left(-X_{s-} + 1 - v_1(u,z_2)\right) \\ &\qquad + \mathbb{1}_{z_1 \in \{(X_{s-}, 1]\}}(z_1) u \left(-X_{s-} + v_2(u,z_2)\right)\Big] \tilde{\xi}(ds, du, dz_1, dz_2) \\ &\quad + \int_0^t \int_{(0,1]^3} \Big[uX_s \left(1 - X_s - v_1(u,z_2)\right) + u(1-X_s)\left(v_2(u,z_2) - X_s\right)\Big] \nu(du) dz_1 dz_2 ds \\ &= X_0 + M_t + \int_0^t \int_{(0,1]^3} -X_s u v_1(u,z_2) + (1-X_s) u v_2(u,z_2) \nu(du) dz_1 dz_2 ds, \end{aligned}$$

where M_t is defined by the first integral against $\tilde{\xi}$. Since M_t is integrated against a centred Poisson point process, it is a square integrable martingale if $\int_0^1 u^2 \nu(du) = \int_0^1 \Lambda(du) < \infty$ is satisfied [Klebaner, 2005], which coincides with the usual criterion for the existence of the Λ-Fleming–Viot process – see Section 3.1. The second integral exists if

$$\int_{(0,1]^2} u \left(v_1(u,z) + v_2(u,z)\right) \nu(du) dz < \infty. \tag{4.46}$$

In particular, if v_1, v_2 are fixed and do not depend on u, as is the case in the model of Section 4.1, then it follows that (4.2) must hold. However, this insight shows how one might generalize the model to obtain the full range of coalescents that we desire. For example, one could let the mutation parameter v_i be a linear function of u.

CHAPTER 5

Selection

This chapter introduces different forms of natural selection into the two-type Λ-Fleming–Viot process. We shall define the differential killing mechanism in Section 5.1, and the differential birth model in Section 5.2. The word 'differential' refers to the fact that, during reproductive events, the two allelic types are not treated equally, leading to one being selectively favoured. We shall call our two alleles a_1 and a_2, the latter being of the stronger type. The parent will, as in the neutral case, be selected uniformly at random, but the extinction and recolonization phases will be modified in the differential killing and birth models respectively. It is possible to define the models with quite general driving measures $\boldsymbol{\nu}$, but ultimately we are interested in a slightly simpler specialization, as it has a rather intuitive interpretation in terms of strong and weak circles (which will be defined on p.111). This specialized

model is characterized through (5.4, p.110), and most of the chapter will be set in this context.

Biologically, the two models have different interpretations – the differential killing model represents a type of viability selection in which the stronger allelic type exhibits some resistance to certain catastrophes such as disease or plague; the differential birth model is a form of fecundity selection whereby the stronger type is able to produce more offspring. In this chapter, our aim is to prove that these models are governed by the same process of allele frequencies, but have different genealogies, even for small sample sizes.

In Section 5.3 we shall compare and contrast the reproductive mechanisms through Table 5.1 and the graphical construction of Figure 5.1. The models behave similarly in some regards – they are governed by the same of process of allele frequencies, for example. However, Table 5.1 highlights some differences in the way parents displace individuals of their own type during reproduction. This will turn out to be of significant importance when comparing their genealogies.

In Section 5.4, we introduce the classical weak mutation mechanism. This is in order to induce a stationary distribution into the processes, allowing us to employ the normalized duality methods of Section 2.2.1, without destroying the property that the forward-in-time allele frequencies are consistent across models.

Section 5.5 studies the genealogies of these processes in detail. We will again be interested in taking a typed sample, and shall characterize how the sample evolves at various levels of detail. At a coarse level, we study the *simple* genealogy in Section 5.5.2. This can be thought of as the dual process to both models, and has a sleek characterization (Theorem 5.3 and Proposition 5.5). However, as we discuss the *real* genealogy in Section 5.5.3, it will emerge that the simple genealogy has some flaws. Firstly, it cannot detect the differences that plainly exist between the true genealogies of the models (highlighted in Figure 5.6, and Theorems 5.10 and 5.11). Secondly, it cannot detect some of the coalescence events that take place, which we shall call *silent events*. The finer-detailed real genealogy, on the other hand, can discern all of this. Such observations should be of interest to population geneticists because it gives a clear example of how algebraic duality may not fully capture the genealogy of the process. Our work here is similar in spirit to that of Slade [2000] who studies the typed, real genealogy of the Wright–

Fisher diffusion with genic selection. His work, however, is set in the context of an importance sampling regime, rather than the more efficient conditioned genealogies paradigm ((3) in Chapter 1) that we utilize here.

Finally, in Section 5.6, we briefly study a model which interpolates between the two major models of this chapter. The main property of interest is that, unlike the other two, it suppresses the silent events, which means that even the simple genealogy is representative of the true genealogical structure of the sample.

We remark that the neutral case simply follows as a corollary of the results in this chapter (by setting the parameter r, which encodes selection, to equal one – see Sections 5.1 and 5.2). By switching selection off, many complications that it brings vanish – silent events are no longer present, and the genealogies of the three models coincide, as do the real and a simple versions of the genealogies. In addition, a simulation study would be far easier to conduct, since the moments of the stationary distribution are easier to obtain under neutrality. With this in mind, one could perform a comparison of the genealogies of the Λ-Fleming–Viot with weak mutation (this chapter with $r = 1$), and strong mutation (Chapter 4).

5.1 Differential Killing

The first of our two selective models differentiates between the two genetic types during the extinction phase of reproduction. To accommodate this, we now think of the driving measure, $\boldsymbol{\nu}$, as providing a *pair* of impact parameters during a (u_1, u_2)-event – the first will be used to kill type-a_1 individuals, the second for those of type a_2. Let $(X_t)_{t\geq 0}$ be the Λ-Fleming–Viot process with differential killing (ΛFVDK) which tracks the proportion of type-a_1 individuals. The generator for this process, acting on bounded, measurable functions on $[0, 1]$, is given by

$$\mathscr{L}_{\Lambda DK} f(x) = \int_{(0,1]^2} \left[x f((1-u_1)x + \bar{u}(x)) + (1-x) f((1-u_1)x) - f(x) \right] \boldsymbol{\nu}(du_1, du_2) \tag{5.1}$$

$$= \int_{(0,1]^2} \left[x f((1-u_2)x + u_2) + (1-x) f((1-u_1)x) - f(x) \right] \boldsymbol{\nu}(du_1, du_2) \tag{5.2}$$

where $\bar{u}(x) := xu_1 + (1-x)u_2$ is the fraction of individuals killed during a (u_1, u_2)-event. To ensure that this process exists, we will enforce the condition

$$\int_{(0,1]^2} (u_2 - u_1)\boldsymbol{\nu}(du_1, du_2) < \infty \tag{5.3}$$

which arises when one takes a Taylor expansion of smooth functions in the generator, in the same way as in Section 4.1.1. As mentioned previously, this model has a natural interpretation: during a reproduction event, the type-a_i mass is reduced by a fraction u_i (differently for each i); the killed proportion $\bar{u}$ is replaced by the offspring of the parent. This parent, as always, is selected 'uniformly at random', which in this high-density limit means that it is of type a_1 with probability x (and type a_2 with probability $1-x$), when the process is in state x.

For the remainder of this exposition, we shall predominantly think about a specialized version of the ΛFVDK process, which we denote by $(X_t^r)_{t\geq 0}$, as it allows a rather simple graphical interpretation. To achieve this, for a fixed parameter $r \in (0,1]$, set $\boldsymbol{\nu}(du_1, du_2) = \nu(du_1)\delta_{ru_1}(du_2)$. Equation (5.2) becomes

$$\mathscr{L}^r_{\Lambda D*} f(x) := \int_0^1 \left[xf((1-ru)x + ru) + (1-x)f((1-u)x) - f(x)\right] \nu(du), \tag{5.4}$$

where ν is a σ-finite measure on $(0,1]$ satisfying $\int_0^1 u\nu(du) < \infty$, and we have removed the subscript of u_1 for clarity. This existence condition was also necessary for the mutation model of Chapter 4.

As for the notation, we have changed the subscript K to a $*$ to emphasize the fact that this model can be interpreted *either* as differential killing, *or* as differential birth (so the $*$ is a placeholder for either K or B), see Section 5.2 for details. The superscript r denotes the fact we have fixed a parameter. One could alternatively rewrite this parameter as $s = 1 - r \in [0,1)$, which really measures the selective difference in this differential killing model. Notice that when $r = 1$, or $s = 0$, we recover the usual ΛFV process, as this represents the case when selection is absent.

We can interpret the specialized model in the following manner. During u-events, which are driven by the measure ν, the parent is chosen 'uniformly at random', which means that it is of type a_1 with probability x, where x is the current frequency of a_1-alleles. Then, independently for each individual

(of which there are infinitely many), excluding the parent (who survives), throw a three-sided coin, for which the following outcomes are possible:

- with probability ru, it dies irrespective of its type;

$\circ$ with probability $(1-r)u$, it dies if it is of type a_1, but survives if of type a_2;

$|$ with probability $1-u$, it survives irrespective of its type.

Finally, the offspring replace the dead, and receive the allelic type of the parent.

The symbols by which the outcomes are labelled are related to Figure 5.1. If the first event occurs, we shall refer to it as a *strong* circle ($\bullet$); if the second occurs, we shall call it a *weak* circle ($\circ$); if neither of these happen, then there is *no* circle ($|$). In this selective model, the type a_2 is stronger than a_1, as an individual of this type can only be killed by strong circles, whereas type-a_1 individuals are killed by both circles.

We can now see why this interpretation yields (5.4), the generator of the process tracking the proportion of a_1-individuals. Firstly, suppose that, during a u-event, the parent is of type a_2, which happens with probability $(1-x)$. Then, some of the a_1-mass is killed, and in particular both strong *and* weak circles will kill these individuals, leaving only the thinned fraction $(1-u)x$ after the extinction phase. In addition, since the offspring will be of type a_2, there is no re-population of this type. This explains the second summand under the integral.

Now suppose that the parent is of type a_1, which happens with probability x. Once again, after the killing stage, the mass of type a_1 will be thinned down to $(1-u)x$, but there is also re-population during the recolonization phase. The contribution from the killed a_1-mass is $ux = x - (1-u)x$. The contribution from the killed a_2-mass is $ru(1-x)$ as the proportion before the event was $(1-x)$ and this was thinned by a factor ru through strong circles only (weak circles had no effect). Hence, after this event, the proportion of type a_1 is equal to

$$(1-u)x + ux + ru(1-x) = (1-ru)x + ru,$$

which explains the first term under the integral. (The final term in the generator, $-f(x)$, as before, is a result of the jump nature of this process.)

A graphical representation of the model can be found in Figure 5.1 (a). In this picture, we focus on only a small segment of the process. We imagine that the individuals of the population reside on lines, and time for the process runs down the page. When there is a u-event, the parent is connected to the circles by arrows. With knowledge of the genetic types, it is possible to say which circles caused a death and which did not. The arrow is unsuccessful (or dashed in the figure) if the circle did not kill the individual on that line. However, if it was successful, an offspring of the parent will be laid there. For example, in the first u-event that we see in the diagram, the parent has three successful offspring in the visible segment of the population, but the fourth was unsuccessful as the type-a_2 on that line was robust to the weak circle. Note that the lines may affect or be affected by other lines that are not shown in the figure. There are also mutations in this picture, which will be formally introduced in Section 5.4.

5.2 Differential Birth

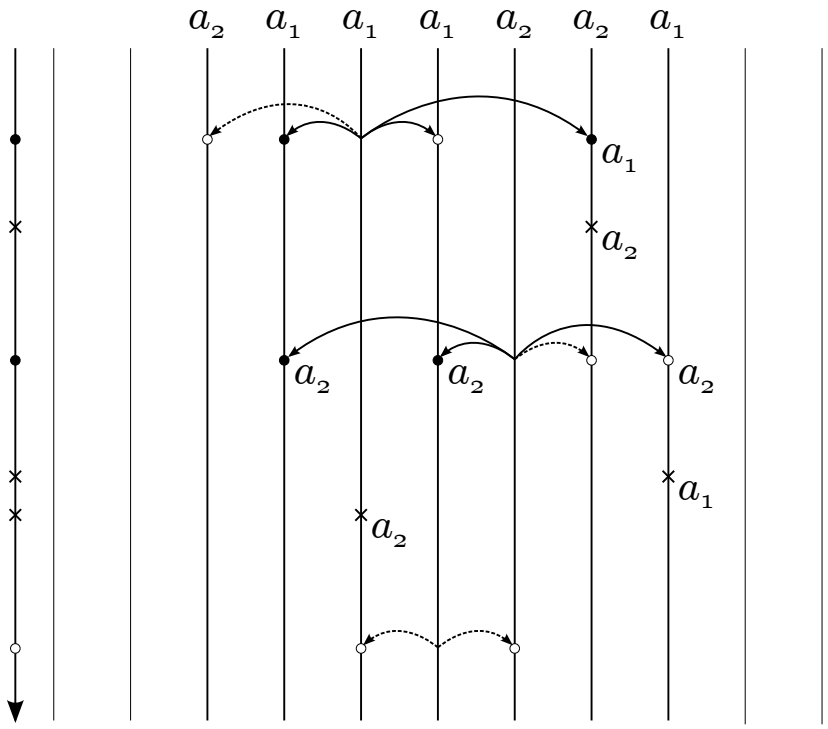
The second selective model, the Λ-Fleming–Viot process with differential birth (ΛFVDB), differentiates between the two types by providing different numbers of offspring depending on the type of the parent. The reproduction events are driven by the measure $\tilde{\boldsymbol{\nu}}$, and its generator is given by

$$\mathscr{L}_{\Lambda DB} f(x) = \int_{(0,1]^2} \left[x f((1-u_1)x + u_1) + (1-x) f((1-u_2)x) - f(x) \right] \tilde{\boldsymbol{\nu}}(du_1, du_2) \tag{5.5}$$

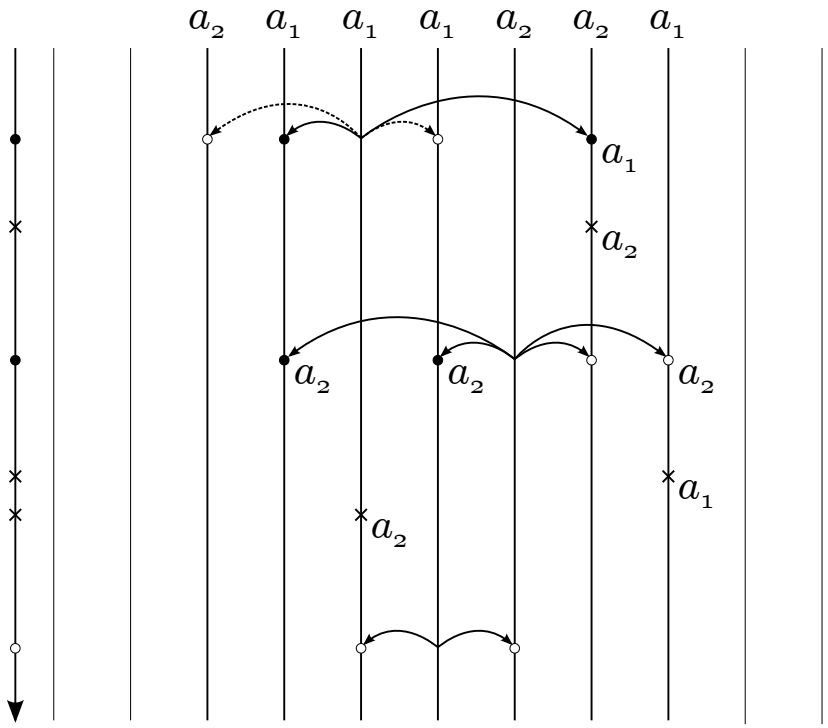
$$= \int_{(0,1]^2} \left[x f((1-u_1)x + u_1) + (1-x) f((1-u_2)x) - f(x) \right] \boldsymbol{\nu}(du_2, du_1), \tag{5.6}$$

where $\tilde{\boldsymbol{\nu}}$ is a measure satisfying (5.3). Clearly, (5.2) and (5.6) are closely related as one can be reached from the other simply by swapping the labels around. This already hints at the close connection between these models, despite the conceptual difference in reproduction mechanisms.

Intuitively, this model works as follows. During a (u_1, u_2)-event, the parent is chosen uniformly at random (as usual), but its type determines which of the impact parameters will be used throughout the event – if the parent is of type a_i, then u_i is used and the other is simply discarded.



(a)



(b)

Figure 5.1: Graphical constructions of the Λ-Fleming–Viot processes with (a) differential killing and (b) differential birth, and weak mutation. Time runs forward down the page, as indicated by the arrow on the left, which also marks when events occur. Reproduction events are indicated by multi-arrows emanating from the parent of the event (solid arrows are successful, dashed arrows are not). Strong circles are indicated by $\bullet$, weak circles by $\circ$ and mutations by $\times$. The types of the individuals are labelled at the top, and subsequently when there is a change of type on a line as a consequence of an event. Notice that the labels are exactly the same in both pictures, despite differences in reproductive success

Once again, we would like to consider a specialized model by setting $\tilde{\boldsymbol{\nu}}(du_1, du_2) = \delta_{ru_2}(du_1)\nu(du_2)$. Upon doing so, (5.6) becomes (5.4), the generator defined as $\mathscr{L}^r_{\Lambda D*}$. This specialization also has an interpretation in terms of weak and strong circles, but it is slightly different to that of differential killing. During a u-event, the parent is again chosen uniformly at random and survives the event. A three-sided coin is thrown independently for each member of the population, excluding the parent, to determine which of the following actions takes place.

- With probability ru, that individual dies irrespective of the *parent's* type;
- with probability $(1-r)u$, it dies if the *parent* is of type a_2, but survives otherwise;
- with probability $1-u$, it survives irrespective of the parent's type.

Finally, as before, the offspring replace the killed individuals and receive the genetic type of their parent. This means that parents of type a_1 can only lay their offspring on strong circles, whilst parents of type a_2 can lay their offspring on both weak *and* strong circles, making them selectively advantageous.

To see that this interpretation does indeed coincide with (5.4), let us evaluate the various scenarios. If the parent is of type a_1 (with probability x), only the strong circles are successful. Then, the fraction of type a_1 lost to strong circles is both lost (through extinction) and regained (through recolonization, since the parent is of type a_1). In addition, the fraction $ru(1-x)$ of type a_2 lost to strong circles is also gained by the a_1-population, which means that the transition

$$x \mapsto x + ru(1-x) = (1-ru)x + ru$$

occurs in this case. This explains the first summand. For the second, suppose that the parent is of type a_2 (with probability $1-x$). Then, from the point of view of the a_1-proportion, we lose the fraction $x((1-r)u + ru) = xu$ through both strong and weak circles (since the parent is of type a_2). Since the offspring are not of this type, the transition

$$x \mapsto x - xu = (1-u)x$$

takes place. The final, negative term under the integral of (5.4) is simply a consequence of the jump nature of this process.

The interpretation of the specialized ΛFVDB model is demonstrated in Figure 5.1 (b). For example, we see that in the final two reproductive events, the parent is of type a_2, and can lay its offspring on every circle. On the other hand, the first event has a type-a_1 parent, and she fails to lay offspring on the weak circles.

5.3 Comparison of the Two Mechanisms

The specialized Λ-Fleming–Viot models with differential killing (Section 5.1) and differential birth (Section 5.2) clearly have distinct reproduction mechanisms and yet, as previously remarked upon, the proportion of type-a_1 individuals can be described by the same generator, $\mathscr{L}^r_{\Lambda D*}$ of (5.4). One can also observe by comparing parts (a) and (b) of Figure 5.1 that, despite the different mechanisms, the label processes along each line are consistent across both models (when appropriately coupled). One of the purposes of this section is to formally prove that this is the case, which would in turn explain why the process X^r_t is described by the same generator in each model. In addition, we shall also point out some crucial differences.

Table 5.1 outlines all the possibilities that could happen to an individual line when affected by a u-event; that is, when it is struck by a weak or strong circle. We have to consider both the type of the parent (parental type), who is trying to displace the individual on the line, and the type of the individual being targeted (target type). The type of the circle (symbol) is also important in determining the outcome, which could potentially be different depending on the model. The final two (split-)columns in the table indicate whether the parent was successful (if so, a ✓ is placed in the displaced column) and what the type of the individual on the line is (resultant type), irrespective of whether there is a new offspring there. The rows iterate this over all eight combinations that could occur in a reproductive event. For the sake of completeness, we include the effects of mutation at the bottom of the table – this will be introduced properly in Section 5.4.

			DK	DB	DK	DB
Parental type	Target type	Symbol	Displaced?		Resultant type	
a_1	a_1	$\circ$	✓	×	a_1	a_1
a_1	a_1	$\bullet$	✓	✓	a_1	a_1
a_1	a_2	$\circ$	×	×	a_2	a_2
a_1	a_2	$\bullet$	✓	✓	a_1	a_1
a_2	a_1	$\circ$	✓	✓	a_2	a_2
a_2	a_1	$\bullet$	✓	✓	a_2	a_2
a_2	a_2	$\circ$	×	✓	a_2	a_2
a_2	a_2	$\bullet$	✓	✓	a_2	a_2
N/A	a_1	×	N/A		a_2	a_2
N/A	a_2	×	N/A		a_1	a_1

Table 5.1: The possible events that may cause a label change in the ΛFVD$*$m models. The type of the targeted individual line after the event is listed (resultant type), along with whether a new offspring was born as a result of the event (displaced?)

That the two models agree on the resultant type for every possible eventuality (reproduction and mutation) proves that the label process along each line is consistent, and explains why the generators $\mathscr{L}^r_{\Lambda DK}$ and $\mathscr{L}^r_{\Lambda DB}$ are in fact the same. However, despite this consistency, it does not change the fact that, at some level, the processes differ at weak circles: this can be seen in the displaced column at rows one and seven. This will have interesting consequences for the genealogical processes, which we shall explore in more detail in Section 5.5.

We still have yet to introduce mutation in order to induce a stationary distribution for the process. It will turn out that this mechanism will not disrupt the observations that we have made in this section – namely that the label processes along the individual lines do not differ between the differential birth and killing models (with the consequence that X^r_t has the same generator in both cases), whilst the placement of offspring *does* differ.

5.4 Adding Mutation

In order to study the genealogies of these models in the paradigm of Section 2.2.1, we would like to introduce a mutation mechanism that will induce a stationary measure for the process. In addition, we would like this mechanism to preserve the consistency between the two models that was discussed at the end of the previous section. For these reasons, we shall model with weak mutation (introduced in Section 2.3.3 for the Wright–Fisher diffusion) as opposed to the strong mutation mechanism of Chapter 4. In the remainder of this section, we shall explain why we have made this choice.

The generator of the Λ-Fleming–Viot process with differential birth/killing and weak mutation is given by the operator

$$\begin{aligned}\mathscr{L}^r_{\Lambda D*m} f(x) &= \int_0^1 \left[xf((1-ru)x+ru) + (1-x)f((1-u)x) - f(x)\right]\nu(du) \\ &\quad + (\nu_2(1-x) - \nu_1 x)\,\frac{df}{dx} \\ &= \mathscr{L}^r_{\Lambda D*} f(x) + \mathscr{L}_m f(x).\end{aligned} \tag{5.7}$$

This ensures that mutation acts as a Poisson point process along each of the lines, independently of reproduction. When a mutation arises in an individual, it will simply change its genetic type to the other one. In particular, mutation behaves the same way in both the differential birth and differential killing models. From this and the comments in Section 5.3, we can deduce that it is possible to couple the processes so that the labels along the lines are *a.s.* the same no matter which model interpretation we choose to consider. The mutations are labelled by a $\times$ in Figure 5.1, and their effects on the label processes of the individual lines are given at the bottom of Table 5.1.

We shall also take this opportunity to explain why the inclusion of the strong mutation mechanism of Chapter 4 would not lead to the same conclusions. Of course, one can define them: let $v_i^K, v_i^B \in (0,1)$, $i \in \{1,2\}$, be fixed mutation parameters. Using the form of (5.1) as our starting point, we may define the Λ-Fleming–Viot process with differential killing and strong mutation through the generator

$$\mathscr{L}_{\Lambda DKM} f(x) = \int_{(0,1]^2} [xf((1-u_1)x + (1-v_1^K)\bar{u}(x)) + (1-x)f((1-u_1)x + v_2^K \bar{u}(x)) - f(x)]\boldsymbol{\nu}(du_1, du_2). \tag{5.8}$$

This has a similar interpretation to before, except that now a fraction v_i^K of the offspring $\bar{u}(x)$ mutate to the other type. Similarly, the Λ-Fleming–Viot process with differential birth and strong mutation can be identified through the generator

$$\mathscr{L}_{\Lambda DBM} f(x) = \int_{(0,1]^2} [xf((1-u_2)x + (1-v_1^B)u_2) + (1-x)f((1-u_1)x + v_2^B u_1) - f(x)]\boldsymbol{\nu}(du_1, du_2). \tag{5.9}$$

It is now possible to see that these processes are *not* the same, since when comparing the first summands in the integrands of the generators, for example, we find that

$$\begin{aligned}(1-u_1)x + (1-v_1^K)\bar{u}(x) &= (1-u_2)x + (1-v_1^K)u_2 + (u_2-u_1)v_1^K x \\ &\neq (1-u_2)x + (1-v_1^B)u_2\end{aligned}$$

in general, even if we set $v_1^K = v_1^B$. This is because the processes will in general take on different values as soon as the first jump of this type occurs. Moreover, this observation still holds even if we specialize to the circle interpretation of Sections 5.1 and 5.2.

We have shown that these two models do not coincide unless all the mutation parameters vanish. This is in contrast to the specialized ΛFVD$*$m processes, which we will concentrate on from this point onwards.

5.5 Genealogies

We would like to understand the genealogy associated with the specialized Λ-Fleming–Viot processes with differential birth/killing and weak mutation introduced in the earlier sections of this chapter. For

the remainder, we shall refer to this process, which is characterized through the generator $\mathscr{L}^r_{\Lambda D*m}$ in (5.7), as $(X^r_t)_{t\geq 0}$.

We shall achieve this goal using a mixture of the duality arguments described in Chapter 2, and in particular Section 2.2.1, and the graphical construction (see Figure 5.1). When we consider the dual process of X^r_t derived through the generator, we shall refer to this as the *weak* genealogy – this was our strategy in Chapter 4, for example. However, we shall also characterize the genealogy obtained through the graphical construction. We will refer to this as the *strong* genealogy. The weak genealogy can be obtained from the strong version by simply removing some of the excess details, and can usually be derived algebraically, through generator arguments.

In this exposition, we shall always condition upon knowing the types in our sample. The weak mutation mechanism that was introduced in Section 5.4 induces a stationary distribution, allowing us to study the normalized duality of the process. Our general strategy is to study the reproduction and mutation mechanisms separately, which is possible since they act independently (their generators are simply added together in (5.7)). The genealogy concerning reproduction alone has an interesting characterization in and of itself (Theorem 5.3 and Proposition 5.5), but we shall later combine it with mutation to obtain a normalized duality (Corollary 5.7).

In the spirit of the ancestral selection graph of Krone and Neuhauser [1997] and Neuhauser and Krone [1997], the genealogies that we uncover will be a system of branching and coalescing lineages. However, the situation is somewhat more complex due to the multiple-merger nature of the Λ-coalescent, and the selection mechanisms involved. As a first step, we shall study the *simple* typed genealogy in Section 5.5.2, which only records the genetic types of the lineages. This is an interesting study in its own right, but in Section 5.5.3 we will realize that there is more information available. In the *real*, typed genealogy, we can track both the genetic type and the *reality* of each lineage. The latter property encodes whether the lineage carries any genetic material, i.e. if it is ancestral to the sample. If so, then it is real, if not, it is virtual. This concept will be defined more carefully in Section 5.5.3. The real genealogy will in principle allow us to infer information concerning the time to most recent common ancestor, and other basic properties concerning the shape of the *true* genealogy, which is ultimately what we would

like to understand.

Remark 5.1. We would like to stress that the simple and real genealogical processes do in fact branch as well as coalesce. This is a slight abuse of the word 'genealogy', but should emphasize that we are studying these processes in order to uncover the true genealogy.

The main result of interest in this chapter is that the simple genealogies of the differential birth and killing interpretations of X_t^r coincide, but their real genealogies do not, even when the events are coupled in the natural way. This is explained more formally through Corollary 5.7, and Theorems 5.10 and 5.11.

Despite the fact that we condition on knowing the types in the sample, one might like to draw a parallel between the classical ancestral selection graph and the simple genealogy that we study in Section 5.5.2. However, we stress that one should not do this because of the existence of silent events (Corollary 5.12). This subtlety means that the simple genealogy that we derive does not represent the true genealogy of either model, as it cannot detect all the coalescence events that take place. To be explicit, the true genealogy of the typed sample is *not* embedded within the branching-coalescing structure of the simple genealogy. The real genealogy, on the other hand, can detect all the coalescence events, making it a worthy object to study.

5.5.1 Terminology

We shall use this section to define some of the terminology that will be used frequently in the sequel. Firstly, we shall recall what we mean by strong and weak circles. During a u-event, each line on which individuals reside receives one of a strong circle ($\bullet$) with probability ru, a weak circle ($\circ$) with probability $(1-r)u$, or no circle ($|$) with the remaining probability $1-u$. No matter which model interpretation one is using, the individual will always die at the behest of a strong circle and will never die if a circle is not received. The weak circle is the driver of selection, and all the individuals with these *may* die depending on the circumstances. The exact rules concerning the weak circle do depend on the model and are described in Sections 5.1-5.3. In any case, the rules ensure that a_2 is the stronger genetic type.

Adjective	Description	Adjective	Description
Typed	Know the types of lineages	Untyped	Do not know types of lineages
Simple	Reality of lineages not detected	Real	Differentiate real and virtual lineages
Weak	Related to the generator	Strong	Related to the graphical construction

Table 5.2: Usage of the word 'genealogy'

When we sample individuals from the population at the current time, we will record their genetic types. The sample will be a vector $\boldsymbol{k} = (k_1, k_2)$ where there are k_i individuals of type a_i. We shall interchangeably refer to this as the *sample configuration*, $|\boldsymbol{k}|$*-sample* or simply *sample*. The sample configuration differs to that of the circle configuration, defined below.

Definition 5.2 (Circle configuration)**.** A $\boldsymbol{k}$*-circle configuration* is an unordered set of strong and weak circles containing at least one and at most $|\boldsymbol{k}|$ symbols. Moreover, it specifies how many of the strong circles fall onto type-a_1 and type-a_2 lineages, and similarly for weak circles. We shall refer to the set of $\boldsymbol{k}$-circle configurations as $\mathcal{C}_{\boldsymbol{k}}$.

Typically, the state of the genealogy will change during a u-event backwards in time. In order to characterize the genealogy through this event when it is in state $\boldsymbol{k}$, we shall be interested understanding each of the $\boldsymbol{k}$-circle configurations at that time. Also, during such an event, and given a circle configuration, we shall study how the circles affect each lineage. Each branching-coalescence event that takes place will involve a parent that is not in the sample – we will refer to this as the *parental lineage*. In addition, weak circles (if there are any) will create extra branches, which we refer to as *continuing lineages*. Note that *branch* is a synonym for lineage in this context. Finally, as we have defined many different kinds of genealogies, we provide a summary in Table 5.2 to clarify the meanings of each one.

5.5.2 Simple Genealogy

To find the simple genealogy of our finite typed sample $\boldsymbol{k}$ (which contains k_i lineages of type a_i), we shall first consider what happens during each possible type of u-event by considering reproduction in isolation. As a first step we shall study examples that become increasingly more complex through a mixture of graphical constructions and duality calculations. The pictures will demonstrate how the

strong simple genealogy behaves, but we will consider these alongside the weak genealogy (which will only record the numbers of each type in the sample). This strategy will mean that we can also define more necessary terminology along the way, before collecting the results in general in Theorem 5.3 and Proposition 5.5. Once we have done all this, we will investigate what role mutation has to play, and use the technique of Section 2.2.1 to construct a normalized duality (Corollary 5.7).

To understand the genealogies through graphical constructions, we need to fix an interpretation of selection, so we arbitrarily choose differential killing for the time being. Any conclusions that we draw about the *weak* genealogy for this model must be equivalent to the differential birth analogue, since they share exactly the same generator, (5.7). Moreover, since the models share the same forwards-in-time label processes along each line, we can extend statements about the *strong* simple genealogy of the ΛFVDKm model to the differential birth model too.

Suppose that, in genealogical time, the sample is currently in state $\boldsymbol{k}$. Then, given a circle configuration which affects it during a u-event, the idea is to find the set of possible sample configurations which, going forwards in time, might have led to the configuration that we see now, under the differential killing model. This will describe all the possible ways that the lineages were related to each other through such an event. We then repeat this for all circle configurations in $\mathcal{C}_{\boldsymbol{k}}$, and finally for each state $\boldsymbol{k}$. Throughout, we shall use the functional notation from Chapter 4, namely that $f_{\boldsymbol{k}}(x) := x^{k_1}(1-x)^{k_2}$ where $\boldsymbol{k} > \boldsymbol{0}$, defined in Appendix A.

To begin, consider the simple case when the sample consists of a single lineage. Recalling that we are only dealing with reproduction, not mutation, at this stage, it is simple to show that

$$\begin{aligned}
\mathscr{L}^r_{\Lambda D*} f_{(1,0)}(x) &= \int_0^1 (1-r)u\nu(du)\left(f_{(2,0)}(x) - f_{(1,0)}(x)\right), \\
\mathscr{L}^r_{\Lambda D*} f_{(0,1)}(x) &= \int_0^1 (1-r)u\nu(du)\left(f_{(0,2)}(x) - f_{(0,1)}(x)\right) \\
&\quad + \int_0^1 2(1-r)u\nu(du)\left(f_{(1,1)}(x) - f_{(0,1)}(x)\right) \\
&\quad + \int_0^1 2(1-r)u\nu(du) f_{(0,1)}(x).
\end{aligned}$$

To understand how this information translates pictorially, consult Figure 5.2. All the sample lineages

are highlighted in bold. There are infinitely many vertical lines which host individuals, but only two which concern us in this scenario. We imagine all the possible events which might result in a change to this sample configuration and, in this simple example, only a single circle hitting the lineage causes an effect.

When that circle is strong (•), there is a change in the strong genealogy as the lineage jumps to the line where its parent was located, but that information is lost when we track only the numbers of types in the sample, which has not changed as a result of this event. This is true whether the single lineage was of type a_1 or a_2, since the strong circle would have killed the parent no matter which type it was. See Figure 5.2 (a), (c).

However, the behaviour differs for a weak circle (∘) on the lineage. When it is of type a_1, we have to check two things to ensure that we obtain the current state $(1, 0)$ – firstly, the parent of the event, or the parental lineage, must have been of type a_1; secondly, we have to check that the individual living on the line inhabited by the current lineage prior to the event (forwards in time), i.e. the continuing lineage, was indeed killed by the weak circle. For this to have happened it must have been of type a_1. (It could not have been of type a_2, otherwise it would have persisted through the event, and the type of the individual on that line would be a_2, which is a contradiction.) This gives rise to a *branching event* since there is now one more lineage whose type we must keep track of. This scenario is depicted in Figure 5.2 (b). The information is recorded in the generator when acting on $f_{(1,0)}(x) = x$, since it yields a generator of a jump process which makes the transition $(1, 0) \to (2, 0)$ at rate $\int_0^1 (1 - r)u\nu(du)$, the rate at which a weak circle hits a single lineage.

The situation is more complex when our single lineage is of type a_2, and is hit by a weak circle (Figure 5.2 (d)), since there is more than one historical sample configuration compatible with this event. (We shall refer to this phenomenon as an *optional event*, in contrast to an *ordinary event*, such as those in (a), (b) and (c) of Figure 5.2. It is also a branching event.) For example, if the continuing lineage were of type a_1, then the individual would have been killed in the event and the parent would have to have been of type a_2 to give rise to our lineage. On the other hand, if the continuing lineage were of type a_2, then the weak circle would *not* have killed the individual residing on that line, and the (potential) parent

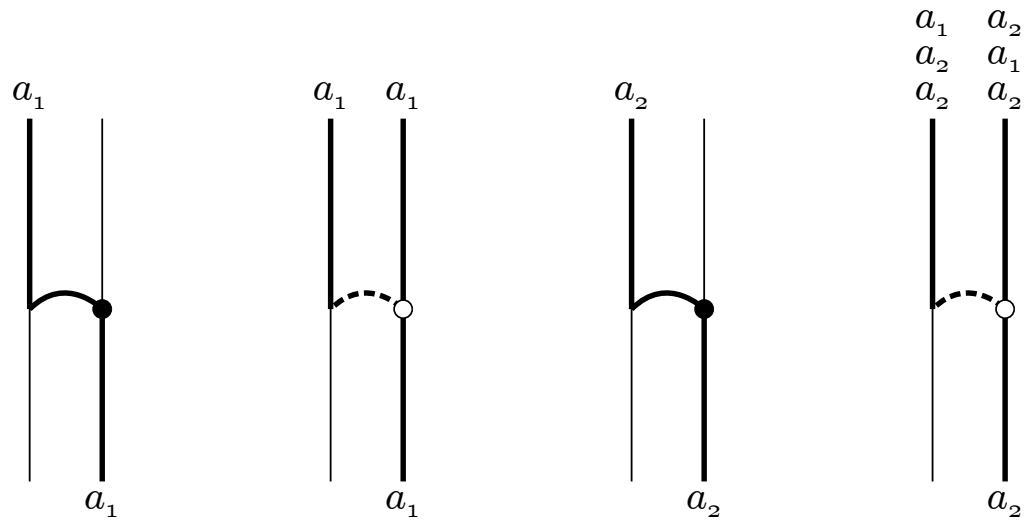


Figure 5.2: Possible transitions in the simple genealogy of a single lineage under the specialized differential killing model (without mutation). The lineage is of type a_1 in (a) and (b), and of type a_2 in (c) and (d). The lineage experiences a strong circle in (a) and (c), and a weak circle in (b) and (d). An optional event is exemplified in (d). The sampled lineages are highlighted in bold

could be of either type. This is interesting because, forwards in time, the potential parent would not have given birth at all, but it is necessary to track this lineage nonetheless.

This is recorded by the generator acting on $f_{(0,1)}(x) = 1-x$, for reasons that we shall describe now. Firstly, the factor $\int_0^1 (1-r)u\nu(du)$ records the rate at which a single weak circle strikes our lineage, as before. There is one way that the transition $(0,1) \to (0,2)$ can occur, but *two* ways that $(0,1) \to (1,1)$ can happen (the fact that these happen in different ways graphically further exemplifies the fact that the strong genealogy contains more information than the weak one). In addition, there is an extra additive term proportional to the input function $f_{(0,1)}(x)$ which handles the fact that there are optional configurations. We will demonstrate later that the prefactor of this extra term can be characterized in terms of the number of options available. The existence of such an additional term should indicate that duality in the broad sense will emerge (see Proposition 2.6 and in particular (2.10)).

Having characterized the single-lineage case, we now consider the scenario when there are two lineages in sample that are both of type a_1. Figure 5.3 summarizes all the possible outcomes of all circle configurations in this case. Notice that branching events are possible (namely when there are no strong

circles), but also genuine coalescence events (as in (c)). It can be shown that

$$\begin{aligned}\mathscr{L}^r_{\Lambda D*} f_{(2,0)}(x) = 2\int_0^1 (1-r)u\cdot(1-u)\nu(du)\left(f_{(3,0)}(x) - f_{(2,0)}(x)\right)\\ + \int_0^1 ru\cdot ru\,\nu(du)\left(f_{(1,0)}(x) - f_{(2,0)}(x)\right)\\ + \int_0^1 (1-r)u\cdot(1-r)u\,\nu(du)\left(f_{(3,0)}(x) - f_{(2,0)}(x)\right).\end{aligned}$$

The right-hand side is a valid generator which describes the possible jumps out of the state $(2,0)$. The first term corresponds to (b) in the figure, which is a branching event that introduces a third a_1 lineage. Its rate consists of the factors $(1-r)u$ and $1-u$ under the integral which respectively correspond to the probabilities of obtaining a weak circle and no circle on the lineages during a u-event. The factor of two is combinatorial and accounts for the number of ways you can lay a single weak circle on two lineages.

The second term is a coalescence event, corresponding to (c) in the figure, in which two strong circles cause the lineages to merge unambiguously into a single type-a_1 lineage. The probability of getting two strong circles during a u-event is $(ru)^2$, which is the integrand. There is only one way to lay this circle configuration onto two lineages (and so the combinatorial factor is equal to one).

The third term corresponds to (d), and is another kind of branching event that happens with a different rate, namely $\int_0^1((1-r)u)^2\nu(du)$, where $(1-r)u$ is the probability of a weak circle striking a lineage during a u-event. Neither (a) nor (e) from the figure contribute to the generator since they do not affect a change in state (in the weak genealogy).

There is no extra term on the right-hand side in this case, and correspondingly no optional events are possible. We will demonstrate later as a corollary of Theorem 5.3 that only ordinary events can occur whenever the sample consists of solely type-a_1 lineages. Intuitively, this is a result of the fact that, in order for a current lineage to be of type a_1, both the continuing and parental lineages must be of type a_1 when it is struck by a weak circle (but this is not true when a lineage is of the stronger type, a_2).

In the following calculation, we have evaluated the generator acting on the function $f_{(0,2)}(x) = (1-x)^2$, corresponding to a sample with two type-a_2 lineages. We omit the intermediate steps, as they are contained in the proof of Theorem 5.3.

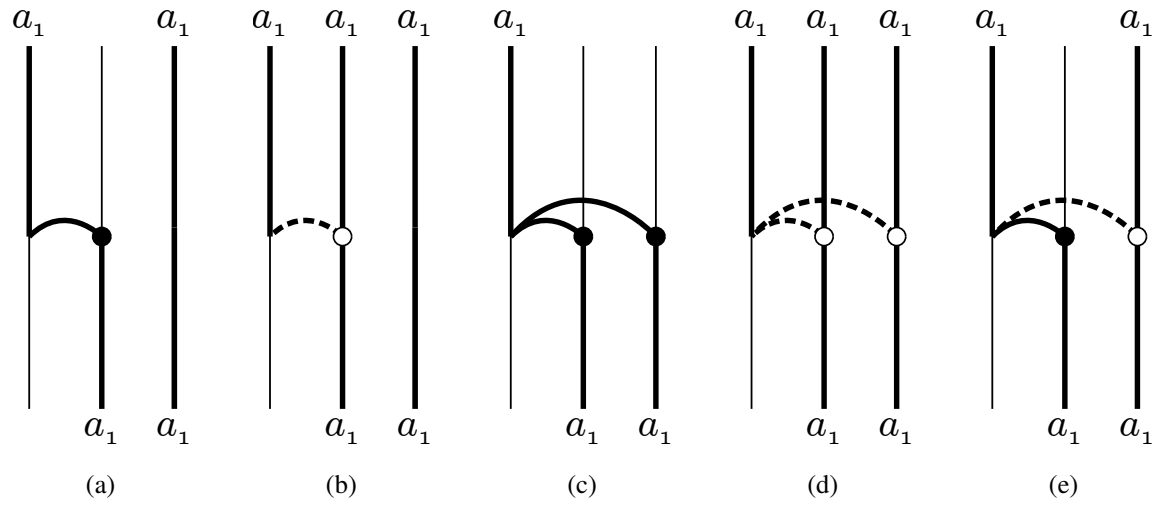


Figure 5.3: Possible transitions in the simple genealogy of a sample in state $(2, 0)$ under the specialized differential killing model (without mutation). The two lineages are struck by (a) one strong circle, (b) one weak circle, (c) two strong circles, (d) two weak circles, (e) one weak and one strong circle. There are two ways to lay the circle configuration onto the lineages in (a), (b) and (e), but only one way in (c) and (d). Events (a) and (e) cause no change to the sample (in the weak genealogy). The sampled lineages are highlighted in bold

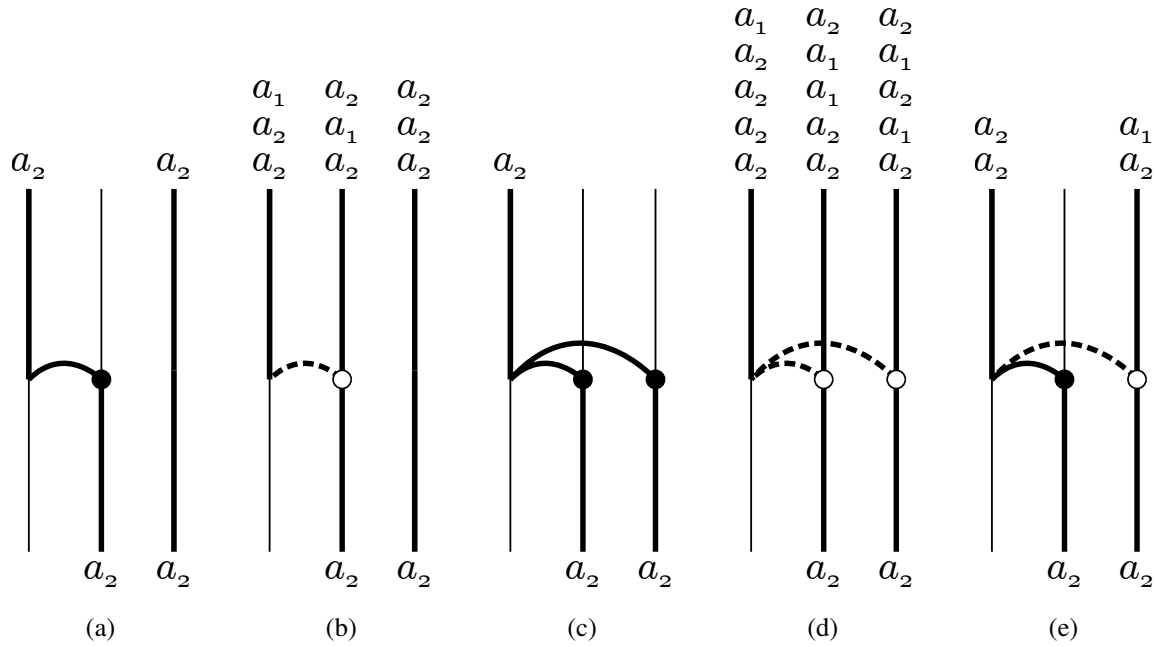


Figure 5.4: Possible transitions in the simple genealogy of a sample in state $(0, 2)$ under the specialized differential killing model (without mutation). The two lineages are struck by (a) one strong circle, (b) one weak circle, (c) two strong circles, (d) two weak circles, (e) one weak and one strong circle. (b), (d) and (e) are optional events. The sampled lineages are highlighted in bold

$$
\begin{aligned}
\mathscr{L}^r_{\Lambda D_*} f_{(0,2)}(x) = \quad & 2\int_0^1 (1-r)u\cdot(1-u)\nu(du)\left(f_{(0,3)}(x)-f_{(0,2)}(x)\right) \\
+ \quad & 2\int_0^1 2\cdot(1-r)u\cdot(1-u)\nu(du)\left(f_{(1,2)}(x)-f_{(0,2)}(x)\right) \qquad \text{(b)} \\
+ \quad & \int_0^1 ru\cdot ru\,\nu(du)\left(f_{(0,1)}(x)-f_{(0,2)}(x)\right) \qquad \text{(c)} \\
+ \quad & \int_0^1 (1-r)u\cdot(1-r)u\,\nu(du)\left(f_{(0,3)}(x)-f_{(0,2)}(x)\right) \\
+ \quad & \int_0^1 3\cdot(1-r)u\cdot(1-r)u\,\nu(du)\left(f_{(1,2)}(x)-f_{(0,2)}(x)\right) \qquad \text{(d)} \\
+ \quad & \int_0^1 (1-r)u\cdot(1-r)u\,\nu(du)\left(f_{(2,1)}(x)-f_{(0,2)}(x)\right) \\
+ \quad & 2\int_0^1 ru\cdot(1-r)u\,\nu(du)\left(f_{(1,1)}(x)-f_{(0,2)}(x)\right) \qquad \text{(e)} \\
+ \quad & \int_0^1 R_{(0,2)}(r,u)\nu(du) f_{(0,2)}(x) \qquad (5.10)
\end{aligned}
$$

Equation (5.10) is rather long, so we have labelled the parts which correspond to Figure 5.4 in order. Note that (a) from the figure is absent from the equation since one strong circle does not affect a change in state. We shall not comment on every single transition, as the ideas are consistent with the earlier ones. One marginally new concept is that the transition terms from the optional events (in (b), (d) and (e)) add up in the way one might expect. We remark that the combinatorial factors (arising from the number of ways to lay the circle configuration onto the sample) are listed to the left of integral, whilst the factors resulting from multiple options (leading to the same transition) are on the inside. A good example of this is the second term: the outer 2 is the number of ways one can lay one weak circle on two lineages, whilst the inner 2 stems from there being two possible transitions to the state $(1,2)$.

An important feature of (5.10) is that there is a remainder, the final term. It is again a multiple of the input function, and so the right-hand side may be interpreted as a duality in accordance with Proposition 2.6. We define the remainder constant to be

$$R_{(0,2)}(r,u) = 2\times 2\cdot(1-r)u\cdot(1-u) + 4\times(1-r)u\cdot(1-r)u + 1\times 2\cdot ru\cdot(1-r)u. \quad (5.11)$$

This term can be seen as a linear combination of the rates (including combinatorial factors) of all optional events, where the (positive) constant for each term is equal to the number of options, minus one. In fact,

this interpretation would extend naturally if we considered ordinary events to be optional with only one option, and they would all therefore contribute zero to the remainder constant.

Finally for $|\boldsymbol{k}| = 2$, consider a mixed sample which has the state $(1, 1)$. The corresponding picture is Figure 5.5. Heuristically, (5.12), the generator acting on $f_{(1,1)}(x)$, contains fewer transitions because it is harder for a sample of mixed type to interact during an event. Indeed, it can happen that some events are not even possible – for example, Figure 5.5 (c). In this example the two lineages may not coalesce because mutation is prohibited during birth (and it means the parent would have given birth to a child not of her type, which is a contradiction). We shall call these *incompatible events*. In (5.12), these events do not appear in the generator part of the right-hand side, but do contribute to the remainder term. For this state, the generator acting on $f_{(1,1)}(x)$ equals

$$\begin{aligned}\mathscr{L}^r_{\Lambda D*} f_{(1,1)}(x) = &\int_0^1 3(1-r)u \cdot (1-u)\nu(du)\left(f_{(2,1)}(x) - f_{(1,1)}(x)\right) \\ &+ \int_0^1 (1-r)u \cdot (1-u)\nu(du)\left(f_{(1,2)}(x) - f_{(1,1)}(x)\right) \\ &+ \int_0^1 (1-r)u \cdot (1-r)u\,\nu(du)\left(f_{(2,1)}(x) - f_{(1,1)}(x)\right) \\ &+ \int_0^1 R_{(1,1)}(r,u)\nu(du) f_{(1,1)}(x), \end{aligned} \tag{5.12}$$

where

$$R_{(1,1)}(r,u) = 2 \times (1-r)u \cdot (1-u) + (-1) \times ru \cdot ru + (-1) \times (1-r)u \cdot ru.$$

This remainder is a linear combination of the rates of *both* optional events ((b, right)) and incompatible events ((c) and (e, right)). If we again allow the combinatorial factors to be part of the rates (they are all 1 in this particular instance), then the constants corresponding to the incompatible events are equal to -1 (with the constants for the optional events as before). This is again consistent if we consider incompatible events to be optional events with *zero* options. In this case, we have shown that incompatible events do not feature in the generator part, but do in the remainder. This is true in general and will be stated formally in Proposition 5.5.

Before stating the general duality result for the reproduction mechanism given any sample $\boldsymbol{k}$, we

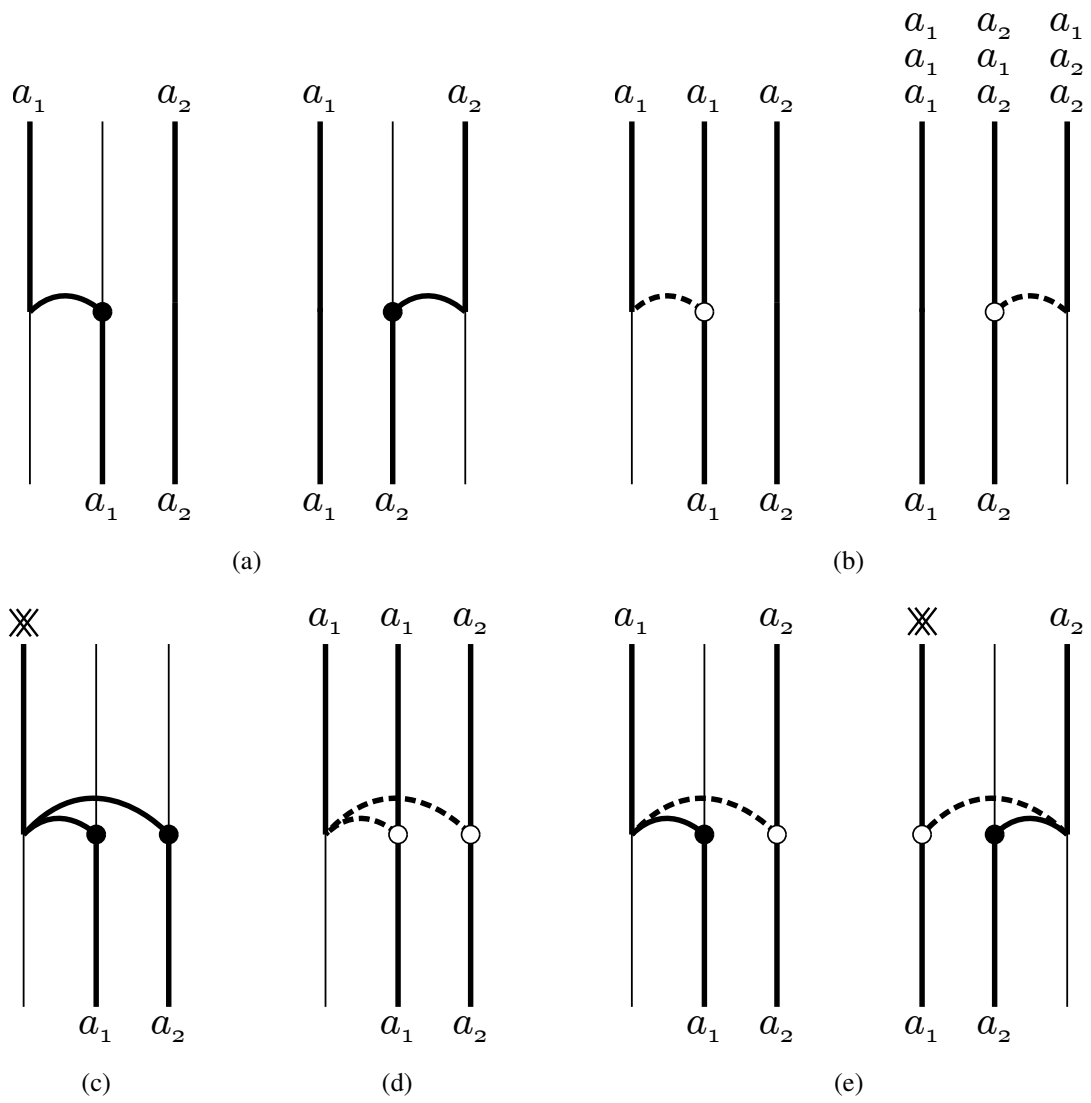


Figure 5.5: Possible transitions in the simple genealogy of a sample in state $(1, 1)$ under the specialized differential killing model (without mutation). The two lineages are struck by (a) one strong circle, (b) one weak circle, (c) two strong circles, (d) two weak circles, (e) one weak and one strong circle. (b, right) is an optional event whilst (c) and (e, right) are incompatible events. The sampled lineages are highlighted in bold

need some more notation. Define the integrals

$$I_1^\nu(s_1, w_1, w_2) := \int_0^1 (ru)^{s_1}((1-r)u)^{w_1+w_2}(1-u)^{|\boldsymbol{k}|-s_1-(w_1+w_2)}\,\nu(du) \tag{5.13}$$

$$I_2^\nu(s_2, w_2) := \int_0^1 (ru)^{s_2}((1-r)u)^{w_2}(1-u)^{|\boldsymbol{k}|-s_2-w_2}\,\nu(du), \tag{5.14}$$

which represent the rates at which compatible circle configurations strike the sample in a single permutation. The subscript refers to the allelic type of the parent of the event.

Theorem 5.3. *For $\boldsymbol{k} \in \mathbb{N}_0^2$ such that $|\boldsymbol{k}| > 0$, the generator of the specialized differential killing/birth model (without mutation) acting on $f_{\boldsymbol{k}}(x) = x^{k_1}(1-x)^{k_2}$ can be decomposed into a generator (in the $\boldsymbol{k}$ coordinate) and a remainder term as follows;*

$$\mathscr{L}^r_{\Lambda D*} f_{\boldsymbol{k}}(x) = \mathscr{L}_1 f_{\boldsymbol{k}}(x) + \mathscr{L}_2 f_{\boldsymbol{k}}(x) + \int_0^1 R_{\boldsymbol{k}}(r,u)\nu(du) f_{\boldsymbol{k}}(x) \tag{5.15}$$

where

$$\mathscr{L}_1 f_{\boldsymbol{k}}(x) := \sum_{(s_1,w_1,w_2)\in\mathcal{C}_{1,\boldsymbol{k}}} \binom{k_1}{s_1}\binom{k_1-s_1}{w_1}\binom{k_2}{w_2} I_1^\nu(s_1,w_1,w_2)\left(f_{(k_1-s_1+1,\,k_2)}(x) - f_{\boldsymbol{k}}(x)\right),$$

$$\mathscr{L}_2 f_{\boldsymbol{k}}(x) := \sum_{(s_2,w_2)\in\mathcal{C}_{2,\boldsymbol{k}}} \binom{k_2}{s_2}\binom{k_2-s_2}{w_2} I_2^\nu(s_2,w_2) \sum_{i=0}^{w_2}\binom{w_2}{i}\left(f_{(k_1+i,\,k_2-s_2-i+1)}(x) - f_{\boldsymbol{k}}(x)\right),$$

$$\mathcal{C}_{1,\boldsymbol{k}} := \{s_1, w_1, w_2 \in \mathbb{N}_0 : 0 \le s_1 + w_1 \le k_1, 0 \le w_2 \le k_2\} \setminus \{\mathbf{0}\},$$

$$\mathcal{C}_{2,\boldsymbol{k}} := \{s_2, w_2 \in \mathbb{N}_0 : 0 \le s_2 + w_2 \le k_2\} \setminus \{\mathbf{0}\}$$

and

$$R_{\boldsymbol{k}}(r,u) = (1-ru)^{k_2} + (1-u)^{k_1}(1+(1-r)u)^{k_2} - (1-u)^{|\boldsymbol{k}|} - 1. \tag{5.16}$$

Corollary 5.4. *If we condition our sample to contain no lineages of type a_2 (by setting $k_2 = 0$), then for $k_1 \ge 1$, $\mathscr{L}^r_{\Lambda D*} x^{k_1}$ forms an exact duality in the* first *coordinate of $\boldsymbol{k}$.*

Proof. Firstly, note that the set $\mathcal{C}_{2,(k_1,0)}$ is empty, so the transitions represented by the second generator cannot increase the number of type-a_2 lineages. In the first sum, there are no generator terms which

increase the second coordinate above zero, as the transitions are of the form $(k_1, 0) \to (k_1 - s_1 + 1, 0)$. The remainder becomes

$$R_{(k_1,0)}(r, u) = 1 + (1-u)^{k_1} \cdot 1 - (1-u)^{k_1+0} - 1 = 0.$$

That the remainder term vanishes means that, by Proposition 2.6, the forward process forms an exact duality with respect to the function $f_{(k_1,0)}(x)$. The second coordinate never changes in the dual process, and is therefore redundant. □

Proof of Theorem 5.3. The strategy is the following. We posit what all the possible transitions from the $\boldsymbol{k}$-sample are by reasoning from the graphical constructions in a similar manner to the above. Then we add and subtract the corresponding generator, denoted $\mathscr{L}^{\tilde{\boldsymbol{K}}} f_{\boldsymbol{k}}(x) := (\mathscr{L}_1 + \mathscr{L}_2) f_{\boldsymbol{k}}(x)$, from the right-hand side of $\mathscr{L}^r_{\Lambda D*} f_{\boldsymbol{k}}(x)$, and show that $(\mathscr{L}^r_{\Lambda D*} - \mathscr{L}^{\tilde{\boldsymbol{K}}}) f_{\boldsymbol{k}}(x)$ simplifies to give the remainder term in (5.15).

We shall use s_i to index sums that count how many strong circles fall on lineages of type a_i, and similarly w_i for weak circles falling on type-a_i lineages. It must be the case that $0 \leq s_i + w_i \leq k_i$, $i \in \{1, 2\}$. We will have to consider each circle configuration in $\mathcal{C}_{\boldsymbol{k}}$ to deduce all the possible transitions of the genealogy from state $\boldsymbol{k}$.

In order to find the transitions, we need to break the set of circle configurations down into smaller cases, and for this we propose the following partition: $\{s_1 > 0, s_2 > 0\}$, $\{s_1 > 0, s_2 = 0\}$, $\{s_1 = 0, s_2 > 0\}$ and $\{s_1 = 0, s_2 = 0\}$, where the upper bounds have been omitted. In this way, we shall count over strong circles first, and weak circles second. The easiest case is the former: there are *no* possible transitions when both s_1 and s_2 are strictly positive. That is, each circle configuration in this part of the partition generates an incompatible event. This is the case since we would require the parent of the event to be of type a_1 and a_2 simultaneously, which is clearly a contradiction. Said another way, going forwards in time, the parent would have had to have given birth to at least one child of each type, which is impossible as mutation during reproduction is not allowed in this model. One instance of this situation is depicted in Figure 5.5 (c).

Secondly, we shall consider the next simplest case: $\{s_1 > 0, s_2 = 0\}$. We observe that the one or

more strong circles on type-a_1 lineages forces the parent to be of type a_1. A more important observation to make is that *all* other (continuing) lineages with either a weak or no circle on them must remain the same type after the event, but for a variety of reasons. If a lineage has no circle on it, be it type a_1 or a_2, then it must remain the same since nothing could have happened to it during the event. If a type-a_2 lineage has a weak circle on it, then it must continue to be of this type after the event – if the individual on that line going forwards in time were a_1, then it would have been killed by the weak circle, and would be replaced by a type a_1 individual as a child of the parent, which would contradict the fact that it is actually an a_2 lineage. If a type-a_1 lineage has a weak circle on it, then it must also continue to be that type after the event – an a_2 individual sitting on that line forwards in time would not be killed by the weak circle, and so would continue to live, causing the sampled lineage to be of type a_2, which is a contradiction.

Conditioning on there being s_1 strong circles on a_1-lineages and w_i weak circles on a_i-lineages in the circle configuration, the genealogy can only make the transition $\boldsymbol{k} \to (k_1 - s_1 + 1, k_2)$ (the s_1 lineages merge into one), so these are ordinary, not optional, events. Notice that when $s_1 = 1$, there is no change in state (as far as the generator is concerned), but it is legitimate to leave that term present as the corresponding difference in the generator will cancel to zero.

It only remains to work out the rates of these transitions. There are $\binom{k_1}{s_1}$ ways to lay s_1 strong circles on k_1 type a_1 lineages (for $1 \leq s_1 \leq k_1$) and $\binom{k_1-s_1}{w_1}$ ways to lay the remaining w_1 weak circles (for $0 \leq w_1 \leq k_1 - s_1$). Similarly, there are $\binom{k_2}{w_2}$ ways to lay w_2 weak circles on k_2 type-a_2 lineages (there are no strong circles on a_2-lineages in this case). The rate of getting a single permutation of this circle configuration is $I_1^\nu(s_1, w_1, w_2)$, defined in (5.13). This is because, during a u-event, the probability of s_1 type-a_1 lineages being struck by strong circles is $(ru)^{s_1}$; similarly, the probability of w_1 (respectively w_2) type-a_1 (resp. type-a_2) lineages being struck by weak circles is $((1-r)u)^{w_1}$ (resp. $((1-r)u)^{w_2}$), and the probability of the remaining lineages receiving no circle is $(1-u)^{|\boldsymbol{k}|-s_1-(w_1+w_2)}$. Hence, the contribution to the expected generator from this part of the partition is

$$\sum_{s_1=1}^{k_1} \sum_{w_1=0}^{k_1-s_1} \sum_{w_2=0}^{k_2} \binom{k_1}{s_1} \binom{k_1-s_1}{w_1} \binom{k_2}{w_2} I_1^\nu(s_1, w_1, w_2) \left(f_{(k_1-s_1+1,\, k_2)}(x) - f_{\boldsymbol{k}}(x) \right). \tag{5.17}$$

Next, we shall consider the case $\{s_1 = 0, s_2 > 0\}$. That there is a least one strong circle on a type-a_2 lineage fixes the parent to be of type a_2. Given this, there cannot be any weak circles on a_1 lineages, as this leads to an incompatibility – if the continuing lineage were of type a_2, then it would not have been killed by the weak circle, and the lineage would be of type a_2; if it were of type a_1, then it would have been killed by the weak circle, but the a_2-parent would have given birth to an a_2-child; both scenarios lead to a contradiction. Hence, we deduce that $w_1 = 0$ in order to avoid incompatibilities. This scenario is exemplified in Figure 5.5 (e, right).

So, for this part, $s_1 = w_1 = 0$, but the other indices are free to vary provided $1 \leq s_2 + w_2 \leq k_2$. Those lineages with strong circles on them will merge together, but the ones with weak circles, if there are any, will continue. Whilst the parental lineage must be of type a_2, each of the continuing lineages may be of either type, causing optional events. Supposing that there are i, $0 \leq i \leq w_2$, continuing lineages that become type a_1 (there are $\binom{w_2}{i}$ ways that this could happen), then the sample will experience the transition $\boldsymbol{k} \to (k_1 + i, k_2 - s_2 - i + 1)$. In total, there are $\sum_{i=0}^{w_2} \binom{w_2}{i} = 2^{w_2}$ options. Overall, the contribution from this part of the partition is

$$\sum_{s_2=1}^{k_2} \sum_{w_2=0}^{k_2-s_2} \binom{k_2}{s_2} \binom{k_2 - s_2}{w_2} I_2^{\nu}(s_2, w_2) \sum_{i=0}^{w_2} \binom{w_2}{i} \left(f_{(k_1+i,\, k_2-s_2-i+1)}(x) - f_{\boldsymbol{k}}(x) \right), \tag{5.18}$$

where I_2^{ν} is defined in (5.14).

Finally, we shall consider the case when there are no strong circles, i.e. when $\{s_1 = 0, s_2 = 0\}$. It is helpful to further partition this into $\{w_1 > 0\}$ and $\{w_1 = 0\}$. In the former case, a similar logic to that of $\{s_1 > 0, s_2 = 0\}$ reasons that all the continuing lineages generated by the weak circles must remain the same type, and the parent must have been of type a_1 (to have given birth to the a_1-lineages which were affected by the weak circles). In this scenario, there is only one option, which is a branching event – see Figure 5.5 (d), for example.

In the other part of the sub-partition, $\{w_1 = 0\}$, there are several options for each circle configuration. If the parent is of type a_2, then similarly to the part $\{s_1 = 0, s_2 > 0\}$, each of the continuing lineages may be of either type. However, now there are no strong circles to fix the type of the parent, and it turns out that the parent of the event may have been of type a_1, provided each of the continuing

lineages remain the same type. This adds one to the number of options, which totals $2^{w_2}+1$ in this case. This scenario is exemplified in Figure 5.5 (b, right). Overall the contribution to the generator from this part of the partition is

$$\begin{aligned}\sum_{w_1=0}^{k_1}\sum_{w_2=0}^{k_2}\binom{k_1}{w_1}\binom{k_2}{w_2}I_1^{\nu}(0,w_1,w_2)\Bigg(&\mathbb{1}_{\{w_1>0\}}\left(f_{\boldsymbol{k}+\boldsymbol{e}_1}(x)-f_{\boldsymbol{k}}(x)\right)\\&+\mathbb{1}_{\{w_1=0\}}\Bigg(f_{\boldsymbol{k}+\boldsymbol{e}_1}(x)-f_{\boldsymbol{k}}(x)+\sum_{i=0}^{w_2}\binom{w_2}{i}\left(f_{(k_1+i,\,k_2-i+1)}(x)-f_{\boldsymbol{k}}(x)\right)\Bigg)\Bigg)\\&+\int_0^1(1-u)^{|\boldsymbol{k}|}\nu(du)f_{\boldsymbol{k}}(x),\end{aligned}\tag{5.19}$$

where the last term accounts for the inclusion of the 'false' event (i.e. $\{w_1=w_2=0\}$) in the sum, in which nothing happens. It is not too difficult to see that the sum of the contributions from each part of the partition (that is, the sum of (5.17), (5.18) and (5.19)) simplifies to $\mathscr{L}^r_{\Lambda D*}f_{\boldsymbol{k}}(x)-\int_0^1 R_{\boldsymbol{k}}(r,u)\nu(du)$, which is the the generator part of the right-hand side of (5.15), $(\mathscr{L}_1+\mathscr{L}_2)f_{\boldsymbol{k}}(x)=\mathscr{L}^{\tilde{\boldsymbol{K}}}f_{\boldsymbol{k}}(x)$. Then, it just remains to show that $\mathscr{L}^r_{\Lambda D*}f_{\boldsymbol{k}}(x)-\mathscr{L}^{\tilde{\boldsymbol{K}}}f_{\boldsymbol{k}}(x)=\int_0^1 R_{\boldsymbol{k}}(r,u)\nu(du)$ in order to complete the proof.

In the following calculation, we shall simplify $\mathscr{L}_1$, but *suppressing the integral*, mainly through expressing the sums as binomial series. We also add in the zero vector term (that is, we enhance $\mathcal{C}_{1,\boldsymbol{k}}$ by including $\mathbf{0}$), since it helps with the simplification.

$$\begin{aligned}&\mathscr{L}_1 f_{\boldsymbol{k}}(x)+(1-u)^{|\boldsymbol{k}|}\left(f_{\boldsymbol{k}+\boldsymbol{e}_1}(x)-f_{\boldsymbol{k}}(x)\right)\\&=\sum_{s_1=0}^{k_1}\sum_{w_1=0}^{k_1-s_1}\sum_{w_2=0}^{k_2}\binom{k_1}{s_1}\binom{k_1-s_1}{w_1}\binom{k_2}{w_2}(ru)^{s_1}((1-r)u)^{w_1+w_2}(1-u)^{|\boldsymbol{k}|-s_1-(w_1+w_2)}\\&\qquad\times\left(f_{(k_1-s_1+1,\,k_2)}(x)-f_{\boldsymbol{k}}(x)\right)\\&=\sum_{s_1=0}^{k_1}\binom{k_1}{s_1}(ru)^{s_1}\left(f_{(k_1-s_1+1,\,k_2)}(x)-f_{\boldsymbol{k}}(x)\right)\sum_{w_1=0}^{k_1-s_1}\binom{k_1-s_1}{w_1}((1-r)u)^{w_1}(1-u)^{k_1-s_1-w_1}\\&\qquad\times\sum_{w_2=0}^{k_2}\binom{k_2}{w_2}((1-r)u)^{w_2}(1-u)^{k_2-w_2}\\&=\sum_{s_1=0}^{k_1}\binom{k_1}{s_1}(ru)^{s_1}(1-ru)^{k_1-s_1}\left(f_{(k_1-s_1+1,\,k_2)}(x)-f_{\boldsymbol{k}}(x)\right)\times(1-ru)^{k_2}\end{aligned}$$

$$= (1-ru)^{k_2}(1-x)^{k_2}\left(x\sum_{s_1=0}^{k_1}\binom{k_1}{s_1}(ru)^{s_1}(1-ru)^{k_1-s_1}x^{k_1-s_1} - x^{k_1}\cdot 1^{k_1}\right)$$
$$= (1-ru)^{k_2}(1-x)^{k_2}\left(x(ru+(1-ru)x)^{k_1} - x^{k_1}\right).$$

The second summation, i.e. $\mathscr{L}_2$, again with the extra zero term and the integral suppressed, simplifies as follows:

$$\begin{aligned}
&\mathscr{L}_2 f_{\boldsymbol{k}}(x) + (1-u)^{|\boldsymbol{k}|}\left(f_{\boldsymbol{k}+\boldsymbol{e}_2}(x) - f_{\boldsymbol{k}}(x)\right)\\
&= \sum_{s_2=0}^{k_2}\sum_{w_2=0}^{k_2-s_2}\binom{k_2}{s_2}\binom{k_2-s_2}{w_2}(ru)^{s_2}((1-r)u)^{w_2}(1-u)^{|\boldsymbol{k}|-s_2-w_2}\sum_{i=0}^{w_2}\binom{w_2}{i}\\
&\qquad\times\left(f_{(k_1+i,\,k_2-s_2-i+1)}(x) - f_{\boldsymbol{k}}(x)\right),\\
&= (1-u)^{k_1}x^{k_1}\sum_{s_2=0}^{k_2}\binom{k_2}{s_2}(ru)^{s_2}\sum_{w_2=0}^{k_2-s_2}\binom{k_2-s_2}{w_2}((1-r)u)^{w_2}(1-u)^{k_2-s_2-w_2}\\
&\qquad\times\left((1-x)^{k_2-s_2+1}\sum_{i=0}^{w_2}\binom{w_2}{i}\left(\frac{x}{1-x}\right)^i - (1-x)^{k_2}\sum_{i=0}^{w_2}\binom{w_2}{i}\right)\\
&= (1-u)^{k_1}x^{k_1}\sum_{s_2=0}^{k_2}\binom{k_2}{s_2}(ru)^{s_2}\sum_{w_2=0}^{k_2-s_2}\binom{k_2-s_2}{w_2}((1-r)u)^{w_2}(1-u)^{k_2-s_2-w_2}\\
&\qquad\times\left((1-x)\cdot(1-x)^{k_2-s_2-w_2} - (1-x)^{k_2}\cdot 2^{w_2}\right)\\
&= (1-u)^{k_1}x^{k_1}\sum_{s_2=0}^{k_2}\binom{k_2}{s_2}(ru)^{s_2}\\
&\qquad\times\left((1-x)((1-r)u+(1-x)(1-u))^{k_2-s_2} - (1-x)^{k_2}(2(1-r)u+(1-u))^{k_2-s_2}\right)\\
&= (1-u)^{k_1}x^{k_1}\left((1-x)(u+(1-x)(1-u))^{k_2} - (1-x)^{k_2}(1+(1-r)u)^{k_2}\right).
\end{aligned}$$

Putting this all together, we obtain

$$\begin{aligned}
\mathscr{L}^r_{\Lambda D^*} f_{\boldsymbol{k}}(x) &= \int_0^1 \big[x(x+(1-x)ru)^{k_1}((1-x)(1-ru))^{k_2} + (1-x)((1-u)x)^{k_1}(1-(1-u)x)^{k_2}\\
&\qquad - x^{k_1}(1-x)^{k_2}\big]\nu(du) + \mathscr{L}^{\tilde{\boldsymbol{K}}} f_{\boldsymbol{k}}(x) - \mathscr{L}^{\tilde{\boldsymbol{K}}} f_{\boldsymbol{k}}(x)\\
&= \mathscr{L}^{\tilde{\boldsymbol{K}}} f_{\boldsymbol{k}}(x) + \int_0^1 \big[x(x+(1-x)ru)^{k_1}((1-x)(1-ru))^{k_2}\nu(du)
\end{aligned}$$

$$
\begin{aligned}
&+ \int_0^1 (1-x)((1-u)x)^{k_1}(1-(1-u)x)^{k_2}\nu(du) - \int_0^1 \nu(du) f_{\boldsymbol{k}}(x) \\
&- \Bigg(\int_0^1 \left[x(ru+(1-ru)x)^{k_1}(1-ru)^{k_2}(1-x)^{k_2} \right] \nu(du) \\
&\quad + \int_0^1 \left[(1-x)(1-u)^{k_1}x^{k_1}(u+(1-x)(1-u))^{k_2} \right] \nu(du) \\
&\quad - \int_0^1 \left[(1-ru)^{k_2} + (1-u)^{k_1}(1+(1-r)u)^{k_2} \right] \nu(du) f_{\boldsymbol{k}}(x) \\
&\quad + \int_0^1 (1-u)^{|\boldsymbol{k}|}\nu(du) f_{\boldsymbol{k}}(x) \Bigg) \\
&= \mathscr{L}^{\tilde{\boldsymbol{K}}} f_{\boldsymbol{k}}(x) + \int_0^1 R_{\boldsymbol{k}}(r,u)\nu(du) f_{\boldsymbol{k}}(x).
\end{aligned}
$$

□

Earlier, we suggested that $R_{\boldsymbol{k}}(r,u)$ has another representation as a linear combination. This is expressed formally in the following proposition.

Proposition 5.5. *Let $\mathcal{C}_{\boldsymbol{k}}$ be the set of circle configurations for the simple genealogy in sample state $\boldsymbol{k}$. For $c \in \mathcal{C}_{\boldsymbol{k}}$, define $n(c)$ to be the number of ways to lay c onto the sample and $p(c)$ to be the probability of seeing c during a u-event (whose r- and u-dependence we suppress). Finally, let $o(c)$ be the number of options generated by c given the sample $\boldsymbol{k}$, which is 1 for ordinary events and 0 for incompatible events. Then,*

$$
R_{\boldsymbol{k}}(r,u) = \sum_{c \in \mathcal{C}_{\boldsymbol{k}}} n(c)p(c)\left(o(c)-1\right). \tag{5.20}
$$

Proof. For the partition that we used in the proof of Theorem 5.3, the following list details the type of event generated by each c in that part, along with the number of options associated to that event (if appropriate) – the bullet points indicate the strength of the circles relating to that part of the partition:

- $\{s_1 = 0, s_2 = 0\}$:
  - $\{w_1 = 0, w_2 = 0\}$: not in $\mathcal{C}_{\boldsymbol{k}}$,
  - $\{w_1 = 0, w_2 > 0\}$: optional $(2^{w_2}+1)$,
  - $\{w_1 > 0, w_2 \geq 0\}$: ordinary (1),
- $\{s_1 = 0, s_2 > 0\}$:
  - $\{w_1 = 0, w_2 = 0\}$: ordinary (1),
  - $\{w_1 = 0, w_2 > 0\}$: optional (2^{w_2}),
  - $\{w_1 > 0, w_2 \geq 0\}$: incompatible (0),
- $\{s_1 > 0, s_2 = 0\}$: ordinary (1),
- $\{s_1 > 0, s_2 > 0\}$: incompatible (0).

We shall partition the set of circle configurations into $\mathcal{C}_{\boldsymbol{k}} = \mathcal{C}^{inc} \cup \mathcal{C}^{ord} \cup \mathcal{C}^{opt}$, which respectively contain the configurations leading to incompatible, ordinary and optional events. For all $c \in \mathcal{C}^{inc}$, there are zero options so $o(c) - 1 = -1$. Similarly, $\forall c \in \mathcal{C}^{ord}$, there is exactly one option, which means the contribution from these configurations vanishes. For $c \in \mathcal{C}^{opt}$, the number of options depends on the circle configuration, and so no factorization occurs. Hence,

$$\sum_{c \in \mathcal{C}_{\boldsymbol{k}}} n(c)p(c)\,(o(c)-1) = \sum_{c \in \mathcal{C}^{opt}} n(c)p(c)\,(o(c)-1) - \sum_{c \in \mathcal{C}^{inc}} n(c)p(c).$$

Then, referring to the above list, there are two sources of both optional and incompatible events, which we must collect together;

$$\begin{aligned}
\sum_{c \in \mathcal{C}_{\boldsymbol{k}}} n(c)p(c)\,(o(c)-1) &= \Sigma_1 + \Sigma_2 - [\Sigma_3 + \Sigma_4] \\
&= \sum_{s_2=1}^{k_2} \sum_{w_2=1}^{k_2-s_2} \binom{k_2}{s_2}\binom{k_2-s_2}{w_2} (ru)^{s_2}((1-r)u)^{w_2}(1-u)^{|\boldsymbol{k}|-s_2-w_2}\,(2^{w_2}-1) \\
&\quad + \sum_{w_2=1}^{k_2} \binom{k_2}{w_2} ((1-r)u)^{w_2}(1-u)^{|\boldsymbol{k}|-w_2}\,(2^{w_2}+1-1) \\
&\quad - \Bigg[\sum_{s_1=1}^{k_1} \sum_{s_2=1}^{k_2} \sum_{w_1=0}^{k_1-s_1} \sum_{w_2=0}^{k_2-s_2} \binom{k_1}{s_1}\binom{k_2}{s_2}\binom{k_1-s_1}{w_1}\binom{k_2-s_2}{w_2} \\
&\qquad \times (ru)^{s_1+s_2}((1-r)u)^{w_1+w_2}(1-u)^{|\boldsymbol{k}|-(s_1+s_2)-(w_1+w_2)} \\
&\quad + \sum_{s_2=1}^{k_2} \sum_{w_2=0}^{k_2-s_2} \sum_{w_1=1}^{k_1} \binom{k_2}{s_2}\binom{k_2-s_2}{w_2}\binom{k_1}{w_1} \\
&\qquad \times (ru)^{s_2}((1-r)u)^{w_1+w_2}(1-u)^{|\boldsymbol{k}|-s_2-(w_1+w_2)} \Bigg].
\end{aligned}$$

With care, the first two terms can be merged, since the second sum somewhat resembles the $s_2 = 0$ term of the first. In so doing, the sum is equal to

$$\begin{aligned}
\Sigma_1 + \Sigma_2 &= \sum_{s_2=0}^{k_2} \binom{k_2}{s_2} (ru)^{s_2} \sum_{w_2=1}^{k_2-s_2} \binom{k_2-s_2}{w_2} (2(1-r)u)^{w_2}(1-u)^{k_1+k_2-s_2-w_2} \\
&\quad - \sum_{s_2=1}^{k_2} \binom{k_2}{s_2} (ru)^{s_2} \sum_{w_2=1}^{k_2-s_2} \binom{k_2-s_2}{w_2} ((1-r)u)^{w_2}(1-u)^{k_1+k_2-s_2-w_2}
\end{aligned}$$

$$= (1-u)^{k_1} \sum_{s_2=0}^{k_2} \binom{k_2}{s_2} (ru)^{s_2} \left((1+u-2ru)^{k_2-s_2} - (1-u)^{k_2-s_2} \right)$$
$$- (1-u)^{k_1} \sum_{s_2=1}^{k_2} \binom{k_2}{s_2} (ru)^{s_2} \left((1-ru)^{k_2-s_2} - (1-u)^{k_2-s_2} \right)$$
$$= (1-u)^{k_1} \left((1+(1-r)u)^{k_2} + (1-ru)^{k_2} - (1-u)^{k_2} - 1 \right).$$

Next, we concentrate on simplifying the sums corresponding to incompatibilities separately. Firstly,

$$\Sigma_3 = \left(\sum_{s_1=0}^{k_1} \tbinom{k_1}{s_1} (ru)^{s_1} (1-ru)^{k_1-s_1} - (1-ru)^{k_1} \right) \left(\sum_{s_2=0}^{k_2} \tbinom{k_2}{s_2} (ru)^{s_2} (1-ru)^{k_2-s_2} - (1-ru)^{k_2} \right)$$
$$= 1 - (1-ru)^{k_1} - (1-ru)^{k_2} + (1-ru)^{k_1+k_2}.$$

The final term (triple sum) is equal to

$$\Sigma_4 = \left((1-ru)^{k_1} - (1-u)^{k_1} \right) \sum_{s_2=1}^{k_2} \binom{k_2}{s_2} (ru)^{s_2} \sum_{w_2=0}^{k_2-s_2} \binom{k_2-s_2}{w_2} ((1-r)u)^{w_2} (1-u)^{k_2-s_2-w_2}$$
$$= \left((1-ru)^{k_1} - (1-u)^{k_1} \right) \left(1 - (1-ru)^{k_2} \right).$$

Adding up these terms and making the necessary cancellations yields the promised $R_{\boldsymbol{k}}(r,u)$. □

We should point out that, with the help of Proposition 2.6, Theorem 5.3 demonstrates that the Λ-Fleming–Viot model with differential killing/birth, but *without* mutation, is dual in the broad sense to a process $\tilde{\boldsymbol{K}}_t$ (implicitly characterized by the generator $\mathscr{L}^{\tilde{\boldsymbol{K}}}$) with respect to the triple of functions

$$\left(f_{\boldsymbol{k}}(x), 0, \int_0^1 R_{\boldsymbol{k}}(r,u)\nu(du) \right).$$

However, in order to form a duality in the narrow sense via the method of Section 2.2.1, there must be some form mutation in order to induce a stationary distribution. This means that we have to slightly deconstruct the characterization of Theorem 5.3 to accommodate the mutation terms. We model the mutation as Poisson point processes acting on the lineages, independently of the reproduction events, as outlined in Section 5.4.

Lemma 5.6. *The generator of the Λ-Fleming–Viot process with differential killing/birth and weak mutation acting on functions of the form $f_{\boldsymbol{k}}(x) = x^{k_1}(1-x)^{k_2}$, $\boldsymbol{k} > \mathbf{0}$, can be written as*

$$\mathscr{L}^r_{\Lambda D * m} f_{\boldsymbol{k}}(x) = \sum_{\boldsymbol{k}' \neq \boldsymbol{k}} c(\boldsymbol{k}, \boldsymbol{k}') f_{\boldsymbol{k}'}(x) - c(\boldsymbol{k}, \boldsymbol{k}) f_{\boldsymbol{k}}(x) \tag{5.21}$$

where $c(\boldsymbol{k}, \boldsymbol{k}') \geq 0$, $\forall \boldsymbol{k}, \boldsymbol{k}' > \mathbf{0}$. The constants are

$$c(\boldsymbol{k}, \boldsymbol{k}) = k_1\nu_1 + k_2\nu_2 + \int_0^1 \left[1 - (1-u)^{|\boldsymbol{k}|} - k_1 r u (1-ru)^{|\boldsymbol{k}|-1} - k_2 u (1-u)^{|\boldsymbol{k}|-1}\right] \nu(du) \tag{5.22}$$

and

$$c(\boldsymbol{k}, \boldsymbol{k}') = \begin{cases} k_1\nu_2, & \boldsymbol{k}' = \boldsymbol{k} - \boldsymbol{e}_1 + \boldsymbol{e}_2, \\ k_2\nu_1, & \boldsymbol{k}' = \boldsymbol{k} + \boldsymbol{e}_1 - \boldsymbol{e}_2, \\ \int_0^1 (1-ru)^{|\boldsymbol{k}|} - (1-u)^{|\boldsymbol{k}|} \nu(du), & \boldsymbol{k}' = \boldsymbol{k} + \boldsymbol{e}_1, \\ \binom{k_1}{s_1} \int_0^1 (ru)^{s_1} (1-ru)^{|\boldsymbol{k}|-s_1} \nu(du), & \boldsymbol{k}' = (k_1 - s_1 + 1, k_2),\ 2 \leq s_2 \leq k_2, \\ \binom{k_2}{s_2}\binom{k_2 - s_2}{w_2} I_2^\nu(s_2, w_2), & \boldsymbol{k}' = (k_1, k_2 - s_2 - w_2 + 1), \\ & \quad 2 \leq s_2 \leq k_2,\ 0 \leq w_2 \leq k_2 - s_2, \\ \left(1 + \frac{r}{1-r} w_2\right) \binom{k_2}{w_2} I_2^\nu(0, w_2), & \boldsymbol{k}' = (k_1, k_2 - w_2 + 1),\ 2 \leq w_2 \leq k_2. \end{cases} \tag{5.23}$$

Corollary 5.7. *Let X be the stationary distribution of the Λ-Fleming–Viot process with differential killing/birth and weak mutation (that is characterized by the generator $\mathscr{L}^r_{\Lambda D * m}$ and its domain). Then it is dual in the narrow sense to a discrete pure-jump process $\boldsymbol{M}_t$ with respect to the duality function*

$$g_{\boldsymbol{k}}(x) = \frac{f_{\boldsymbol{k}}(x)}{m_{\boldsymbol{k}}}, \qquad m_{\boldsymbol{k}} = \mathbb{E}[X^{k_1}(1-X)^{k_2}]. \tag{5.24}$$

The dual process can be characterized by the generator

$$\mathscr{L}^{\boldsymbol{M}} h(\boldsymbol{k}) = \sum_{\boldsymbol{k}' \neq \boldsymbol{k}} c(\boldsymbol{k}, \boldsymbol{k}') \frac{m_{\boldsymbol{k}'}}{m_{\boldsymbol{k}}} \left(h(\boldsymbol{k}') - h(\boldsymbol{k})\right), \tag{5.25}$$

where $h(\boldsymbol{k}) \in \mathscr{D}(\mathscr{L}^{\boldsymbol{M}})$ and $c(\boldsymbol{k}, \boldsymbol{k}')$ is defined in (5.23).

Proof. Using Lemma 5.6, we have written $\mathscr{L}^r_{\Lambda D*m} f_{\boldsymbol{k}}(x)$ in the form of (2.11) through (5.21), and can therefore use the machinery of Section 2.2.1 to deduce the result. □

Proof of Lemma 5.6. Recall that the generator of this process is defined by the sum of two others, $\mathscr{L}^r_{\Lambda D*} + \mathscr{L}_m$, the first of which corresponds to reproduction, the second to mutation. Then, observe that

$$\mathscr{L}_m f_{\boldsymbol{k}}(x) = k_1\nu_2 f_{\boldsymbol{k}-\boldsymbol{e}_1+\boldsymbol{e}_2}(x) + k_2\nu_1 f_{\boldsymbol{k}+\boldsymbol{e}_1-\boldsymbol{e}_2}(x) - (k_1\nu_1 + k_2\nu_2) f_{\boldsymbol{k}}(x). \tag{5.26}$$

We shall now write $\mathscr{L}^r_{\Lambda D*m} f_{\boldsymbol{k}}(x)$, given by the sum of (5.15) and (5.26), in the form of (5.21), a special case of (2.11). We must extract all those terms which are proportional to $f_{\boldsymbol{k}}(x)$. Since the generator terms in (5.15) are usually of the form $f_{\boldsymbol{k}'}(x) - f_{\boldsymbol{k}}(x)$, we can simply split these in half to obtain many of them. Upon doing so, major simplifications ensue because $f_{\boldsymbol{k}}(x)$ factorizes out of the sums, and they can again be realized as binomial series, iterated three times. It follows that

$$\begin{aligned}
\mathscr{L}^r_{\Lambda D*m} f_{\boldsymbol{k}}(x) &= \left(\int_0^1 R_{\boldsymbol{k}}(r,u)\nu(du) - \int_0^1 (1-ru)^{k_2} + (1-u)^{k_1}(1+(1-r)u)^{k_2}\nu(du) \right) f_{\boldsymbol{k}}(x) \\
&+ \int_0^1 \Bigg[\sum_{(s_1,w_1,w_2)\in\mathcal{C}_{1,\boldsymbol{k}}} \tbinom{k_1}{s_1}\tbinom{k_1-s_1}{w_1}\tbinom{k_2}{w_2} (ru)^{s_1}((1-r)u)^{w_1+w_2}(1-u)^{|\boldsymbol{k}|-s_1-(w_1+w_2)} f_{(k_1-s_1+1,\, k_2)}(x) \\
&+ \sum_{(s_2,w_2)\in\mathcal{C}_{2,\boldsymbol{k}}} \tbinom{k_2}{s_2}\tbinom{k_2-s_2}{w_2} (ru)^{s_2}((1-r)u)^{w_2}(1-u)^{|\boldsymbol{k}|-s_2-w_2} f_{(k_1,\, k_2-s_2-w_2+1)}(x) \Bigg] \nu(du) \\
&+ k_1\nu_2 f_{\boldsymbol{k}-\boldsymbol{e}_1+\boldsymbol{e}_2}(x) + k_2\nu_1 f_{\boldsymbol{k}+\boldsymbol{e}_1-\boldsymbol{e}_2}(x) - (k_1\nu_1 + k_2\nu_2) f_{\boldsymbol{k}}(x).
\end{aligned} \tag{5.27}$$

Some supplementary simplifications have been performed on the second summation to remove the sum that was indexed by i. There are now some cancellations in the first line that simplify the constant in front of $f_{\boldsymbol{k}}(x)$. However, more work needs to be done to extract further terms that are proportional to $f_{\boldsymbol{k}}(x)$ that remain in the sums. For example, observe what happens upon setting $s_1 = 1$ in the first sum, and $s_2 + w_2 = 1$ in the second.

With this in mind, the first sum in (5.27) simplifies to

$$\sum_{s_1=2}^{k_1}\binom{k_1}{s_1}(ru)^{s_1}(1-ru)^{|\boldsymbol{k}|-s_1}f_{(k_1-s_1+1,\,k_2)}(x)+\left((1-ru)^{|\boldsymbol{k}|}-(1-u)^{|\boldsymbol{k}|}\right)f_{\boldsymbol{k}+\boldsymbol{e}_1}(x)$$
$$+k_1ru(1-ru)^{|\boldsymbol{k}|-1}f_{\boldsymbol{k}}(x).$$

The first line gives some of the constants in (5.23), whilst the second line contributes to $c(\boldsymbol{k},\boldsymbol{k})$. Evidently,

$$(1-ru)^{|\boldsymbol{k}|}-(1-u)^{|\boldsymbol{k}|}\geq 0,\quad \forall|\boldsymbol{k}|\geq 1,\ r\in(0,1],\ u\in(0,1],$$

and equality is attained if and only if $r=1$. Making the necessary extractions in the second sum and simplifying yields

$$\sum_{s_2=2}^{k_2}\sum_{w_2=0}^{k_2-s_2}\binom{k_2}{s_2}\binom{k_2-s_2}{w_2}(ru)^{s_2}((1-r)u)^{w_2}(1-u)^{|\boldsymbol{k}|-s_2-w_2}f_{(k_1,\,k_2-s_2-w_2+1)}(x)$$
$$+\sum_{w_2=2}^{k_2}\left(1+\frac{r}{1-r}w_2\right)\binom{k_2}{w_2}((1-r)u)^{w_2}(1-u)^{|\boldsymbol{k}|-w_2}f_{(k_1,\,k_2-w_2+1)}(x)$$
$$+k_2u(1-u)^{|\boldsymbol{k}|-1}f_{\boldsymbol{k}}(x).$$

The first two lines provide constants for (5.23), and the last will be incorporated into $c(\boldsymbol{k},\boldsymbol{k})$. We have now listed all the constants in (5.23), and it is not difficult to see that they are non-negative and that the vectors $\boldsymbol{k}'$ genuinely differ from $\boldsymbol{k}$. It now only remains to prove that $c(\boldsymbol{k},\boldsymbol{k})>0$, which is perhaps less obvious.

Since $\nu_1,\nu_2>0$, and we are considering vectors $\boldsymbol{k}$ for which $|\boldsymbol{k}|\geq 1$ holds, implying that $k_1\nu_1+k_2\nu_2>0$, it is enough to prove the inequality

$$(1-u)^{|\boldsymbol{k}|}+k_1ru(1-ru)^{|\boldsymbol{k}|-1}+k_2u(1-u)^{|\boldsymbol{k}|-1}\leq 1,\ \forall r\in(0,1],\ u\in(0,1],$$

which is achieved in Lemma C.5.1. It follows that $c(\boldsymbol{k},\boldsymbol{k})>0$ and the proof is complete. □

5.5.3 Real Genealogy

We showed in Section 5.5.2 that the simple genealogies of the differential killing and birth models with weak mutation coincide, in both the weak and strong senses, when the events are appropriately coupled. However, the matter cannot be that simple, as we observe disagreements between them in the displaced column of Table 5.1. This implies that the true genealogical structure relating the lineages behaves differently under each model. In this section, we shall introduce the *real* genealogy, which is able to capture a much finer level of detail than its simple counterpart. The real genealogies of the two models are genuinely different, even when coupled. The ideas herein are similar in spirit to those of the ancestral selection graph of Krone and Neuhauser [1997] and Neuhauser and Krone [1997].

The key concept in building a real genealogy is that of the *reality* of a lineage: each lineage can be in one of two states – real or virtual. Every lineage starts off as real when it is sampled at the present time, and, initially, represents an actual individual living in the population. Through a succession of reproduction events, the lineage may remain real but represent a relative in an earlier generation. On the other hand, it may become virtual; if it does so, it stays in that state forevermore. Virtual lineages can be lost through coalescence, however. If we remove the information about the realities of the lineages in our sample, then we recover the simple genealogy discussed in Section 5.5.2.

The real lineages trace out the true genealogical structure of our sample. Using the very fundamental observation that each child may have exactly one parent, we deduce that the number of real lineages cannot increase, but can of course decrease through coalescence events. However, we observed that pure branching events were possible in the simple genealogy, and so the increase must be accounted for by the proliferation of virtual lineages. A lineage is virtual whenever the individual that would have been living on that line is not related to any of the individuals in the original sample. For this reason, it may seem as though it would not be necessary to keep track of them, but their presence yields some essential information about the allelic frequencies of the forwards-in-time process. We should note, however, that there is no notion of reality in the forwards-in-time models, and consequently there is no generator nor duality to work with in this setting. For this reason, it does not make sense to talk about a weak, real genealogy, only a strong one.

We shall say more about the motivation behind this more detailed version of the genealogy. Typically, one would be interested in tracing back until the *ultimate ancestor* in this branching-coalescing structure in order to later infer details about the true genealogy, such as the time to the most recent common ancestor (MRCA) – this is how the ancestral selection graph may be utilized. However, there are two reasons why this is not a good idea in our regime.

Firstly, it is quite computationally costly to keep track of the whole structure of the strong, simple genealogy, especially when iterating the process over many realizations. It would be much easier to track a smaller amount of information, such as that in the weak, simple genealogy of Theorem 5.3. However, making such a reduction throws away essential information that would allow us to make statements about the MRCA. It would be possible to find the time to the ultimate ancestor, but it is not clear how useful that information is if the graph structure is lost.

Secondly, and perhaps more importantly, there are changes that occur in the real genealogy that cannot be detected by the simple one. We shall coin these *silent events*, and discuss them at length in the sequel. For this reason, it is not obvious exactly what the simple genealogy represents. The real genealogy, on the other hand, can detect all the changes that affect our sample, and is therefore a worthy object to study. In addition, we shall show that it is possible to discard most of the branching-coalescing graph of the real genealogy so that we only keep track of the number of real and virtual lineages of each type. From this, it is possible to infer the time to MRCA – it is simply the stopping time at which the process contains exactly one real lineage.

In order to characterize the real genealogy, we need to understand how events affect the distribution of real (typed) lineages. That is, if we know types and realities of lineages before an event (in genealogical time), what are they after the event? The types are given by the simple genealogy, so we now just need to consider the realities of lineages. After a reproductive event, there will be one parental lineage, and possibly some continuing lineages (generated by weak circles). Any lineage struck by a strong circle will coalesce into the parental lineage. We also need to understand how mutation affects the real genealogy, but it turns out to be straightforward to describe. Once again we reiterate that all lineages are real when the sample is taken (at time zero in genealogical time).

To begin, we shall state what happens during reproduction and mutation events in words. Firstly, consider the differential killing model. Recall that, after a reproduction event, a strong circle will kill either allelic type, but a weak circle will only kill type-a_1 individuals. Hence, if the continuing lineage is of type a_1, it will become virtual if it was not already, since the a_1-individual that was residing on that line would have been killed by the weak circle, and so the parent of that event laid an offspring there. The individual that was killed in this scenario therefore has no relation to anyone in the sample. On the other hand, a type-a_2 continuing lineage will retain its reality; that is, if it was real before the event, it was not affected by the weak circle, since the individual that it represents was of type a_2. The same is true if it was virtual before the event, but the individual represented by the lineage lost its connection to the sample in a previous event. Recall that once a lineage has become virtual, it will stay that way for the rest of the genealogical process.

The parental lineage will be real if she successfully gave birth to an individual who is currently represented by a real lineage in the sample (or virtual if she fails to do so). There are two ways in which this can be achieved: if a strong circle strikes a real lineage of either type, or a weak circle strikes a real lineage of type a_1. In the former case, the strong circle would have killed either allelic type, resulting in a successful birth for the parent. In the latter, the weak circle can only kill type-a_1 individuals, so the parent could have only given birth to an offspring represented by an a_1-lineage. It is important that real lineages, not virtual, are hit by the circles, because the individuals represented by virtual lineages do not carry any genetic material relating to the original sample. Note that when a set of virtual lineages coalesce, the resultant parent remains virtual, but a merger containing both real and virtual lineages will lead to a real parent.

Finally, when a mutation strikes a lineage, it does not affect its reality, but does swap its genetic type.

For the differential birth model, the focus is on the parent: a parent of type a_1 may only lay offspring on strong circles, whereas parents of type a_2, the stronger type, may lay offspring on either weak or strong circles. As before, we shall consider the continuing lineages first. If a (potential) parent is of type a_1, then all the continuing lineages retain their reality as the parent cannot lay its offspring on weak circles (forwards in time, all the individuals struck by weak circles would have survived). If the parent

is of type a_2, all continuing lineages become virtual if they were not before, since the parent would have had successful offspring on every circle, weak or strong. This type of event would have led to a coalescence of every lineage struck by a circle, and all the genetic material of those lineages relating to the sample (if there was any) would be transferred to the parent.

Consider the parental lineage. Firstly, observe that if a strong circle strikes a real lineage then a parent of either type would have given birth to the individual represented by that real lineage, and so receives the genetic material. Consequently, the parent is real in this case. The only other way that the parent could be real is if it is of type a_2 and a weak circle struck a real lineage. In any other scenario, the parent failed to give birth to a real lineage, so must be virtual.

Once again, when a mutation falls upon a lineage, it will not affect the reality of that lineage, just the type.

This description fully explains each sort of event that can take place in the real genealogies, but we have to do some more work if we want to derive the rates at which such events occur. Before this, we shall describe an explicit example of how the real genealogies differ in the two models. It should be evident from the above characterization that the processes behave in different ways, but Figure 5.6 demonstrates this. It illustrates what may happen to a sample consisting of two real type-a_1 lineages. Real lineages are depicted as black, bold lines, and virtual lineages by red, dashed lines. The top layer ((a), (b) and (c)) correspond to the differential killing model, whilst the bottom layer ((d), (e) and (f)) represent the exact same events for the differential birth model. The second and third columns provide examples of where the models differ – for differential killing, the events both lead to a coalescence of the lineages, and therefore virtual continuing lineages. This is not the case for differential birth. Note also that the numbers, not just the positions, of real and virtual lineages are unequal after the event. In (f), the parent is virtual because, forwards in time, it could not successfully displace the individuals struck by weak circles, so the real lineages retain their reality.

Remark 5.8. A similar phenomenon has been noted by Taylor [2009]. He considers a diploid version of the Wright–Fisher diffusion with weak mutation, recombination and fluctuating selection and proves that multiple coalescent processes exist for a single forwards-in-time process of allele frequencies. In-

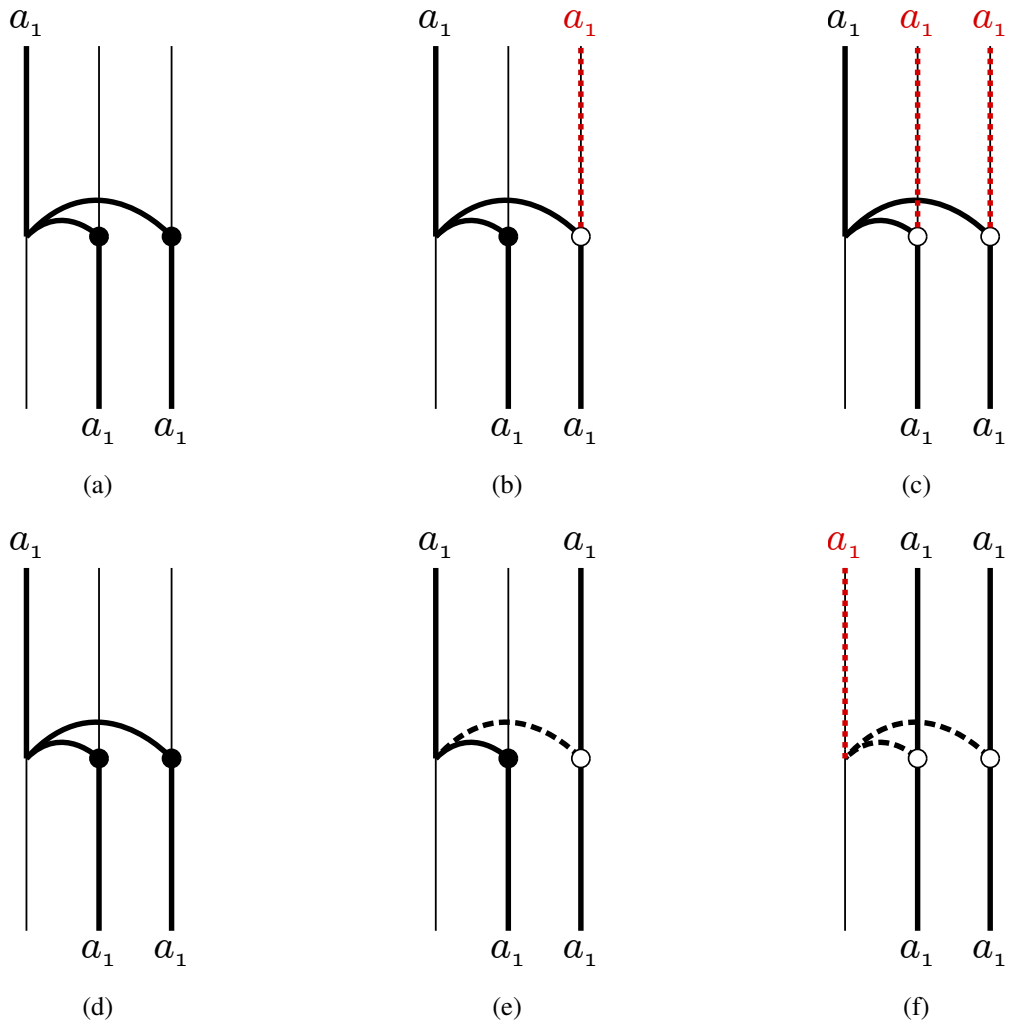


Figure 5.6: Examples of the differences between the real genealogies of the differential killing (top row) and differential birth (bottom row) models. The black bold lines correspond to real lineages, and red dashed lines to virtual ones. (b) exemplifies a silent event

terestingly, this fails to be the case in the corresponding haploid model. To the best of this author's knowledge, this thesis is the first time that such a phenomenon has been discovered in a haploid model.

Figure 5.6 (b) is a particularly interesting example as it is a *silent event* (defined formally in Definition 5.9). Let's suppose we summarize the state of the strong, real genealogy by a 2×2 matrix

$$K = \begin{pmatrix} k_1^r & k_2^r \\ k_1^v & k_2^v \end{pmatrix}, \tag{5.28}$$

where k_i^r and k_i^v are respectively the number of real and virtual lineages of type a_i. Then it is clear that the process in this example experiences the transition $\left(\begin{smallmatrix}2&0\\0&0\end{smallmatrix}\right) \to \left(\begin{smallmatrix}1&0\\1&0\end{smallmatrix}\right)$, a genuine change of state. However, if we go back to the weak, simple genealogy by forgetting the realities of the lineages (i.e. by taking column totals), then the state stays at $(2, 0)$ before and after the event. This is a significant observation because it means that there exist coalescence events that are undetected by the simple genealogy. This phenomenon has the consequence that the simple genealogy does not really represent the true genealogy, for the differential killing model at least. Later, we will show that silent events also occur for the differential birth model, but under alternative circumstances. Because of their importance, we shall define them formally.

Definition 5.9. A *silent event* is a reproductive event which causes a change of state in the real genealogy, whilst leaving the weak, simple genealogy unaffected.

We would now like to derive a detailed description of the transitions that the real genealogies make, and the rates at which they happen. We will in some sense mimic the proof of Theorem 5.3, by carefully considering the circle configurations that might fall onto our real sample, and the effects they have. Unfortunately, in this setting, we do not have a duality to use. We need some more notation in order to handle the more detailed circle configurations under consideration. Let s_i^r and $s_i^v = s_i - s_i^r$ respectively be the number of strong circles that strike real and virtual lineages of type a_i. Similarly, define w_i^r and $w_i^v = w_i - w_i^r$ to be the number of weak circles that strike real and virtual lineages of type a_i.

Suppose that our real genealogy is in state K, the 2×2 matrix given in (5.28), with the constraint $k_1^r + k_2^r > 0$ – there must be at least one real lineage otherwise there is no relevant genetic material. The rate at which we see a circle configuration striking in any given permutation is independent of the real and virtual lineages, but does depend on the column totals k_1 and k_2, as before. They are given by the integrals I_1^ν and I_2^ν defined in (5.13) and (5.14), depending on whether the parent was of type a_1 or a_2 respectively.

However, we need to perform the combinatorics more carefully, as the number of ways of laying the finer circle configuration does depend on the numbers of real and virtual lineages. Starting with the

events in which the parent is of type a_1, then, for a fixed s_1, there are

$$\binom{k_1^r}{s_1^r}\binom{k_1 - k_1^r}{s_1 - s_1^r} = \binom{k_1^r}{s_1^r}\binom{k_1^v}{s_1^v}$$

ways of laying s_1^r strong circles onto real a_1-lineages and s_1^v strong circles onto virtual a_1-lineages. As a convention, we shall let $s_1^r \in [s_1]_0$ be a free variable, forcing $s_1^v = s_1 - s_1^r$ to be fixed. Note that, making use of the Vandermonde identity,

$$\binom{k_1}{s_1} = \sum_{s_1^r=0}^{s_1} \binom{k_1^r}{s_1^r}\binom{k_1 - k_1^r}{s_1 - s_1^r},$$

which accounts for the first binomial coefficient of $\mathscr{L}_1$ of Theorem 5.3. Given this placement, there are $k_1^r - s_1^r$ real and $k_1^v - s_1^v$ virtual a_1-lineages remaining on which to lay w_1 weak circles. Hence, the number of ways of laying w_1^r (respectively w_1^v) weak circles onto real (resp. virtual) a_1-lineages is

$$\binom{k_1^r - s_1^r}{w_1^r}\binom{k_1^v - s_1^v}{w_1^v},$$

where $w_1^r \in [w_1]_0$. A similar Vandermonde identity yields

$$\binom{k_1 - s_1}{w_1} = \sum_{w_1^r=0}^{w_1} \binom{k_1^r - s_1^r}{w_1^r}\binom{k_1 - s_1 - (k_1^r - s_1^r)}{w_1 - w_1^r} = \sum_{w_1^r=0}^{w_1} \binom{k_1^r - s_1^r}{w_1^r}\binom{k_1^v - s_1^v}{w_1^v},$$

which is the second binomial term in $\mathscr{L}_1$. Completely analogous arguments can be used to explain how the other binomial coefficients in Theorem 5.3 break down, so those calculations will be omitted.

Each of the rates that we list below will also contain a ratio of moments, a Bayes factor, of the type that we have seen so often throughout this thesis. This factor corrects the conditioning upon knowing the *types* in our sample, which gives us some information about how the allelic proportions must have behaved in the forwards-in-time model. By inspecting the real genealogy instead of the simple one, we have not gained any more information about the forwards-in-time process, as reality is not a forwards-in-time concept. For this reason, the Bayes factor must only depend the numbers of types in our sample, not the realities, and hence must be equal to

$$\frac{m_{\boldsymbol{k}'}}{m_{\boldsymbol{k}}},$$

as in the simple genealogy. The vectors $\boldsymbol{k}, \boldsymbol{k}'$ represent the matrices K, K' after taking the column totals, and the dash represents a new state. Notice that for silent events, for which the transition does not result in a change of the column totals, the Bayes factor will be equal to one, since $\boldsymbol{k}' = \boldsymbol{k}$.

The analysis so far has been consistent between the two models, but the following two theorems, which characterize the real genealogies, will outline how they disagree.

Theorem 5.10. *The real genealogy of the Λ-Fleming–Viot model with differential killing and weak mutation, when in state K, undergoes the transition (reproduction with a_1-parent)*

$$\begin{pmatrix} k_1^r & k_2^r \\ k_1^v & k_2^v \end{pmatrix} \longrightarrow \begin{pmatrix} k_1^r - s_1^r - w_1^r + \mathbb{1}\{s_1^r + w_1^r > 0\} & k_2^r \\ k_1^v - s_1^v + w_1^r + \mathbb{1}\{s_1^r = w_1^r = 0\} & k_2^v \end{pmatrix} \tag{5.29}$$

at rate

$$\frac{m_{(k_1 - s_1 + 1, k_2)}}{m_{\boldsymbol{k}}} \binom{k_1^r}{s_1^r} \binom{k_1^v}{s_1^v} \binom{k_1^r - s_1^r}{w_1^r} \binom{k_1^v - s_1^v}{w_1^v} \binom{k_2^r}{w_2^r} \binom{k_2^v}{w_2^v} I_1^{\nu}(s_1, w_1, w_2)$$

where $(s_1, w_1, w_2) \in \mathcal{C}_{1,\boldsymbol{k}}$, $s_1^r \in [s_1]_0, w_1^r \in [w_1]_0$ and $w_2^r \in [w_2]_0$, and the transition (reproduction with a_2-parent)

$$\begin{pmatrix} k_1^r & k_2^r \\ k_1^v & k_2^v \end{pmatrix} \longrightarrow \begin{pmatrix} k_1^r & k_2^r - s_2^r - i^r + \mathbb{1}\{s_2^r > 0\} \\ k_1^v + i^r + i^v & k_2^v - s_2^v - i^v + \mathbb{1}\{s_2^r = 0\} \end{pmatrix} \tag{5.30}$$

at rate

$$\frac{m_{(k_1 + i, k_2 - s_2 - i + 1)}}{m_{\boldsymbol{k}}} \binom{k_2^r}{s_2^r} \binom{k_2^v}{s_2^v} \binom{k_2^r - s_2^r}{w_2^r} \binom{k_2^v - s_2^v}{w_2^v} \binom{w_2^r}{i^r} \binom{w_2^v}{i^v} I_2^{\nu}(s_2, w_2),$$

where $(s_2, w_2) \in \mathcal{C}_{2,\boldsymbol{k}}$, $s_2^r \in [s_2]_0, w_2^r \in [w_2]_0, i^r \in [w_2^r]_0, i^v \in [w_2^v]_0$ and $i = i^r + i^v$. Finally, due to mutation, the state K could transition to

$$\begin{pmatrix} k_1^r - 1 & k_2^r + 1 \\ k_1^v & k_2^v \end{pmatrix}, \quad \begin{pmatrix} k_1^r & k_2^r \\ k_1^v - 1 & k_2^v + 1 \end{pmatrix}, \quad \begin{pmatrix} k_1^r + 1 & k_2^r - 1 \\ k_1^v & k_2^v \end{pmatrix}, \quad \begin{pmatrix} k_1^r & k_2^r \\ k_1^v + 1 & k_2^v - 1 \end{pmatrix}$$

at rates

$$\frac{m_{\boldsymbol{k} - \boldsymbol{e}_1 + \boldsymbol{e}_2}}{m_{\boldsymbol{k}}} k_1^r \nu_2, \quad \frac{m_{\boldsymbol{k} - \boldsymbol{e}_1 + \boldsymbol{e}_2}}{m_{\boldsymbol{k}}} k_1^v \nu_2, \quad \frac{m_{\boldsymbol{k} + \boldsymbol{e}_1 - \boldsymbol{e}_2}}{m_{\boldsymbol{k}}} k_2^r \nu_1, \quad \frac{m_{\boldsymbol{k} + \boldsymbol{e}_1 - \boldsymbol{e}_2}}{m_{\boldsymbol{k}}} k_2^v \nu_1$$

respectively.

Theorem 5.11. *The real genealogy of the Λ-Fleming–Viot model with differential birth and weak mutation, when in state K, undergoes the transition (reproduction with a_1-parent)*

$$\begin{pmatrix} k_1^r & k_2^r \\ k_1^v & k_2^v \end{pmatrix} \longrightarrow \begin{pmatrix} k_1^r - s_1^r + \mathbb{1}\{s_1^r > 0\} & k_2^r \\ k_1^v - s_1^v + \mathbb{1}\{s_1^r = 0\} & k_2^v \end{pmatrix} \tag{5.31}$$

at rate

$$\frac{m_{(k_1 - s_1 + 1, k_2)}}{m_{\boldsymbol{k}}} \binom{k_1^r}{s_1^r} \binom{k_1^v}{s_1^v} \binom{k_1^r - s_1^r}{w_1^r} \binom{k_1^v - s_1^v}{w_1^v} \binom{k_2^r}{w_2^r} \binom{k_2^v}{w_2^v} I_1^\nu(s_1, w_1, w_2)$$

where $(s_1, w_1, w_2) \in \mathcal{C}_{1,\boldsymbol{k}}$, $s_1^r \in [s_1]_0, w_1^r \in [w_1]_0$ and $w_2^r \in [w_2]_0$, and the transition (reproduction with a_2-parent)

$$\begin{pmatrix} k_1^r & k_2^r \\ k_1^v & k_2^v \end{pmatrix} \longrightarrow \begin{pmatrix} k_1^r & k_2^r - s_2^r - w_2^r + \mathbb{1}\{s_2^r + w_2^r > 0\} \\ k_1^v + i^r + i^v & k_2^v - s_2^v + w_2^r - i^r - i^v + \mathbb{1}\{s_2^r = w_2^r = 0\} \end{pmatrix} \tag{5.32}$$

at rate

$$\frac{m_{(k_1 + i, k_2 - s_2 - i + 1)}}{m_{\boldsymbol{k}}} \binom{k_2^r}{s_2^r} \binom{k_2^v}{s_2^v} \binom{k_2^r - s_2^r}{w_2^r} \binom{k_2^v - s_2^v}{w_2^v} \binom{w_2^r}{i^r} \binom{w_2^v}{i^v} I_2^\nu(s_2, w_2),$$

where $(s_2, w_2) \in \mathcal{C}_{2,\boldsymbol{k}}$, $s_2^r \in [s_2]_0, w_2^r \in [w_2]_0, i^r \in [w_2^r]_0, i^v \in [w_2^v]_0$ and $i = i^r + i^v$. Finally, due to mutation, the state K could transition to

$$\begin{pmatrix} k_1^r - 1 & k_2^r + 1 \\ k_1^v & k_2^v \end{pmatrix}, \quad \begin{pmatrix} k_1^r & k_2^r \\ k_1^v - 1 & k_2^v + 1 \end{pmatrix}, \quad \begin{pmatrix} k_1^r + 1 & k_2^r - 1 \\ k_1^v & k_2^v \end{pmatrix}, \quad \begin{pmatrix} k_1^r & k_2^r \\ k_1^v + 1 & k_2^v - 1 \end{pmatrix}$$

at rates

$$\frac{m_{\boldsymbol{k} - \boldsymbol{e}_1 + \boldsymbol{e}_2}}{m_{\boldsymbol{k}}} k_1^r \nu_2, \quad \frac{m_{\boldsymbol{k} - \boldsymbol{e}_1 + \boldsymbol{e}_2}}{m_{\boldsymbol{k}}} k_1^v \nu_2, \quad \frac{m_{\boldsymbol{k} + \boldsymbol{e}_1 - \boldsymbol{e}_2}}{m_{\boldsymbol{k}}} k_2^r \nu_1, \quad \frac{m_{\boldsymbol{k} + \boldsymbol{e}_1 - \boldsymbol{e}_2}}{m_{\boldsymbol{k}}} k_2^v \nu_1$$

respectively.

Corollary 5.12 (Existence of silent events)**.** *The differential killing and birth models both possess silent events in their genealogies with positive probability. For differential killing, they only occur when the parent of the reproduction event is of type a_1. For differential birth on the other hand, they may only occur when the parent of the event is of type a_2.*

Proof. By checking transition (5.29) for differential killing in the special case $s_1 = s_1^v = 1$, i.e. $s_1^r = 0$, and $w_1^r \geq 2$, we see that this yields a genuine change of state in the real genealogy. However, for the corresponding term $\mathscr{L}_1$ of Theorem 5.3, a circle configuration for which $s_1 = 1$ does not generate a change of state. This is an example of a silent event when the parent is of type a_1.

We also need to show that they cannot occur when the parent is of type a_2. In $\mathscr{L}_2$ of Theorem 5.3, the conditions for no change are $i = 0$ (which implies that $i^r = i^v = 0$) and $s_2 = 1$. We need to show that when these conditions are met, there are also no changes in the real genealogy (that is, all the entries in (5.30) remain the same). There are two cases: (a) $s_2^r = 1$ with $s_2^v = 0$, and (b) $s_2^r = 0$ with $s_2^v = 1$. In the first case, (a), the parent is real and cancels with s_2^r in the top-right entry; the bottom-right entry evaluates to k_2^v since $s_2^v = 0 = i^v$ and the indicator condition fails. In the second case, (b), $s_2^v = 1$ cancels with the successful indicator in the bottom-right entry, and everything besides k_2^r equals zero in the top-right. Hence there are no silent events.

For the differential birth model, we essentially have to prove the opposite. We hope to find a genuine transition in (5.32) subject to the conditions $i = 0$ and $s_2 = 1$. This can be achieved by letting $s_2^r = 1$, $s_2^v = 0$ and $w_2^r > 0$. Hence, silent events exist in the differential birth model where they do not in the differential killing counterpart.

Finally, we would like to prove that no silent events are possible in (5.31). In $\mathscr{L}_1$ of Theorem 5.3, we require that $s_1 = 1$ for there to be no change. It is not difficult to then check that the s_1^r or s_1^v cancels the appropriate indicator function in the top-left and top-right entries of (5.31) to leave everything unchanged. □

Proof of Theorem 5.10. Having already explained the dynamics of the real genealogy using prose (beginning p.144), it just remains to apply it to the details of the transitions. In addition, we have already outlined where all the factors in the rates of transitions of reproductive events come from, so we shall not repeat that here.

We shall start with the transition (5.29), supposing that there is an a_1-parent. Recall that continuing lineages retain their reality if of type a_2, but become virtual if of the weaker type. Hence there is a transfer of w_1^r real a_1-lineages to the virtual state. The s_1^r real and s_1^v virtual a_1-lineages that were struck

by strong circles coalesce into the parental lineage. The parental lineage is real if a strong circle hit a real lineage of either type, or a weak circle hit a real a_1-lineage, i.e. if $s_1^r + w_1^r > 0$; it is virtual if $s_1^r + w_1^r = 0$. There is no change in the number of real or virtual a_2-lineages in this scenario, as they are robust to weak circles and the parent is of type a_1.

For transitions of type (5.30), we have that $w_1 = 0$, and so weak circles only fall on a_2-lineages, which retain their reality. There is a loss of s_2^r real and s_2^v virtual a_2-lineages through coalescence. There may be a transfer of i^r real and i^v virtual a_2-lineages to $i^r + i^v$ virtual a_1-lineages. Since $w_1 = 0$, the parental lineage can only be real if a strong circle struck a real lineage, which happens if $s_2^r > 0$; otherwise, it is virtual.

If a mutation occurs, it affects exactly one of the lineages, which swaps its type but not its reality. □

Proof of Theorem 5.11. As in the proof of Theorem 5.10, we shall just explain the nature of the transitions that the real genealogy makes, starting with (5.31). As the parent of this kind of event is of type a_1, the continuing lineages all retain their reality. As usual, there is a loss of s_1^r real and s_1^v virtual a_1-lineages through coalescence. The parent can only be real if a strong circle struck a real lineage, i.e. if $s_1^r > 0$, otherwise it is virtual.

In (5.32), since the parent is of type a_2, all the continuing lineages will become virtual if they were not before. Since $w_1 = 0$, only the w_2^r real a_2-lineages are transferred to the virtual set, with $w_2^r - i^r$ remaining as type a_2, and i^r becoming of type a_1. There are also i^v lineages which switch to type a_1, meaning that, overall, there is an increase of $i^r + i^v$ virtual a_1-lineages. There is a loss of s_2^r real and s_2^v virtual a_2-lineages through coalescence. Finally, the parent is real if *any* circle struck a real lineage. Hence, it is virtual if $s_2^r = w_2^r = 0$.

If a mutation occurs, it affects exactly one of the lineages, which swaps its type but not its reality. □

Remark 5.13. In principle, simulation of the real genealogies of these processes, at least up to the matrix summary, is now possible. This is enough to probe the genealogies for basic shape information, such as the time to MRCA. It would be very interesting indeed to ask if these times differ in distribution between the models.

In order to be able to simulate, one would need to calculate the moments of the stationary distribution of the ΛFVD∗m process. Using (4.17), it is not difficult to derive the non-linear difference equation

$$\left(\int_0^1 \left[(1-ru)^k - (1-u)^k\right]\nu(du)\right) m_{(k+1,0)} = \left(\int_0^1 \left[1-(1-u)^k\right]\nu(du) + k(\nu_1+\nu_2)\right) m_{(k,0)} - \sum_{s=1}^{k}\binom{k}{s}\int_0^1 \left[(ru)^s(1-ru)^{k-s}\right]\nu(du) m_{(k-s+1,0)} - k\nu_2 m_{(k-1,0)}.$$

which the moments satisfy. This is valid for $k \in \mathbb{N}$ with the convention that $m_{(0,0)} = 1$. However, this system is not closed, and therefore cannot be used recursively. There may be hope to solve it in very special cases, or by linearizing when $r \approx 1$ (weak selection), but it may be better to simulate the moments using Monte Carlo methods for cases of interest.

Another practical point to be wary of when attempting to simulate the process is that there are some 'false' events in Theorems 5.10 and 5.11 which do not change the state of the process. For completeness, we shall list them here. For differential killing, they occur in (5.29) when $s_1 = 1$ and $s_1^r + w_1^r \leq 1$ and, in (5.30), when $s_2 = 1$ and $i = 0$. For differential birth, they occur in (5.31) when $s_1 = 1$ and, in (5.32), when $s_2 = 1$, $i^r = i^v = 0$ and $s_2^r + w_2^r \leq 1$.

5.6 The Differential Middle

The existence of silent events in the real genealogy of the models we have studied so far has the consequence that the simple genealogy is not representative of the true genealogical structure of the sample, as there are some coalescence events that go undetected. This prompts the question of whether there is an adaptation of the selective model whose forward-in-time allelic frequencies are unchanged, and whose genealogy possesses no silent events. The answer is in fact yes, as we shall demonstrate now.

If we study silent events in more detail, we see that they occur when the parent displaces individuals of its own type at weak circles. Looking back at Table 5.1, or forward to Table 5.3, we observe that the differential killing model does this in the first row, and differential birth in the seventh. Since this is

			DK	DM	DB	DK	DM	DB
Parental type	Target type	Symbol	Displaced?			Resultant type		
a_1	a_1	○	✓	×	×	a_1	a_1	a_1
a_1	a_1	●	✓	✓	✓	a_1	a_1	a_1
a_1	a_2	○	×	×	×	a_2	a_2	a_2
a_1	a_2	●	✓	✓	✓	a_1	a_1	a_1
a_2	a_1	○	✓	✓	✓	a_2	a_2	a_2
a_2	a_1	●	✓	✓	✓	a_2	a_2	a_2
a_2	a_2	○	×	×	✓	a_2	a_2	a_2
a_2	a_2	●	✓	✓	✓	a_2	a_2	a_2
N/A	a_1	×	N/A			a_2	a_2	a_2
N/A	a_2	×	N/A			a_1	a_1	a_1

Table 5.3: The possible events that may cause a label change in the ΛFVD$*$m models. The type of the targeted individual line after the event is listed (resultant type), along with whether a new offspring was born as a result of the event (displaced?). The differential middle has been included in this extended table

exactly where the models disagree, we can conclude that the silent events drive the differences between the models.

In order to create a third model, we can introduce the simple but perhaps arbitrary rule that, in addition to the ordinary rules, a parent may not displace individuals of its own type at weak circles. This creates a model which interpolates between the two previous ones, and so we call it the differential middle (DM). Table 5.3 shows the core features of the model and compares it to the others. Clearly, the new rule will not affect the allelic frequencies, and so these are consistent amongst all three models. The differential middle has the additional feature that there are no silent events, which means its simple genealogy encapsulates the true genealogical structure of a sample, even if it does not contain the same level of detail as the real genealogy.

One property of the genealogy of the differential middle is that its time to MRCA is *a.s.* at least as long as that of the genealogies of the differential killing and birth models, when coupled appropriately. This is because its genealogy omits the silent events that cause coalescences in the other models. However, the relationship between the time to MRCA of the original two models remains to be established.

Appendix A: Notation

$\mathbb{N}$ is the set of natural numbers ($\mathbb{N}_0$ with zero included).

$\mathbb{Z}$ is the set of integers.

$\mathbb{R}$ is the set of real numbers ($\mathbb{R}_+$ for positive real numbers).

$\mathbb{P}$ is a probability measure (sometimes $\mathbb{P}^X$ to emphasize its relation to X).

$\mathbb{E}$ is an expectation (sometimes $\mathbb{E}^X$ to emphasize its relation to X).

$\mathbb{E}_x[X_t] := \mathbb{E}[X_t|X_0 = x]$ is a conditional expectation for a random process X_t with initial value $X_0 = x$ (sometimes $\mathbb{E}_x^X$).

$\mathbb{1}$ is the indicator function.

$[n] := \{1, 2, \ldots, n\}$, $[n]_0 := [n] \cup \{0\}$, $n \in \mathbb{N}$.

$!e_i := e_1\mathbb{1}\{e_i = e_2\} + e_2\mathbb{1}\{e_i = e_1\}$ is the negation operator for a two-point set $\{e_1, e_2\}$. That is, $!e_1 = e_2$ and $!e_2 = e_1$.

Vectors will be denoted in bold face, e.g. $\boldsymbol{k}$, and sometimes underlined for emphasis, e.g. $\underline{\boldsymbol{C}}$.

$\boldsymbol{e}_i$ is the unit vector, with zeros in every entry except for a 1 in the ith position.

$\mathbf{0}$ is the zero vector, with zeros in every entry.

$|\boldsymbol{k}| := \sum_{i=1}^{d} k_i$ is the L^1-norm of the d-dimensional vector $\boldsymbol{k} = (k_1, k_2, \ldots, k_d)$.

$|\boldsymbol{k}|_w := \sum_{i=1}^{d} ik_i$ is the weighted norm of the d-dimensional vector $\boldsymbol{k} = (k_1, k_2, \ldots, k_d)$.

$\boldsymbol{k} > \mathbf{0}$ means that each vector entry satisfies $k_i \geq 0$ with the restriction that $|\boldsymbol{k}| > 0$.

$\boldsymbol{k} \geq \boldsymbol{l}$ means that each vector entry satisfies $k_i \geq l_i$ with no further restrictions.

$z^{\underline{k}} := z(z-1) \cdot \ldots \cdot (z-k+1)$ is the kth falling factorial power of z.

$z^{\overline{k}} := z(z+1) \cdot \ldots \cdot (z+k-1)$ is the kth rising factorial power of z.

$\Gamma(z)$ is the Gamma function.

$B(x, y)$ is the Beta function.

$\Lambda_{a,b}(dx) = B(a,b)^{-1} x^{a-1}(1-x)^{b-1} dx$ is the two-parameter Beta measure, supported on $(0, 1)$.

$\Lambda^{\alpha}(dx) = \Lambda_{2-\alpha,\alpha}(dx)$, $\alpha \in (0, 2)$, is the measure for the Beta coalescent.

$w_\bullet$ is shorthand for the sequence $\{w_k\}_{k \in \mathbb{N}}$ of non-negative integers.

$B_n(v_\bullet, w_\bullet)$ is the nth complete Bell polynomial with sequences $v_\bullet, w_\bullet$.

$B_{n,k}(v_\bullet)$ is the (n,k)th partial Bell polynomial with sequence $w_\bullet$.

$f \sim g$ is equivalent to $\lim_{n\to\infty} \frac{f(n)}{g(n)} = 1$.

$\stackrel{d}{=}$ is equality in distribution.

$||f||_\infty$ is the supremum norm of the function f.

$\mathcal{U}(S)$ is the uniform distribution on the set S.

$\mathscr{L}$ is a linear operator, often a generator of a Markov process. It will include a subscript or superscript for more specificity, e.g. $\mathscr{L}_{WF}$ for the generator of the Wright–Fisher diffusion.

$\mathscr{P}_A$ is the set of partitions of the set A ($\mathscr{P}_{A,k}$ is the set of partitions of A into k blocks).

Appendix B: Preliminaries

B.1 Bell Polynomials

In this section, we shall give a brief introduction to Bell polynomials. For more details, see Chapter 1 of Pitman [2006].

Definition B.1.1 (Partial Bell polynomial)**.** Let $k, n \in \mathbb{N}$ be such that $k \leq n$. Further, let $\pi \in \mathscr{P}_{[n],k}$ be a partition of $[n]$ into k blocks (that is, $|\pi| = k$). Let $B \in \pi$ be a block in this partition, which has size $|B|$. For a sequence $w_\bullet = \{w_k\}_{k\in\mathbb{N}}$ of non-negative integers,

$$B_{n,k}\left(w_\bullet\right) := \sum_{\pi\in\mathcal{P}_{[n],k}} \prod_{B\in\pi} w_{|B|}$$

is the (n,k)th *partial Bell polynomial*.

Example B.1.2. The partial Bell polynomial of the sequence of factorials is given explicitly by

$$B_{n,k}(\bullet!) = \binom{n-1}{k-1}\frac{n!}{k!} =: L_{n,k}.$$

These are known as the *Lah numbers*.

Remark B.1.3. The coefficient in front of the term $\prod_{j=1}^{n} w_j^{m_j}$ in $B_{n,k}(w_\bullet)$ is the number of partitions $\pi \in \mathscr{P}_{[n],k}$ such that m_j parts equal $j \in [n]$. By Equation (1.8) of Pitman [2006], this number is given by

$$\left[\prod_j w_j^{m_j}\right] B_{n,k}(w_\bullet) = \frac{n!}{\prod_j (j!)^{m_j} m_j!}.$$

Definition B.1.4 (Complete Bell polynomial)**.** For $n \in \mathbb{N}$ and two sequences $v_\bullet = \{v_k\}_{k\in\mathbb{N}}$, $w_\bullet = \{w_k\}_{k\in\mathbb{N}}$ of non-negative integers,

$$B_n(v_\bullet, w_\bullet) := \sum_{k=1}^{n} v_k B_{n,k}(w_\bullet)$$

is the nth *complete Bell polynomial*.

Suppose that the sequences of the complete Bell polynomial have generating functions

$$v(\theta) := \sum_{k=1}^{\infty} \frac{\theta^k}{k!} v_k, \quad w(\eta) := \sum_{k=1}^{\infty} \frac{\eta^k}{k!} w_k.$$

Then, the formula for the composition of their generating functions states that the generating function for B_n is given by

$$\sum_{k=1}^{\infty} \frac{x^k}{k!} B_k(v_\bullet, w_\bullet) = v(w(x)).$$

This is closely linked to the *Faà di Bruno formula* for the derivatives of compositions. For $n \in \mathbb{N}$,

$$\frac{d^n}{dx^n} f(g(x)) = \sum_{\pi \in \mathscr{P}_{[n]}} f^{(|\pi|)}(g(x)) \prod_{B\in\pi} g^{(|B|)}(x).$$

Definition B.1.5 (Polynomials of binomial type)**.** Given a sequence of non-negative integers $w_\bullet$, the sequence of polynomials $B_n(x) := B_n(x^\bullet, w_\bullet)$ (with $B_0 = 1$) is of *binomial type* if it satisfies

$$\sum_{j=0}^{n} \binom{n}{j} B_j(x) B_{n-j}(y) = B_n(x+y).$$

Example B.1.6. The rising and falling factorials, $x^{\overline{n}}$ and $x^{\underline{n}}$ defined in Appendix A, are known to be of binomial type.

Appendix C: Technical Lemmas

This appendix will record some of the more technical results that are subsidiary to the main themes of this thesis. The results are partitioned into their relevant chapters.

C.1 Chapter 1

There are no technical lemmas for this chapter.

C.2 Chapter 2

Lemma C.2.1. *Let $X \stackrel{d}{=} Beta(\alpha, \beta)$, where $\alpha, \beta > 0$. Then, for $k_1, k_2 \geq 0$,*

$$\mathbb{E}\left[X^{k_1}(1-X)^{k_2}\right] = \frac{\alpha^{\overline{k_1}}\beta^{\overline{k_2}}}{(\alpha+\beta)^{\overline{k_1+k_2}}}.$$

Proof. Firstly,

$$\begin{aligned}\mathbb{E}\left[X^{k_1}(1-X)^{k_2}\right] &= \int_0^1 x^{k_1}(1-x)^{k_2}\Lambda_{\alpha,\beta}(dx) \\ &= \frac{B(k_1+\alpha, k_2+\beta)}{B(\alpha,\beta)}\left(\frac{1}{B(k_1+\alpha, k_2+\beta)}\int_0^1 x^{k_1+\alpha-1}(1-x)^{k_2+\beta-1}dx\right),\end{aligned}$$

where $B(\alpha,\beta)$ is the Beta function and $\Lambda_{\alpha,\beta}$ is the Beta measure defined in Appendix A. The quantity within the brackets is precisely $\int_0^1 \Lambda_{k_1+\alpha,k_2+\beta}(dx) = 1$. Hence, converting the remaining ratio into gamma functions,

$$\mathbb{E}\left[X^{k_1}(1-X)^{k_2}\right] = \frac{\Gamma(k_1+\alpha)}{\Gamma(\alpha)}\frac{\Gamma(k_2+\beta)}{\Gamma(\beta)}\frac{\Gamma(\alpha+\beta)}{\Gamma(k_1+k_2+\alpha+\beta)} = \frac{(\alpha)^{\overline{k_1}}\,(\beta)^{\overline{k_2}}}{(\alpha+\beta)^{\overline{k_1+k_2}}}. \qquad \square$$

C.3 Chapter 3

Lemma C.3.1. *For $a,b \in \mathbb{Z}_+$ and $0 < \zeta \leq n$,*

$$\zeta^{\underline{a}}(n-\zeta)^{\underline{b}} = \sum_{k=0}^{b}(-1)^k\,\zeta^{\underline{a+k}}\binom{b}{k}\frac{(n-(a+k))!}{(n-(a+b))!}.$$

Proof. We will use a method similar in flavour to that of a result in the appendix of Etheridge *et al.* [2010], though a little more work is needed. Firstly, considering the generating functions

$$(1+t)^{\zeta} = \sum_{a=0}^{\zeta}\zeta^{\underline{a}}\frac{t^a}{a!}, \qquad (1+s)^{n-\zeta} = \sum_{b=0}^{n-\zeta}(n-\zeta)^{\underline{b}}\frac{s^b}{b!},$$

we see that the desired expression will be the coefficient of ${}^{t^a s^b}/_{a!b!}$ in the product generating function. We can rewrite this function and perform some binomial expansions to obtain

$$\begin{aligned}(1+t)^{\zeta}(1+s)^{n-\zeta} &= (1+s)^n\left(1+\frac{t-s}{1+s}\right)^{\zeta} = \sum_{l=0}^{\zeta}\binom{\zeta}{l}(t-s)^l(1+s)^{n-l}\\ &= \sum_{l=0}^{\zeta}\sum_{a=0}^{l}\sum_{b=0}^{n-l}\binom{\zeta}{l}\binom{l}{a}\binom{n-l}{b}(-1)^{l-a}\,t^a\,s^{b+l-a}\\ &= \sum_{a=0}^{\zeta}\sum_{b=0}^{n-a}\frac{t^a s^b}{a!b!}\left\{\sum_{l=a}^{a+b}\binom{\zeta}{l}\binom{l}{a}\binom{n-l}{b-(l-a)}(-1)^{l-a}\,a!b!\right\},\end{aligned}$$

where we have swapped the order of summation and shifted the index b in the last equality. The coefficient we require is the sum inside the braces, and after manipulating the binomial coefficients and shifting the index l, the result follows. $\square$

Lemma C.3.2. *Let $n \in \mathbb{N}$, $z, z' \in \mathbb{Z}$ and $a, b \in \mathbb{R} \setminus \{-\mathbb{N}\}$. Then, as $n \to \infty$,*

$$\frac{a^{\overline{n+z}}}{b^{\overline{n+z'}}} \sim \frac{\Gamma(b)}{\Gamma(a)} n^{a+z-b-z'},$$

where the meaning of $f \sim g$ is defined in Appendix A.

Proof. Using properties of the rising factorial,

$$\frac{a^{\overline{n+z}}}{b^{\overline{n+z'}}} = \frac{\Gamma(a+n+z)}{\Gamma(a)} \frac{\Gamma(b)}{\Gamma(b+n+z')} \sim \frac{\Gamma(b)}{\Gamma(a)} n^{(a+z)-(b+z')},$$

using the approximation

$$\frac{\Gamma(n+c)}{\Gamma(n)n^c} \sim 1$$

for $c \in \mathbb{R}$ as $n \to \infty$. □

Before proceeding with the next lemmas, recall equation (3.31), a definition that we shall restate here, along with another. For $k \in \mathbb{N}_0$,

$$\gamma_n^{(k)} := \sum_{l=2}^{n} \binom{n}{l} (l-1)^k \lambda_{n,l}, \quad \gamma_n^{(\underline{k})} := \sum_{l=2}^{n} \binom{n}{l} l^{\underline{k}} \lambda_{n,l}.$$

Lemma C.3.3. *Let $a \in (0,1)$ and $b > 0$. Then, as $n \to \infty$,*

$$\gamma_n^{(1)} \sim \frac{\Gamma(a+b)}{(a-1)(a-2)\Gamma(b)} n^{2-a}.$$

Proof. Firstly, note the relation

$$\gamma_n^{(1)} = \gamma_n^{(\underline{1})} - \gamma_n^{(0)}.$$

Then, using Lemma C.3.5 with $d = k = 1$, we obtain

$$\gamma_n^{(\underline{1})} = \frac{n}{a-1} \left(\frac{(a+b-1)^{\overline{n-1}}}{(a+b)^{\overline{n-2}}} - \frac{b^{\overline{n-1}}}{(a+b)^{\overline{n-2}}} \right) =: \frac{n}{a-1} \left(R_n^{neg} - R_n^{dom} \right).$$

The first ratio will turn out to be negligible, but we have to consider cases to show its asymptotic behaviour. Firstly, suppose that $a + b \neq 1$, in which case, using Lemma C.3.2, we see that

$$R_n^{neg} = \frac{(a+b-1)^{\overline{n-1}}}{(a+b)^{\overline{n-2}}} \sim \frac{\Gamma(a+b)}{\Gamma(a+b-1)}$$

as $n \to \infty$. On the other hand, if $a + b = 1$, then $R_n^{neg} \equiv 0$ for all $n \geq 2$. Then,

$$R_n^{dom} = \frac{b^{\overline{n-1}}}{(a+b)^{\overline{n-2}}} \sim \frac{\Gamma(a+b)}{\Gamma(b)} n^{1-a},$$

which dominates the other ratio as $n \to \infty$, since $a < 1$. Hence, putting this together, the leading-order behaviour of this quantity is

$$\gamma_n^{(1)} \sim \frac{\Gamma(a+b)}{(1-a)\Gamma(b)} n^{2-a}.$$

To calculate the asymptotics of $\gamma_n^{(0)}$, we use that the rising factorial is a polynomial of binomial type (see Appendix B), yielding

$$\begin{aligned}
\gamma_n^{(0)} = \sum_{l=2}^{n} \binom{n}{l} \lambda_{n,l} &= \frac{1}{(a+b)^{\overline{n-2}}} \sum_{l=2}^{n} \binom{n}{l} a^{\overline{l-2}} b^{\overline{n-l}} \\
&= \frac{1}{(a-1)(a-2)(a+b)^{\overline{n-2}}} \sum_{l=2}^{n} \binom{n}{l} (a-2)^{\overline{l}} b^{\overline{n-l}} \\
&= \frac{1}{(a-1)(a-2)} \left(\frac{(a+b-2)^{\overline{n}}}{(a+b)^{\overline{n-2}}} - \frac{(a-2)nb^{\overline{n-1}} + b^{\overline{n}}}{(a+b)^{\overline{n-2}}} \right) \\
&=: \frac{1}{(a-1)(a-2)} \left(\tilde{R}_n^{neg} - \tilde{R}_n^{dom} \right).
\end{aligned}$$

Once again, we have to perform a case analysis on the first, negligible ratio. Firstly, suppose $a + b \neq 2$, then using Lemma C.3.2 once more,

$$\tilde{R}_n^{neg} = \frac{(a+b-2)^{\overline{n}}}{(a+b)^{\overline{n-2}}} \sim \frac{\Gamma(a+b)}{\Gamma(a+b-2)}$$

as $n \to \infty$. If $a + b = 2$, then $\tilde{R}_n^{neg} \equiv 0$ for all $n \in \mathbb{N}$. The dominant behaviour comes from the other ratio,

$$\tilde{R}_n^{dom} = \frac{(a-2)nb^{\overline{n-1}} + b^{\overline{n}}}{(a+b)^{\overline{n-2}}} \sim \frac{(a-1)\Gamma(a+b)}{\Gamma(b)} n^{2-a} \quad \Rightarrow \quad \gamma_n^{(0)} \sim -\frac{\Gamma(a+b)}{(a-2)\Gamma(b)} n^{2-a},$$

as $n \to \infty$. Once we put this all together, the claim will follow. □

Lemma C.3.4. *As $n \to \infty$ we have that $\gamma_n^{(2)} \sim n^2$, where $\gamma_n^{(k)}$ is defined in (3.31).*

Proof. We can write $\gamma_n^{(2)} = \gamma_n^{(\underline{2})} - \gamma_n^{(1)}$, since $(l-1)^2 = l^{\underline{2}} - (l-1)$. Applying Lemma C.3.5 (with $d = 1$ and $k = 2$), we find that

$$\gamma_n^{(\underline{2})} = \sum_{l=2}^{n} \binom{n}{l} l^{\underline{2}} \frac{a^{\overline{l-2}} b^{\overline{n-l}}}{(a+b)^{\overline{n-2}}} = \frac{1}{(a-2)^{\overline{2}}(a+b)^{\overline{n-2}}} \sum_{l=2}^{n} \binom{n}{l} l^{\underline{2}} (a-2)^{\overline{l}} b^{\overline{n-l}} = n(n-1).$$

From this and Lemma C.3.3 we conclude $\gamma_n^{(2)} = \gamma_n^{(\underline{2})} - \gamma_n^{(1)} \sim n^2$ as $n \to \infty$. □

Lemma C.3.5. *Let $d \in \mathbb{N}$ and $\boldsymbol{k}, \boldsymbol{n} \in \mathbb{N}_0^d$. For $a, b \in \mathbb{R}$,*

$$\sum_{\boldsymbol{l} \in \mathbb{N}_0^d} \boldsymbol{l}^{\underline{\boldsymbol{k}}} a^{\overline{|\boldsymbol{l}|}} b^{\overline{|\boldsymbol{n}|-|\boldsymbol{l}|}} \prod_{i=1}^{d} \binom{n_i}{l_i} = a^{\overline{|\boldsymbol{k}|}} \boldsymbol{n}^{\underline{\boldsymbol{k}}} (a+b+|\boldsymbol{k}|)^{\overline{|\boldsymbol{n}|-|\boldsymbol{k}|}},$$

where $\boldsymbol{n}^{\underline{\boldsymbol{k}}} := \prod_{i=1}^{d} (n_i)^{\underline{k_i}}$.

Proof. Firstly, we have to prove the result for the vector $\boldsymbol{k} = \boldsymbol{0}$ separately. To do so, we shall use induction on d. For $\boldsymbol{k} = \boldsymbol{0}$ and $d = 1$, the statement reads

$$\sum_{l=0}^{n} \binom{n}{l} a^{\overline{l}} b^{\overline{n-l}} = (a+b)^{\overline{n}},$$

which is true as the rising factorial is a polynomial of binomial type (see Appendix B). Now suppose that the statement holds for some $d \in \mathbb{N}$. Then,

$$\begin{aligned}
\sum_{\boldsymbol{l} \in \mathbb{N}_0^{d+1}} \prod_{i=1}^{d+1} \binom{n_i}{l_i} a^{\overline{|\boldsymbol{l}|}} b^{\overline{|\boldsymbol{n}|-|\boldsymbol{l}|}} &= \sum_{\boldsymbol{l} \in \mathbb{N}_0^{d}} \prod_{i=1}^{d} \binom{n_i}{l_i} a^{\overline{|\boldsymbol{l}|}} b^{\overline{n_1+\ldots+n_d-|\boldsymbol{l}|}} \\
&\quad \times \sum_{l_{d+1}=0}^{n_{d+1}} \binom{n_{d+1}}{l_{d+1}} (a+|\boldsymbol{l}|)^{\overline{l_{d+1}}} (b+n_1+\ldots+n_d-|\boldsymbol{l}|)^{\overline{n_{d+1}-l_{d+1}}} \\
&= (a+b)^{\overline{\sum_{i=1}^{d} n_i}} (a+b+n_1+\ldots+n_d)^{\overline{n_{d+1}}} = (a+b)^{\overline{\sum_{i=1}^{d+1} n_i}},
\end{aligned}$$

where we used the induction hypothesis in the second equality. Now suppose that $\boldsymbol{k}$ has at least one positive entry, i.e. $|\boldsymbol{k}| \geq 1$. In the following calculation, in the third equality, we will utilize the identity

$$(m+k)^{\underline{k}} \binom{n}{m+k} = n^{\underline{k}} \binom{n-k}{m}$$

which we state without proof. Although the sum in the statement of the result ranges over vectors $\boldsymbol{l}$ in the space $\mathbb{N}_0^d$, in reality, any vector for which there is an entry satisfying $l_i < k_i$ will make a zero contribution to the sum because of the falling factorial term. Hence, in the first equality, we shall make the substitution $\boldsymbol{m} = \boldsymbol{l} - \boldsymbol{k}$. Upon doing so,

$$\begin{aligned}\sum_{\boldsymbol{l}\in\mathbb{N}_0^d} \boldsymbol{l}^{\underline{\boldsymbol{k}}} a^{\overline{|\boldsymbol{l}|}} b^{\overline{|\boldsymbol{n}|-|\boldsymbol{l}|}} \prod_{i=1}^{d} \binom{n_i}{l_i} &= \sum_{\boldsymbol{m}\in\mathbb{N}_0^d} (\boldsymbol{m}+\boldsymbol{k})^{\underline{\boldsymbol{k}}} a^{\overline{|\boldsymbol{m}|+|\boldsymbol{k}|}} b^{\overline{|\boldsymbol{n}|-|\boldsymbol{k}|-|\boldsymbol{m}|}} \prod_{i=1}^{d} \binom{n_i}{m_i+k_i} \\ &= a^{\overline{|\boldsymbol{k}|}} \sum_{\boldsymbol{m}\in\mathbb{N}_0^d} (a+|\boldsymbol{k}|)^{\overline{|\boldsymbol{m}|}}\, b^{\overline{|\boldsymbol{n}|-|\boldsymbol{k}|-|\boldsymbol{m}|}} \prod_{i=1}^{d} (m_i+k_i)^{\underline{k_i}} \binom{n_i}{m_i+k_i} \\ &= a^{\overline{|\boldsymbol{k}|}} \boldsymbol{n}^{\underline{\boldsymbol{k}}} \sum_{\boldsymbol{m}\in\mathbb{N}_0^d} \prod_{i=1}^{d} \binom{n_i-k_i}{m_i} (a+|\boldsymbol{k}|)^{\overline{|\boldsymbol{m}|}}\, b^{\overline{|\boldsymbol{n}|-|\boldsymbol{k}|-|\boldsymbol{m}|}}.\end{aligned}$$

The result follows by applying the $\boldsymbol{k} = \boldsymbol{0}$ case to the remaining sum. □

C.4 Chapter 4

Lemma C.4.1. *For each $n \in \mathbb{N}$, the generator of the ΛFVM process with the Beta measure Λ^α satisfies the estimate*

$$|\mathscr{L}_{\Lambda M} y^n| \leq \left(1 + \frac{v_1+v_2}{1-\alpha}\right) 2^n, \quad \forall y \in [0,1].$$

Proof. Using the generator decomposition of Lemma 4.3, we observe that

$$|\mathscr{L}_{\Lambda M} y^n| \leq |\mathscr{L}_{\Lambda} y^n| + |\mathscr{L}_{M} y^n|$$

which allows us to consider each part separately. Using (3.8),

$$\begin{aligned}|\mathscr{L}_{\Lambda} y^n| &= \left| \int_0^1 \sum_{k=2}^{n} \binom{n}{k} \frac{u^{k-1-\alpha}(1-u)^{n-k+\alpha-1}}{B(2-\alpha,\alpha)} du (y^{n-k+1} - y^n) \right| \\ &\leq \sum_{k=2}^{n} \binom{n}{k} \frac{1}{B(2-\alpha,\alpha)} \int_0^1 u^{k-\alpha-1}(1-u)^{n-k+\alpha-1} du |y^{n-k+1}(1-y^{k-1})| \\ &\leq \sum_{k=2}^{n} \binom{n}{k} \frac{B(2-\alpha+k-2, \alpha+n-k)}{B(2-\alpha,\alpha)} \leq \sum_{k=2}^{n} \binom{n}{k} \leq 2^n,\end{aligned}$$

where we used the triangle inequalities for the first bound, and that $y^{n-k+1}(1-y^{k-1}) \leq 1$ for all

$y \in [0,1]$ in the second. In the third inequality, we used the fact that the Beta function is *decreasing* in both of its arguments, so that the ratio of Beta functions can be bounded by one. The final inequality follows from a standard combinatorial identity.

For the mutation part,

$$\begin{aligned}
|\mathscr{L}_M y^n| &= \left|(v_2 - (v_1+v_2)y)\int_0^1 [((1-u)y+u)^n - ((1-u)y)^n]\,\frac{\Lambda^\alpha(du)}{u^2}\right| \\
&= |v_2 - (v_1+v_2)y|\left|\int_0^1 \sum_{k=1}^n \binom{n}{k} u^k (1-u)^{n-k} y^{n-k} \frac{u^{1-\alpha}(1-u)^{\alpha-1}}{u^2 B(2-\alpha,\alpha)} du\right| \\
&\le |v_2 - (v_1+v_2)y| \sum_{k=1}^n \binom{n}{k} \frac{|y|^{n-k}}{B(2-\alpha,\alpha)} \int_0^1 u^{k-\alpha-1}(1-u)^{n-k+\alpha-1} du \\
&\le |v_2 - (v_1+v_2)y| \sum_{k=1}^n \binom{n}{k} \frac{B(k-\alpha, n-k+\alpha)}{B(2-\alpha,\alpha)} \\
&= |v_2 - (v_1+v_2)y| \sum_{k=1}^n \binom{n}{k} \frac{B(1-\alpha+k-1, \alpha+n-k)}{(1-\alpha)B(1-\alpha,\alpha)} \\
&\le \frac{|v_2 - (v_1+v_2)y|}{1-\alpha} \sum_{k=1}^n \binom{n}{k} \le \left(\frac{v_1+v_2}{1-\alpha}\right) 2^n.
\end{aligned}$$

Combining these two inequalities completes the result. □

Lemma C.4.2. *If $\alpha \in (0,1)$ and $\Lambda^\alpha(du)$ is the Beta probability measure defined in Appendix A, then*

$$\int_0^1 \frac{\Lambda^\alpha(du)}{u} = \frac{1}{1-\alpha}.$$

Moreover, it is not integrable for $\alpha \notin (0,1)$.

Proof. Set

$$I_{\epsilon_0,\epsilon_1} = \int_{\epsilon_0}^{\epsilon_1} \frac{\Lambda^\alpha(du)}{u} \propto \int_{\epsilon_0}^{\epsilon_1} u^{-\alpha}(1-u)^{\alpha-1} du.$$

Since the behaviour of $I_{\epsilon_0,\epsilon_1}$ near $u = 0$ is governed by $\int_{\epsilon_0}^{\epsilon_1} u^{-\alpha} du$, it is clear that I_{0,ϵ_1} is integrable for $\alpha < 1$, but non-integrable for $\alpha \ge 1$. Similarly, near $u = 1$, $I_{\epsilon_0,1}$ is integrable for $\alpha > 0$ and non-integrable otherwise. Putting this together, we find that $I_{0,1}$ is integrable only for $\alpha \in (0,1)$.

We now we concentrate on the case $\alpha \in (0,1)$. It remains to show that

$$\begin{aligned}\int_0^1 \frac{\Lambda^\alpha(du)}{u} &= \frac{\Gamma(2)}{\Gamma(2-\alpha)\Gamma(\alpha)} \int_0^1 u^{(1-\alpha)-1}(1-u)^{\alpha-1}du \\ &= \frac{B(1-\alpha,\alpha)}{\Gamma(2-\alpha)\Gamma(\alpha)} \left\{ B(1-\alpha,\alpha)^{-1} \int_0^1 u^{(1-\alpha)-1}(1-u)^{\alpha-1}du \right\} \\ &= \frac{\Gamma(1-\alpha)\Gamma(\alpha)}{\Gamma(2-\alpha)\Gamma(\alpha)} = \frac{1}{1-\alpha},\end{aligned}$$

where we have used the fact that $\Gamma(z+1) = z\Gamma(z)$. The quantity inside the braces on the second line is $\int_0^1 \Lambda_{1-\alpha,\alpha}(du) = 1$. □

Lemma C.4.3. *Suppose $\alpha \neq 1$. Then*

$$S := \sum_{k=0}^{\infty} \frac{(-z)^k}{(k+1)!} (2-\alpha)^{\overline{k}} = \frac{(1+z)^{\alpha-1}-1}{z(\alpha-1)},$$

holds with radius of convergence 1.

Proof. We shall make use of the following identity,

$$\binom{k-\alpha+1}{k} = \binom{k-\alpha}{k}\left(1+\frac{k}{1-\alpha}\right) = (-1)^k \frac{k+1-\alpha}{1-\alpha}\binom{\alpha-1}{k}.$$

which can be proved using various standard results on binomial coefficients. Using this expression, along with the binomial expansion $(1+z)^{\alpha-1} = \sum_{k=0}^{\infty}\binom{\alpha-1}{k}z^k$, which has radius of convergence 1,

$$S = \frac{(1+z)^{\alpha-1}}{1-\alpha} - \frac{\alpha}{z(1-\alpha)}\tilde{S} \qquad \text{where} \qquad \tilde{S} = \sum_{k=0}^{\infty}\binom{\alpha-1}{k}\frac{z^{k+1}}{k+1}.$$

We can differentiate $\tilde{S}$ term-by-term, which yields the differential equation

$$f'(z) = \sum_{k=0}^{\infty}\binom{\alpha-1}{k}\frac{(k+1)z^k}{k+1} = (1+z)^{\alpha-1}.$$

This is readily solved to give $\tilde{S} = \int^z f'(u)du = \frac{1}{\alpha}(1+z)^\alpha + C$. Using the fact that $\tilde{S} = 0$ when $z = 0$ solves for the constant $C = -\frac{1}{\alpha}$. Finally,

$$S = \frac{(1+z)^{\alpha-1}}{1-\alpha} - \frac{(1+z)^\alpha}{z(1-\alpha)} + \frac{1}{z(1-\alpha)} = \frac{(1+z)^{\alpha-1}-1}{z(\alpha-1)}.$$ □

Lemma C.4.4. *Let Λ^α be the Beta probability measure defined in Appendix A. For $\alpha \in (0,1) \cup (1,2)$ and $a, b \in \mathbb{R} \setminus \{0\}$ such that $|a| < |b|$,*

$$\int_0^1 \frac{\Lambda^\alpha(du)}{au+b} = \frac{(a+b)^{\alpha-1} - b^{\alpha-1}}{ab^{\alpha-1}(\alpha-1)}.$$

Proof. Without loss of generality, we shall set $b = 1$ for the remainder of this proof. Define the function

$$f(u,k) = (-1)^k (au)^k \frac{u^{1-\alpha}(1-u)^{\alpha-1}}{B(2-\alpha,\alpha)}, \quad u \in (0,1),\ k \in \mathbb{Z}.$$

Next, define the integral

$$\begin{aligned} I^+ &:= \int_0^1 \sum_{k=0}^\infty |f(u,k)| du = \sum_{k=0}^\infty \frac{|a|^k}{B(2-\alpha,\alpha)} \int_0^1 u^{k+1-\alpha}(1-u)^{\alpha-1} du \\ &= \sum_{k=0}^\infty |a|^k \frac{B(2+k-\alpha,\alpha)}{B(2-\alpha,\alpha)} \left\{ \frac{1}{B(2+k-\alpha,\alpha)} \int_0^1 u^{(k+2-\alpha)-1}(1-u)^{\alpha-1} du \right\} \\ &\leq \sum_{k=0}^\infty |a|^k < \infty. \end{aligned}$$

In the first line, we made use of Fubini's theorem, which allows the swapping of the integral and summation since $|f(u,k)|$ is clearly a non-negative function. In the second line, the quantity in braces is equal to $\int_0^1 \Lambda_{k+2-\alpha,\alpha}(du)$ and therefore cancels to unity. In the third line, we bounded the ratio of Beta functions term-wise by one since the Beta function is decreasing in its first argument. The remaining sum clearly converges as it is a geometric series with $|a| < 1$.

Then, making use of a binomial series expansion (in the denominator of the following integral), we find that

$$I := \int_0^1 \frac{\Lambda_{2-\alpha,\alpha}(du)}{au+1} = \int_0^1 \sum_{k=0}^\infty f(u,k) du = \sum_{k=0}^\infty \int_0^1 f(u,k) du = \sum_{k=0}^\infty (-a)^k \frac{B(2+k-\alpha,\alpha)}{B(2-\alpha,\alpha)},$$

where we have applied Fubini's theorem once more (using the fact that $I^+ < \infty$) to swap the integral and sum, and to deduce that I is finite. We also used very similar arguments to those above to obtain the sum on the right-hand side. Finally, making use of identities of Gamma functions, we find that

$$I = \sum_{k=0}^\infty (-a)^k \frac{1}{\Gamma(k+2)} \frac{\Gamma(2-\alpha+k)}{\Gamma(2-\alpha)} = \sum_{k=0}^\infty (-a)^k \frac{(2-\alpha)^{\overline{k}}}{(k+1)!},$$

from which we can apply Lemma C.4.3 (with $z = a$, $a \in (-1, 1)$) to close the proof. A completely analogous argument holds when $b \neq 1$. □

Lemma C.4.5. *Let $S(t)$ be the integral transform defined in (4.14). For $k \geq 2$,*

$$S^{(k)}(t) = \frac{(1-\alpha)(p(t)S(t))^{(k-1)} - \sum_{i=0}^{k-2} \binom{k-1}{i} q^{(k-1-i)}(t) S^{(i+1)}(t)}{q(t)}. \tag{C.1}$$

Proof. We proceed is by induction. The base case, $k = 2$, is satisfied by comparison with (4.21). Now, supposing (C.1) holds, and suppressing the dependence on t,

$$\begin{aligned}
S^{(k+1)} &= \frac{(1-\alpha)(p \cdot S)^{(k)}}{q} - \frac{1}{q} \sum_{i=0}^{k-2} \binom{k-1}{i} \left(q^{(k-i)} S^{(i+1)} + q^{(k-i-1)} S^{(i+2)} \right) \\
&\quad - \frac{q'}{q} \left(\frac{(1-\alpha)(p \cdot S)^{(k-1)} - \sum_{i=0}^{k-2} \binom{k-1}{i} q^{(k-1-i)} S^{(i+1)}}{q} \right) \\
&= \frac{1}{q} \left((1-\alpha)(p \cdot S)^{(k)} - \sum_{i=0}^{k-1} \binom{k-1}{i} q^{(k-i)} S^{(i+1)} - \sum_{i=0}^{k-2} \binom{k-1}{i} q^{(k-(i+1))} S^{((i+1)+1)} \right) \\
&= \frac{1}{q} \left((1-\alpha)(p \cdot S)^{(k)} - q^{(k)} S' - \sum_{i=1}^{k-1} \left(\binom{k-1}{i} + \binom{k-1}{i-1} \right) q^{(k-i)} S^{(i+1)} \right) \\
&= \frac{(1-\alpha)(p \cdot S)^{(k)} - \sum_{i=0}^{k-1} \binom{k}{i} q^{(k-i)} S^{(i+1)}}{q}.
\end{aligned} \tag{C.2}$$

In the first equality we used the quotient rule; in the second we again exploited the inductive assumption; in the third we adjusted the index of the second sum; finally, in the fourth, we used the recursive identity of binomial coefficients. □

Following on from the final line of the proof of Lemma C.4.5, we can use the Leibniz rule to expand $(p \cdot S)^{(k)}$. To evaluate the moments, we ultimately need an expression for $S^{(k)}(0)$, and we shall use an argument similar to the one used to derive $S''(0)$ to find it. Since $q(0) = 0$, and the moment is finite, the numerator of $S^{(k+1)}(t)$, evaluated at zero, should vanish too. Suppose it does for the time being. Then, the numerator of the right-hand side of (C.2) satisfies

$$(1-\alpha) \sum_{i=0}^{k} \binom{k}{i} p^{(k-i)}(0) S^{(i)}(0) - \sum_{i=0}^{k-1} \binom{k}{i} q^{(k-i)}(0) S^{(i+1)}(0) = 0$$

if and only if

$$\left(kq'(0) - (1-\alpha)p(0)\right) S^{(k)}(0) = (1-\alpha)\sum_{i=0}^{k-1}\binom{k}{i}p^{(k-i)}(0)S^{(i)}(0) - \sum_{i=0}^{k-2}\binom{k}{i}q^{(k-i)}(0)S^{(i+1)}(0).$$

This motivates the following lemma.

Lemma C.4.6. *Let $S(t)$ be the integral transform defined in (4.14). For $k \geq 2$,*

$$S^{(k)}(0) = \frac{k(1-\alpha)p'(0)S^{(k-1)}(0) + \sum_{i=0}^{k-2}\binom{k}{i}\left((1-\alpha)p^{(k-i)}(0)S^{(i)}(0) - q^{(k-i)}(0)S^{(i+1)}(0)\right)}{kq'(0) - (1-\alpha)p(0)}. \tag{C.3}$$

Proof. We shall again proceed by induction. The base case, $k = 2$, was shown in (4.22). For the inductive step, as we showed above, supposing (C.3) to be true is equivalent to the numerator of $S^{(k+1)}(t)$, evaluated at zero, equalling zero. (An expression for $S^{(k+1)}(t)$ can be found in (C.2).) Therefore, on this supposition, $\lim_{t\downarrow 0} S^{(k+1)}(t)$ is indeterminate, and we apply l'Hôpital's rule. That is,

$$\lim_{t\downarrow 0} S^{(k+1)}(t) = \lim_{t\downarrow 0}\frac{n_{k+1}(t)}{q(t)} = \lim_{t\downarrow 0}\frac{n'_{k+1}(t)}{q'(t)},$$

where

$$\begin{aligned} n_{k+1}(t) = &\left((1-\alpha)p(t) - kq'(t)\right) S^{(k)}(t) + k(1-\alpha)p'(t)S^{(k-1)}(t) \\ &+ \sum_{i=0}^{k-2}\binom{k}{i}\left((1-\alpha)p^{(k-i)}(t)S^{(i)}(t) - q^{(k-i)}(t)S^{(i+1)}(t)\right). \end{aligned}$$

Using the product rule,

$$\begin{aligned} n'_{k+1}(t) = &\left((1-\alpha)p(t) - kq'(t)\right) S^{(k+1)}(t) + \left((k+1)(1-\alpha)p'(t) - kq''(t)\right) S^{(k)}(t) \\ &+ k(1-\alpha)p''(t)S^{(k-1)}(t) + \sum_{i=0}^{k-2}\binom{k}{i}\left((1-\alpha)p^{(k+1-i)}(t)S^{(i)}(t) - q^{(k+1-i)}(t)S^{(i+1)}(t)\right) \\ &+ \sum_{i=0}^{k-2}\binom{k}{i}\left((1-\alpha)p^{(k-(i+1)+1)}(t)S^{(i+1)}(t) - q^{(k-(i+1)+1)}(t)S^{((i+1)+1)}(t)\right) \\ = &\left((1-\alpha)p(t) - kq'(t)\right) S^{(k+1)}(t) + (k+1)(1-\alpha)p'(t)S^{(k)}(t) \\ &+ \sum_{i=0}^{k-1}\left(\binom{k}{i} + \binom{k}{i-1}\right)\left((1-\alpha)p^{(k+1-i)}(t)S^{(i)}(t) - q^{(k+1-i)}(t)S^{(i+1)}(t)\right). \end{aligned}$$

Using the algebra of limits, and the fact that $q'(0) = 1 - \alpha w \neq 0$ (see Lemma C.4.7), $\lim_{t\downarrow 0} S^{(k+1)}(t)$ equals

$$\lim_{t\downarrow 0}\left[\frac{(k+1)(1-\alpha)p'(t)S^{(k)}(t) + \sum_{i=0}^{k-1}\binom{k+1}{i}\left((1-\alpha)p^{(k+1-i)}(t)S^{(i)}(t) - q^{(k+1-i)}(t)S^{(i+1)}(t)\right)}{(k+1)q'(t) - (1-\alpha)p(t)}\right].$$

To prove that we can push the limit inside, it only remains to check that the denominator is non-zero (to ensure that we do not need l'Hôpital's rule again). This follows from the fact that $p(0) = 1$, and $\alpha, v_1, v_2 \in (0,1)$. □

Lemma C.4.7. *For $k \geq 1$, the derivatives of $p(t)$ and $q(t)$, evaluated at zero, satisfy $p(0) = 1$, $q(0) = 0$,*

$$p^{(k)}(0) = (\alpha)^{\underline{k}}\, v_2 \quad \textit{and} \quad q^{(k)}(0) = k\,(\alpha)^{\underline{k-1}} - (\alpha)^{\underline{k}}\, w.$$

The proof of this lemma is omitted as it is a simple exercise in differentiation and induction.

C.5 Chapter 5

Lemma C.5.1. *For all vectors $\boldsymbol{k} \in \mathbb{N}_0^2$ such that $|\boldsymbol{k}| \geq 1$, $\forall r \in (0,1]$ and $\forall u \in (0,1]$,*

$$(1-u)^{|\boldsymbol{k}|} + k_1 ru(1-ru)^{|\boldsymbol{k}|-1} + k_2 u(1-u)^{|\boldsymbol{k}|-1} \leq 1. \tag{C.4}$$

Proof. Fix $k_2 \in \mathbb{N}$ and proceed using induction on k_1. We shall postpone the proof of the base case for a moment, and begin by assuming that (C.4) holds as the inductive hypothesis. Then, using the fact that $1 - u \leq 1 - ru < 1$,

$$\begin{aligned}
&(1-u)^{|\boldsymbol{k}|+1} + (k_1+1)ru(1-ru)^{|\boldsymbol{k}|} + k_2 u(1-u)^{|\boldsymbol{k}|} \\
&\quad\leq (1-ru)\left((1-u)^{|\boldsymbol{k}|} + k_1 ru(1-ru)^{|\boldsymbol{k}|-1} + k_2 u(1-u)^{|\boldsymbol{k}|-1}\right) + ru(1-ru)^{|\boldsymbol{k}|} \\
&\quad\leq 1 - ru + ru(1-ru)^{|\boldsymbol{k}|} \leq 1,
\end{aligned}$$

where the inductive assumption was invoked in the second step. Returning to the base case ($k_1 = 0$), we

need to prove that, for $k_2 \in \mathbb{N}$,

$$k_2 u(1-u)^{k_2-1} + (1-u)^{k_2} \leq 1,$$

which can itself be proved by induction on k_2. This base case ($k_2 = 1$) is trivial and

$$\begin{aligned}(k_2+1)u(1-u)^{k_2} + (1-u)^{k_2+1} &= (1-u)\left(k_2 u(1-u)^{k_2-1} + (1-u)^{k_2}\right) + u(1-u)^{k_2} \\ &\leq 1 - u\left(1-(1-u)^{k_2}\right) \leq 1.\end{aligned}$$

In order to finish the proof, we need to show that it is true when $k_2 = 0$ and $k_1 \in \mathbb{N}$. This can again be done by induction in a very similar manner. □

References

Aldous, D. J. (1999). Deterministic and stochastic models for coalescence (aggregation and coagulation): a review of the mean-field theory for probabilists. *Bernoulli*, pages 3–48. 63, 64

Barbour, A., Ethier, S., and Griffiths, R. (2000). A transition function expansion for a diffusion model with selection. *Ann. Appl. Probab.*, pages 123–162. 10, 14, 15, 23, 95

Barton, N. H., Etheridge, A. M., and Sturm, A. K. (2004). Coalescence in a random background. *Ann. Appl. Probab.*, **14**(2), 754–785. 9

Barton, N. H., Etheridge, A. M., and Véber, A. (2010). A new model for evolution in a spatial continuum. *Electron. J. Probab.*, **15**, 162–216. 2, 49

Barton, N. H., Etheridge, A. M., and Véber, A. (2013). Modelling evolution in a spatial continuum. *Journal of Statistical Mechanics: Theory and Experiment*, **2013**(01), P01002. 11, 50

Berestycki, N. (2009). Recent progress in coaelscent theory. *Ensaios Matemáticos*, **16**. 5, 45, 56

Berestycki, N., Etheridge, A. M., and Hutzenthaler, M. (2009). Survival, extinction and ergodicity in a spatially continuous population model. *Markov Processes and Related Fields*, **15**, 265–288. 47, 49

Bertoin, J. and Le Gall, J.-F. (2003). Stochastic flows associated to a coalescent process. *Prob. Theor. Rel. Fields*, **126**, 261–288. 10, 11, 44, 45

Bhaskar, A., Clark, A. G., and Song, Y. S. (2014). Distortion of genealogical properties when the sample is very large. *Proceedings of the National Academy of Sciences*, **111**(6), 2385–2390. 6, 37

Birkner, M., Blath, J., Capaldo, M., Etheridge, A. M., Möhle, M., Schweinsberg, J., and Wakolbinger, A. (2005). Alpha-stable branching and Beta-coalescents. *Elect. J. Probab.*, **10**, 303–325. 56

Darwin, C. (1859). On the origin of species by means of natural selection, or the preservation of favoured races in the struggle for life. 4

Donnelly, P. J. and Kurtz, T. G. (1999). Particle representations for measure-valued population models. *Ann. Probab.*, **27**, 166–205. 3, 11, 43, 52

Eldon, B. and Wakeley, J. (2006). Coalescent processes when the distribution of offspring number among individuals is highly skewed. *Genetics*, **172**(4), 2621–2633. 10, 44, 56, 57, 77

Etheridge, A. M. (2000). *An introduction to superprocesses*, volume 20 of *University lecture notes*. Amer. Math. Soc. 45

Etheridge, A. M. (2008). Drift, draft and structure: some mathematical models of evolution. *Banach Center Publ.*, **80**, 121–144. 49

Etheridge, A. M. (2011). Some mathematical models from population genetics. In *École d'été de probabilités de Saint Flour XXXXIX-2009*. Springer-Verlag. 7, 29, 37

Etheridge, A. M. and Griffiths, R. (2009). A coalescent dual process in a Moran model with genic selection. *Theor. Pop. Biol.*, **75**, 320–330. 23, 25, 34

Etheridge, A. M. and Kurtz, T. G. (2014). Genealogical constructions of population models. *arXiv preprint arXiv:1402.6724*. 47, 48, 49, 50

Etheridge, A. M., Griffiths, R., and Taylor, J. E. (2010). A coalescent dual process in a Moran model with genic selection, and the lambda coalescent limit. *Theor. Pop. Biol.*, **78**, 77–92. 25, 62, 95, 160

Ethier, S. N. and Griffiths, R. C. (1993). The transition function of a Fleming–Viot process. *Ann. Probab.*, **21**(3), 1571–1590. 7

Ethier, S. N. and Kurtz, T. G. (1986). *Markov processes: characterization and convergence*. Wiley. 8, 19, 20, 21, 23, 24, 46, 87, 88, 89

Feller, W. (1951). Diffusion processes in genetics. In *Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability*, pages 227–246. University of California Press. 30, 31, 35

Felsenstein, J. (1975). A pain in the torus: some difficulties with the model of isolation by distance. *Amer. Nat.*, **109**, 359–368. 49

Gillespie, J. (2000). Genetic drift in an infinite population: the pseudohitchhiking model. *Genetics*, **155**, 909–919. 5, 44

Gillespie, J. H. (1984). The status of the neutral theory. *Science*, **224**(4650), pp. 732–733. 4

Griffiths, R. C. (2006). Coalescent lineage distributions. *Advances in applied probability*, **38**(2), 405–429. 61

Handa, K. (2014). Stationary distributions for a class of generalized Fleming–Viot processes. *Ann. Probab.*, **42**(3), 1257–1284. 15, 78, 87, 89

Hedgecock, D. (1994). Does variance in reproductive success limit effective population sizes of marine organisms? *Genetics and evolution of aquatic organisms*, pages 122–134. 10, 56

Karlin, S. and Taylor, H. M. (1981). *A second course in stochastic processes*. Academic Press. 31

Kimura, M. (1953). Stepping stone model of population. *Ann. Rep. Nat. Inst. Genetics Japan*, **3**, 62–63. 2

Kimura, M. (1985). *The neutral theory of molecular evolution*. Cambridge University Press. 3

Kingman, J. F. C. (1982). The coalescent. *Stoch. Proc. Appl.*, **13**, 235–248. 6, 59, 60

Klebaner, F. C. (2005). *Introduction to Stochastic Calculus with Applications*. Imperial College Press, second edition. 106

Krone, S. M. and Neuhauser, C. (1997). Ancestral processes with selection. *Theor. Pop. Biol.*, **51**, 210–237. 13, 119, 142

Malécot, G. (1948). *Les Mathématiques de l'hérédité*. Masson et Cie, Paris. 6, 8, 49

Moran, P. A. P. (1958). Random processes in genetics. *Proc. Camb. Phil. Soc.*, **54**, 60–71. 13, 36

Neuhauser, C. and Krone, S. M. (1997). Genealogies of samples in models with selection. *Genetics*, **145**, 519–534. 119, 142

Pitman, J. (1999). Coalescents with multiple collisions. *Ann. Probab.*, **27**, 1870–1902. 11, 43, 52

Pitman, J. (2006). *Combinatorial stochastic processes*, volume 1875. Springer-Verlag. 157, 158

Saadi, H. (2014). On a voter model on $\mathbb{R}^d$: cluster growth in the spatial Λ-Fleming–Viot process. *arXiv preprint arXiv:1401.6721*. 52

Sagitov, S. (1999). The general coalescent with asynchronous mergers of ancestral lines. *J. Appl. Probab.*, **26**, 1116–1125. 11, 43, 52

Schweinsberg, J. (2000). Coalescents with simultaneous multiple collisions. *Electron. J. Probab.*, **5**, 1–50. 11

Schweinsberg, J. (2003). Coalescents obtained from supercritical Galton-Watson processes. *Stoch. Proc. Appl.*, **106**, 107–139. 56

Slade, P. F. (2000). Simulation of selected genealogies. *Theor. Pop. Biol.*, **57**(1), 35–49. 108

Smoluchowski, M. (1916). Drei Vorträge über Diffusion, Brownsche Molekularbewegung und Koagulation von Kolloidteilchen. *Z. Phys.*, **17**, 557–585. 14, 58, 63

Steinrücken, M., Birkner, M., and Blath, J. (2013). Analysis of DNA sequence variation within marine species using Beta-coalescents. *Theor. Pop. Biol.*, **87**, 15–24. 44, 57

Stephens, M. and Donnelly, P. (2003). Ancestral inference in population genetics models with selection (with discussion). *Australian & New Zealand Journal of Statistics*, **45**(4), 395–430. 9, 41

Tavaré, S. (1984). Line-of-descent and genealogical processes, and their applications in population genetics models. *Theoretical population biology*, **26**(2), 119–164. 38

Taylor, J. E. (2009). The genealogical consequences of fecundity variance polymorphism. *Genetics*, **182**(3), 813–837. 16, 145

Varadhan, S. S. (2007). *Stochastic processes*, volume 16. American Mathematical Society. 22

Wakeley, J. (2009). *Coalescent theory: an introduction*, volume 1. Roberts & Company Publishers Greenwood Village, Colorado. 3, 7

Wright, S. (1943). Isolation by distance. *Genetics*, **28**, 114–138. 2, 6, 49

Wright, S. (1949). Adaptation and selection. In G. Jepson, G. Simpson, and E. Mayr, editors, *Genetics, Paleontology and Evolution*, pages 365–389. Princeton University Press. 32

Yano, Y. (2006). On the occupation time on the half line of pinned diffusion processes. *Publications of the Research Institute for Mathematical Sciences*, **42**(3), 787–802. 93