Recombination models forward and backward in time

Dissertation

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

vorgelegt an der Fakultät für Mathematik der Universität Bielefeld

eingereicht von

Dipl.-Math. Mareike Esser

am 21. Dezember 2016

Danksagung

An dieser Stelle habe ich die Möglichkeit mich bei einigen Personen zu bedanken, die mir auf verschiedenste Art und Weise geholfen haben diese Arbeit zu schreiben.

Zunächst gilt mein aufrichtiger Dank Ellen Baake. Nicht nur für die außergewöhnlich intensive Betreuung dieser Arbeit, sondern auch für ihre Geduld, ihr Vertrauen und ihren Zuspruch. Ellen Baake hat in mir nicht nur das Interesse für die Biomathematik geweckt, sondern mich durch zahlreiche Impulse und Diskussionen auch maßgeblich darin unterstützt diese Arbeit anzufertigen. Ebenfalls möchte ich Reinhard Bürger für die Bereitschaft der Zweitkorrektur dieser Arbeit danken. Sehr dankbar bin ich auch allen derzeitigen, ehemaligen und assoziierten Mitgliedern der Arbeitsgruppe 'Biomathematik und theoretische Bioinformatik' für die familiäre Arbeitsatmosphäre sowie für die hilfreichen Rückfragen, Diskussionen und Kommentare in unserem Seminar. Insbesondere danke ich Sebastian Probst für die enge Zusammenarbeit in weiten Teilen der Kapitel drei und vier. Weiterhin danke ich Michael Baake für seine wertvollen Hinweise und Anmerkungen. Zu großem Dank bin ich auch Sebastian Hummel, Jakob Landwehr und Fernando Cordero für das Korrekturlesen von weiten Teilen der Arbeit verpflichtet. Martin Dieckmann möchte ich für die geduldige Unterstützung bei mathematischen Angelegenheiten aller Art danken.

Darüber hinaus möchte ich mich bei der Deutschen Forschungsgemeinschaft (Priority Programme SPP 1590 'Probabilistic Structures in Evolution', grant no. BA 2469/5-1) für die finanzielle Förderung bedanken. Der enge Kontakt zu anderen Arbeitsgruppen des Schwerpunktprogrammes im Rahmen von Workshops und Konferenzen erwies sich als perfektes Arbeitsumfeld für die Promotionszeit.

Ich möchte mich bei meiner Familie für die bedingungslose Unterstützung und den Rückhalt während meiner gesamten Ausbildungszeit bedanken. Das Wissen um diese Unterstützung ist für mich von unschätzbarem Wert und hat ungemein dazu beigetragen dass ich mich auf mein Studium konzentrieren konnte. Von Herzen möchte ich noch einigen Personen danken, die mich in der Zeit der Promotion begleitet, abgelenkt und immer wieder aufgemuntert haben (in alphabetischer Reihenfolge): Andrea, Andrea, Annika, Basia, Carol-Ann, Demet, Felix, Ines, Jakob, Jana, Julia, Lotte, Lukas, Martin, Marvin, Matthias, Rafael, Saskia, Steph, Tobias, Ute, ... An alle ein großes Dankeschön. Matthias Zinram kann ich nicht genug danken! Ohne seine Unterstützung, seine Ermutigung und seine Zuversicht hätte ich es nicht geschafft.

Contents

1 Introduction

1.1 Motivation and overview

Theoretical population genetics describes the evolution of the genetic composition of populations driven by forces such as selection, mutation, migration and recombination. In this thesis, we are particularly interested in the impact of recombination, which can briefly be described as an exchange of genetic material from maternal and paternal gene-sequences during sexual reproduction. Such physical swaps between maternal and paternal chromosomes are called crossovers and were discovered by Morgan [101] in 1911. Since the exchange is in general as likely to create a change for the worse as a change for the better, there is a controversial discussion about the selective advantage of recombination, see for example [24, 85, 92, 106, 107]. Agreement, on the other hand, is found in the observation that recombination is conserved in virtually all cells on earth [2, Chap. 19]. In many organisms, there is at least one recombination event per chromosome per reproduction step [84, 90, 108, 115].

Driven by the ongoing technological developments in DNA sequencing, it is now possible to analyse large genomic data-sets, whereas not long ago it was challenging enough to obtain sequence data from a single locus [126]. While it might have been reasonable to neglect recombination events in the single-locus analysis, large data-sets now prompt population geneticists to incorporate recombination into the basic models and to study how recombination affects the genetic composition of a population over time [86]. In particular, there is a considerable interest in how the correlations between sites (known as linkage disequilibria) develop under recombination. The interaction between individuals caused by recombination, however, adds a challenging layer to mathematical models.

Recombination models come in various flavours. One major difference is the number of loci and the number of alleles per loci considered. Many models, in particular the early ones, are restricted to two loci and two alleles per locus [21, 42, 43, 54, 58, 73, 75, 76, 77, 78]. More recent models are usually either multi-locus models [17, 19, 22, 59, 69, 79] or continuous-sequence models, where chromosomes are identified with unit intervals [30, 60, 61, 79, 95, 131]. A second difference in recombination models is the recombination pattern. Pioneering models, such as the ones introduced in [14, 53], allow a very general recombination set-up, that is, they allow for any number of crossover events to occur per chromosome per reproduction step. Some models even allow for an arbitrary number of parents [5, 6, 28, 53, 94]. On the contrary, although rarely made explicit, most of the

recent recombination models assume *single-crossover* recombination (also known as simple crossovers [89, Chap. 6]), which allows only one crossover event per reproduction step per chromosome. Due to the observation that one crossover decreases the probability of a second crossover nearby [67], single-crossover recombination is indeed a biologically relevant case even for fairly large genomic regions.

In this thesis, we describe a general multi-locus, multi-allele and multi-crossover recombination model based on set-partitions. Some results are obtained for the single-crossover case only. We will throughout the thesis compare discrete and continuous-time formulations of our models. Even though there are no particular advantages in considering one or the other, it seems that the overwhelming part of the literature deals with nonoverlapping generations (discrete time), whereas overlapping generations (continuous time) are often easier to treat mathematically.

To begin with, we collect some important facts about posets, partitions and Möbius functions in Section 1.2, which will be used frequently in the main part of the thesis. We then introduce the general recombination model in Section 1.3 and point out simplifications in the single-crossover case. Chapter 2 describes the forward dynamics of a haploid population evolving under recombination. After a short review of well-known results in the deterministic setting, we focus attention to the evolution of finite populations. For the stochastic setting forward in time, we describe the Moran model (continuous time) and the Wright-Fisher model (discrete time) with multi-crossover recombination similar to previous models that appeared in [8, 9, 10, 73].

In line with modern population genetics, we then shift perspective from the forward, prospective, view to the backward, retrospective, view and trace back the ancestry of a sample of individuals taken from a present population. The dispersal of genetic material of the present sample to the ancestors in the past is described by an ancestral process, also called coalescence process. Hudson (1982) was the first who incorporated recombination into the coalescence analysis [68]. Ever since, various expansions of the model have appeared [60, 61, 69, 95, 116, 122, 131, 132, 133]. Most commonly, these processes assume the diffusion limit, in which time is sped up by population size and population size tends to infinity. The corresponding graphical picture is best known under the name ancestral recombination graph (ARG). Since different parts of the chromosome may have different ancestries, obtaining results for the evolution of a sample of individuals over time remains a major challenge and is often restricted to certain regions of the parameter space. Approximating the sampling distribution, for instance, works well if population size is large and recombination rate is high [75, 76, 77]. Genealogical-based inference methods, on the other hand, are computationally intensive and will work well only if population size is large and recombination rate is low [95].
In Chapter 3, we present an alternative route that differs from common approaches in two ways. On the one hand, we describe the *finite* process instead of the diffusion limit. The finite model allows one to draw conclusions about the entire parameter space and to consider various limits in the end. On the other hand, we consider a marginalised version of the ancestral process in which each locus is followed in one individual only. The respective process (to be called *partitioning process*) takes values in the set of partitions of loci and is described in detail in Section 3.2.1. The marginalised approach is rich enough to answer questions of interest such as the evolution of correlations of sites.

The formal duality between the Moran model forward in time and the marginal ancestral process backward in time is proved in Chapter 4. The starting point was a paper by Bobrowski et al. [22] whose setting is entirely forward in time, and thereby hides the genealogical structure. Similar duality results in the diffusion limit can be found in [43, 59, 73, 91]. We are then interested in the time course of our (finite) Moran model. Based on the duality relation, expected type frequencies and linkage disequilibria of all orders can be calculated by studying certain quantities of the partitioning process. Explicit results are obtained in the two-site and three-site case. Since there is no mutation, a single type will go to fixation in the long run. On the grounds of the duality statement, we reveal the relationship between the fixation probabilities of the Moran model and the stationary distribution of the partitioning process.

In Chapter 5, we will again rely on the interplay between the forward and the backward picture of our processes, this time for sufficiently large populations. If population size tends to infinity, the partitioning process from Chapter 3 turns into a process of pure refinements, called segmentation process. Studying the probability distribution of the segmentation process will lead to the solution of the deterministic recombination equation described in Chapter 2. We give a conceptual proof for the probability distribution in the single-crossover case, for which an explicit solution was previously stated by Baake and von Wangenheim [10]. The solution in [10] was obtained from a technical calculation and hinted at an underlying inclusion-exclusion principle that could not be made concrete so far. We will show that this inclusion-exclusion expression appears as a consequence of a Möbius inversion on a suitable poset of rooted forests that we construct in Section 5.2.

We summarise our findings in Chapter 6.

Chapter 3, Chapter 4 and the Moran model in Chapter 2 are built on a joint project together with PhD student Sebastian Probst, supervised by Prof. Dr. Ellen Baake. The joint work captures the forward and backward perspective of single-crossover recombination in continuous time and has been published in [40]. Both PhD students contributed equally to the manuscript. It was first planned that Sebastian Probst concentrates on the forward-time process together with the corresponding half of the duality result, whereas the author of this thesis was supposed to focus on the respective backward-time counterparts (Sect. 3.2.1). Nevertheless, it finally appeared that the employed techniques are more strongly connected as expected, so that the contribution of the authors can not be disentangled in detail. The considerations in this thesis extend the ones in [40] with respect to several aspects. Firstly, the models and properties are generalised to multi-crossover recombination. The models are secondly supplemented by the corresponding discrete-time counterparts and are thirdly complemented by several limit results in forward time (Sect. 2.2.3) and by a more detailed investigation of the stationary distribution of the partitioning process and its correspondence to the fixation probabilities of the Moran model (Sect. 4.4).
The results in Chapter 5 are submitted [7].

1.2 Preliminaries

Working with partitions will be essential to our approach, and we will rely throughout on the powerful concept of *Möbius functions* and *Möbius inversion*. Let us briefly collect the basic notations and standard results starting with a partially ordered set. We follow the description in [18, Chap. 1] and [123, Chap. 3], see also [31, 57].

A *partially ordered set*, or *poset* $P = (X, \preccurlyeq)$, is a set X equipped with a binary relation $\preccurlyeq$ on $X \times X$ that satisfies

- $x \preccurlyeq x$ (reflexivity),
- if $x \preccurlyeq y$ and $y \preccurlyeq x$, then $x = y$ (antisymmetry),
- if $x \preccurlyeq y$ and $y \preccurlyeq z$, then $x \preccurlyeq z$ (transitivity)

for all $x, y, z \in X$. A poset P is called *finite* if the (cardinal) number of elements in P is finite. If all intervals of the form

$$[x, y] := \{z \in X \mid x \preccurlyeq z,\, z \preccurlyeq y\}, \quad x, y \in X,$$

are finite, then P is called a *locally finite poset*. As usual, we write $x \prec y$ if $x \preccurlyeq y$ and $x \neq y$. We write $x \succcurlyeq y$ if $y \preccurlyeq x$. Two elements x and y are said to be *comparable* if $x \preccurlyeq y$ or $y \preccurlyeq x$; otherwise they are said to be *incomparable*. A subset of P in which all of its elements are comparable is called a *chain*. A subset in which all of its elements are incomparable is an *antichain*. An element x in $P = (X, \preccurlyeq)$ is *minimal* (*maximal*) if there is no $y \in X$ such that $y \prec x$ ($y \succ x$). If there exists an element $\mathbf{0}$ such that $\mathbf{0} \preccurlyeq x$ for all $x \in X$, then $\mathbf{0}$ is called the *minimal element* of P. Dually, the *maximal element*, if it exists, is denoted by $\mathbf{1}$.

We say that y covers x, or respectively that x is covered by y, if $x \prec y$ and if there is no $z \in X$ such that $x \prec z \prec y$. The elements that cover $\mathbf{0}$ are called *atoms*. The elements that are covered by $\mathbf{1}$ are called *co-atoms*. We call w an *upper bound* (*lower bound*) of x and y if $x \preccurlyeq w$ and $y \preccurlyeq w$ ($w \preccurlyeq x$ and $w \preccurlyeq y$). If there is a (unique) smallest upper bound z, i.e. if there is an upper bound z that satisfies $z \preccurlyeq w$ for all upper bounds w of x and y, we call this the *join* of x and y (sometimes also denoted as the *least upper bound* or the *supremum*) and denote it by $x \vee y$. Dually, the *meet* or *greatest lower bound*, if it exists, is denoted by $x \wedge y$. If a least upper bound and a greatest lower bound exist for all pairs of elements of a poset, we call the poset a lattice [18, p. 6].

Finite posets are often represented by graphs whose vertices are the elements of that poset. Elements that are minimal (maximal) are placed at the bottom (top) of the graph. A vertex y is placed above a vertex x if $x \prec y$. Two vertices x and y with $x \prec y$ are connected if y covers x. The resulting graph is called *Hasse diagram*.

We call I an *(order) ideal* of a poset $P = (X, \preccurlyeq)$ if I is a subset of X and if for every $x \in I$ and $y \in X$ with $y \preccurlyeq x$ follows that $y \in I$ [18, p. 8]. If $J(P)$ denotes the set of all ideals of P, then $(J(P), \subseteq)$ is a (distributive) lattice [123, p. 106].

Two posets $P = (X_P, \preccurlyeq_P)$ and $Q = (X_Q, \preccurlyeq_Q)$ are *isomorphic* if there is an order-preserving bijection $\phi\colon P \to Q$ that satisfies $\phi(x) \preccurlyeq_Q \phi(y)$ if and only if $x \preccurlyeq_P y$, $x, y \in X_P$. The poset Q is a *subposet* of P if $x \preccurlyeq_Q y$ holds for all $x, y \in X_Q$ whenever $x \preccurlyeq_P y$. The *direct product* $P \times Q$ of two posets P and Q is defined on the set of all tuples $\{(x, y) : x \in X_P,\ y \in X_Q\}$ such that $(x_1, y_1) \preccurlyeq_{P\times Q} (x_2, y_2)$ precisely if $x_1 \preccurlyeq_P x_2$ and $y_1 \preccurlyeq_Q y_2$.

1.2.1 Partitions

A *partition* $\mathcal{A}$ of a finite set $W \subset \mathbb{N}_0$ is a collection of nonempty subsets $A_1, \ldots, A_m$ such that $A_i \cap A_j = \varnothing$ for all $j \neq i$ and $A_1 \cup \cdots \cup A_m = W$. We call $A_i = \{a_{i_1}, \ldots, a_{i_{n_i}}\}$ with $a_{i_j} < a_{i_{j+1}}$ a block of $\mathcal{A}$ and denote by $|\mathcal{A}|$ the number of blocks in $\mathcal{A}$. A partition into k blocks is called a *k-partition*. Unless specified otherwise, the blocks will be listed in increasing order so that A_1 is the block containing the smallest element of W, A_2 is the block containing the smallest element not in the block A_1 and so on. The notions are based on [1, p. 69-70], [3, Chap. 13.3], [16, Chap. 1-2], [117] and [123, Chap. 3.10].

Let $\mathbb{P}(W)$ denote the set of all partitions of W. The set $\mathbb{P}(W)$ equipped with the *refinement relation* $\preccurlyeq$ forms a poset, where $\mathcal{A} \preccurlyeq \mathcal{B}$ means that every block of $\mathcal{A}$ is a subset of a block of $\mathcal{B}$. In the described case, we call $\mathcal{A}$ a refinement of $\mathcal{B}$ or, vice versa, $\mathcal{B}$ a *coarsening* of $\mathcal{A}$. $\mathbb{P}(W)$ has a minimal element $\mathbf{0} = \{\{x\} \,|\, x \in W\}$ and a maximal coarsest element $\mathbf{1} = \{W\}$. The greatest lower bound or meet of two partitions $\mathcal{A}$ and $\mathcal{B}$ will be denoted by $\mathcal{A} \wedge \mathcal{B}$. Analogously, denote the least upper bound or join of $\mathcal{A}$ and $\mathcal{B}$ by $\mathcal{A} \vee \mathcal{B}$.

If U and V are two disjoint (finite) sets and if $\mathcal{A} \in \mathbb{P}(U)$ and $\mathcal{B} \in \mathbb{P}(V)$, then $\mathcal{A} \cup \mathcal{B}$ is a partition of $\mathbb{P}(U \cup V)$. A partition $\mathcal{A} \in \mathbb{P}(W)$ *induces* a unique partition $\mathcal{A}|_C$ on a subset $C \subseteq W$ by restriction, that is, $\mathcal{A}|_C$ consists precisely of all nonempty sets of the form $A_i \cap C$, $i = 1, \ldots, |\mathcal{A}|$.

Example 1.1. Consider the set $W = \{1, \ldots, 5\}$, a partition $\mathcal{A} = \{\{1, 3, 4\}, \{2, 5\}\}$, some other partition $\mathcal{B} = \{\{1, 4\}, \{2, 3\}, \{5\}\}$ and a subset $C = \{1, 3, 5\} \subseteq W$. Here, $\mathcal{A} \wedge \mathcal{B} = \{\{1, 4\}, \{2\}, \{3\}, \{5\}\}$, $\mathcal{A} \vee \mathcal{B} = \{\{1, \ldots, 5\}\}$, $\mathcal{A}_{|_C} = \{\{1, 3\}, \{5\}\}$ and $\mathcal{B}_{|_C} = \{\{1\}, \{3\}, \{5\}\}$. ♢

For a given partition $\mathcal{A} = \{A_1, \ldots, A_m\} \in \mathbb{P}(W)$, let $M := \{1, 2, \ldots, m\} = M(\mathcal{A})$ be the corresponding index set. Obviously, M depends on $\mathcal{A}$, but we suppress this dependence when there is no risk of confusion. For $J \subseteq M$, we define $\mathcal{A}_J := \{A_j\}_{j \in J}$ and $A_J := \cup_{j \in J} A_j$. Clearly, $\mathcal{A}_J$ is a partition of A_J. In particular, $\mathcal{A}_M = \mathcal{A}$, $A_M = W$, $\mathcal{A}_{\{j\}} = \{A_j\}$ and $\mathcal{A}_{M \setminus \{j\}} = \mathcal{A} \setminus \{A_j\}$ for any $j \in M$. We will throughout abbreviate $J \setminus j := J \setminus \{j\}$ and $J \cup k := J \cup \{k\}$.

There are some basic isomorphic relations for the poset of partitions. First, $(\mathbb{P}(W), \preccurlyeq)$ is isomorphic to $(\mathbb{P}(\{1, \ldots, |W|\}), \preccurlyeq)$. Secondly, any coarsening of a partition $\mathcal{B} \in \mathbb{P}(W)$ is obtained by merging complete blocks of $\mathcal{B}$. On the other hand, any refinement of $\mathcal{B}$ may be obtained by refining the blocks of $\mathcal{B}$ separately. Hence

$$([\mathcal{B}, \mathbf{1}_W], \preccurlyeq) \simeq (\mathbb{P}(\{1, \ldots, |\mathcal{B}|\}), \preccurlyeq), \qquad ([\mathbf{0}_W, \mathcal{B}], \preccurlyeq) \simeq \bigtimes_{i=1}^{|\mathcal{B}|} (\mathbb{P}(\{1, \ldots, |B_i|\}), \preccurlyeq). \tag{1.1}$$

Combining the two previous statements yields for all $\mathcal{A}, \mathcal{B} \in \mathbb{P}(W)$ the correspondence

$$([\mathcal{A}, \mathcal{B}], \preccurlyeq) \simeq \bigtimes_{i=1}^{|\mathcal{B}|} (\mathbb{P}(\{1, \ldots, |n_i|\}), \preccurlyeq), \quad \mathcal{A} \preccurlyeq \mathcal{B}, \tag{1.2}$$

where n_i is the number of blocks of $\mathcal{A}$ within block B_i, $|\mathcal{A}| = \sum_{i=1}^{|\mathcal{B}|} n_i$.

Subsets of $\mathbb{P}(W)$ There are some subsets of $\mathbb{P}(W)$ that have a specific relevance for biological applications. Since every offspring descents from exactly two parents, it is convenient to define the subset $\mathbb{P}_2(W) := \{\mathcal{A} \in \mathbb{P}(W) \,|\, |\mathcal{A}| = 2\}$ of all partitions of W into exactly two blocks. The set of all partitions of W into at most two blocks is $\mathbb{P}_{\leqslant 2}(W)$. The number of elements in $\mathbb{P}_2(W)$ is $2^{|W|-1} - 1$.

Secondly, there is the set $\mathbb{O}(W)$ of all *contiguous* or *ordered* partitions of W. A partition is called ordered in W if every block is of the form $A_i = \{x \in W \mid \min A_i \leq x \leq \max A_i\}$. Any ordered partition may alternatively be described by a set of 'break points' that separate the blocks. Since for a given $\mathcal{A} \in \mathbb{O}(W)$, there are $|\mathcal{A}| - 1$ possible break positions, $(\mathbb{O}(W), \preccurlyeq)$ is isomorphic to $(\wp(\{1, \ldots, |W| - 1\}), \subseteq)$, where $\wp(\{1, \ldots, |W| - 1\})$ is the set of subsets of $\{1, \ldots, |W| - 1\}$. $\mathbb{O}(W)$ has cardinality $2^{|W|-1}$. The set of all ordered partitions of W into exactly two blocks is $\mathbb{O}_2(W)$. The set of all ordered partitions of W into at most two blocks is $\mathbb{O}_{\leq 2}(W)$.

We now investigate some counting functions related to $\mathbb{P}(W)$. For a detailed survey of this topic, we refer to [16, Chap. 1.10-1.11] or [56, Chap. 6]. Due to the isomorphic relations in (1.1) and (1.2), it suffices to consider the set of all partitions of a contiguous set, say $S := \{1, \ldots, n\}$.

Cardinalities related to $\mathbb{P}(S)$ The number of partitions of $S = \{1, \ldots, n\}$ is B_n, known as the n-th *Bell number*, and recursively defined via $B_{n+1} = \sum_{k=0}^{n} \binom{n}{k} B_k$ with initial value $B_0 = 1$. B_n can be expressed as $B_n = \frac{1}{e} \sum_{k=0}^{\infty} \frac{k^n}{k!}$. The number of k-partitions of S, namely the number of partitions in $\mathbb{P}(S)$ into exactly k blocks, is given by $S(n, k)$, the *Stirling number of the second kind* (sometimes also referred to as $\left\{ {n \atop k} \right\}$ or $S_{n,k}$). For the Stirling number of the second kind there is a recurrence relation of the form $S(n + 1, k) = S(n, k - 1) + k\, S(n, k)$, $1 < k < n$, with boundary conditions $S(n, 1) = S(n, n) = 1$. As for the Bell number, there is again a closed expression given by $S(n, k) = \frac{1}{k!} \sum_{i=0}^{n} (-1)^{k-i} \binom{k}{i} i^n$. The Bell number and the Stirling number of the second kind are related in an obvious way via $B(n) = \sum_{k=1}^{n} S(n, k)$.

1.2.2 Möbius inversion

Möbius inversion is a fundamental principle in combinatorics that allows to invert a finite series ranging over a locally finite poset. In 1935, almost a century after the first publication, a far-reaching generalisation of the classical Principle of Inclusion-Exclusion has been established independently by Weisner [130] and shortly thereafter by Hall [62]. Both have been motivated by group theoretical questions and until then did not see the combinatorial implications of their group-theoretical framework. Only much later (1964), Rota [117] incorporated their results into the extensive theory of combinatorics on posets and lattices. Among other things, Rota drew the connection between a general inversion statement from Weisner and Hall and the specific number-theoretical inversion formula

$$g(n) = \sum_{d|n} f(d) \quad \Leftrightarrow \quad f(n) = \sum_{d|n} \mu\left(\frac{n}{d}\right) g(d) = \sum_{d|n} \mu(d)\, g\Big(\frac{n}{d}\Big), \tag{1.3}$$

which has been identified by Möbius [98] in 1932. Here, $d|n$ means that n is divisible by d, and the function μ is the number theoretical Möbius function defined by $\mu(n) = (-1)^k$

if $n = p_1 \cdot p_2 \cdot \dots \cdot p_k$, $p_1, \dots, p_k$ being distinct prime numbers, and $\mu(n) = 0$ otherwise. The specific inversion in (1.3), called Möbius inversion by Hardy and Wright [63], lended its name to the fundamental and unifying principle of inversion on partially ordered sets. We briefly introduce the concept in the following. Background material and generalisations can be found in [1, Chap. 4], [3, Chap. 13], [16, Chap. 3], [123, Chap. 3] or [117].

Let $P = (X, \preccurlyeq)$ be a locally finite poset and $\mathrm{Int}(P)$ be the set of all intervals of P. The set $I(P) := \{f \colon \mathrm{Int}(P) \to \mathbb{R}\}$ equipped with the standard addition, scalar multiplication and convolution of functions is called the *incidence algebra* of P over $\mathbb{R}$.

A famous element of the incidence algebra is the well-known *delta function*, or *Kronecker function*, δ with $\delta_{x,y} := \delta(x,y) = 1$ if $x = y$, and $\delta_{x,y} = 0$ otherwise. The delta function serves as the multiplicative two-sided identity element of $I(P)$. A second famous element is the *zeta function* ζ, also called *Riemann function*, which is for any two objects $x, y \in X$ given by $\zeta(x,y) = 1$ if $x \leqslant y$, and $\zeta(x,y) = 0$ otherwise. The inverse of ζ is denoted by μ and called *Möbius function*. Based on the property $\mu\zeta = \delta$, one can define μ inductively for all $x, y, z \in X$ via

$$\mu(x,x) = 1 \quad \text{and} \quad \mu(x,y) = -\sum_{x \preccurlyeq \dot{z} \prec y} \mu(x,z) = -\sum_{x \prec \dot{z} \preccurlyeq y} \mu(z,y), \quad x \prec y, \tag{1.4}$$

where the underdot indicates the summation variable. As a direct consequence of (1.4), one obtains

$$\sum_{x \preccurlyeq \dot{z} \preccurlyeq y} \mu(x,z) = \begin{cases} 1, & \text{if } x = y, \\ 0, & \text{otherwise,} \end{cases} \tag{1.5}$$

which will be used frequently in the main part of the thesis.

The inverse property for μ and ζ ensures that for any two functions $f, g : P \to \mathbb{R}$, the relation $f\,\zeta = g$ holds if and only if $f = g\,\mu$. This equivalence expression leads to the very general and powerful inversion theorem called *Möbius inversion*.

Theorem 1.1 (Möbius inversion, [1, Prop. 4.18])**.** *Let P be a locally finite poset and let f and g be two functions with $f, g \colon P \to \mathbb{R}$.*

- *Inversion from below: If there exists a minimal element $\mathbf{0} \in P$, then*

$$g(x) = \sum_{\mathbf{0} \preceq \dot{y} \preceq x} f(y) \quad \Leftrightarrow \quad f(x) = \sum_{\mathbf{0} \preceq \dot{y} \preceq x} g(y)\,\mu(y,x).$$

- *Inversion from above: If there exists a maximal element $\mathbf{1} \in P$, then*

$$g(x) = \sum_{x \preceq \dot{y} \preceq \mathbf{1}} f(y) \quad \Leftrightarrow \quad f(x) = \sum_{x \preceq \dot{y} \preceq \mathbf{1}} g(y)\,\mu(x,y).$$

It is important to note that Möbius inversion is not restricted to functions. It also holds for bounded operators.

To apply Möbius inversion on any poset, it is essential to compute the Möbius function of that poset first. Given a concrete, small poset, this is an easy task due to the recursive definition in (1.4). For more general results, there are certain elaborated techniques available, see [1, Chap. 4.3] or [123, Chap. 3.8]. The most simple one is the product structure.

Proposition 1.1 (The product theorem, [123, Prop. 3.8.2]). *Let P and Q be two locally finite posets with Möbius functions μ_P and μ_Q, and let $P \times Q$ be their direct product with Möbius function $\mu_{P\times Q}$. Then*

$$\mu_{P\times Q}((x,y),(x',y')) = \mu_P(x,x') \cdot \mu_Q(y,y')$$

if $(x,y) \preccurlyeq (x',y')$ in $P \times Q$.

Within this thesis, we will rely on two known expressions for Möbius functions: the first with respect to the powerset poset $(\wp(S), \subseteq)$, where $\wp(S)$ denotes the set of subsets of S, the second with respect to $(\mathbb{P}(S), \preccurlyeq)$, the poset on the set of partitions of S.

Example 1.2 (Set of subsets). The Möbius function for the powerset poset $(\wp(S), \subseteq)$ is for any two subsets $A, B \subseteq S$ given by

$$\mu(A,B) = \begin{cases} (-1)^{|B|-|A|}, & \text{if } A \subseteq B, \\ 0, & \text{otherwise.} \end{cases}$$

The corresponding Möbius inversion formula is the Principle of Inclusion-Exclusion in its purest form [123, p. 64]. ♢

Example 1.3 (Lattice of partitions). The Möbius function for the poset $(\mathbb{P}(S), \preccurlyeq)$ is for any pair of partitions $\mathcal{A}, \mathcal{B} \in \mathbb{P}(S)$ given by

$$\mu(\mathcal{A},\mathcal{B}) = \prod_{j=1}^{|\mathcal{B}|} \mu(\mathcal{A}|_{B_j}, \mathbf{1}|_{B_j}) = \prod_{j=1}^{|\mathcal{B}|} (-1)^{n_j-1}(n_j-1)!\,, \quad \mathcal{A} \preccurlyeq \mathcal{B}, \tag{1.6}$$

where n_j is the number of blocks of $\mathcal{A}$ within block B_j of $\mathcal{B}$, that is, n_j is the number of blocks in $\mathcal{A}|_{B_j}$, $1 \leqslant j \leqslant |\mathcal{B}|$. The Möbius function for $(\mathbb{P}(S), \preccurlyeq)$ was discovered independently by Schützenberger [119] and by Frucht and Rota [52]. ♢

1.3 Recombination model

Within this section, we present a general multi-locus, multi-allele and multi-crossover recombination model, which is the generalisation of the recombination model introduced by E. Baake and M. Baake [12].

For the beginning, let us collect some basic genetic vocabulary. By a *gene* we understand a contiguous sequence of DNA that codes proteins and that is inherited from one generation to the next. Every gene may have certain distinguishable forms, called *alleles*. The specific location of a gene on a chromosome is called *locus* (plural *loci*). Since we do not want to distinguish between loci or positions of single nucleotides, we will work with *sites* instead. Each site may represent an allele, a specific base (adenine, cytosine, guanine, or thymine) or any other variable one could think of.

In sexual reproducing individuals, the majority of body cells is *diploid*, that is, they carry two sets of chromosomes, one maternal and one paternal. In contrast, *gamete cells* (eggs and sperms) are *haploid*; every gene is present in a single copy only.

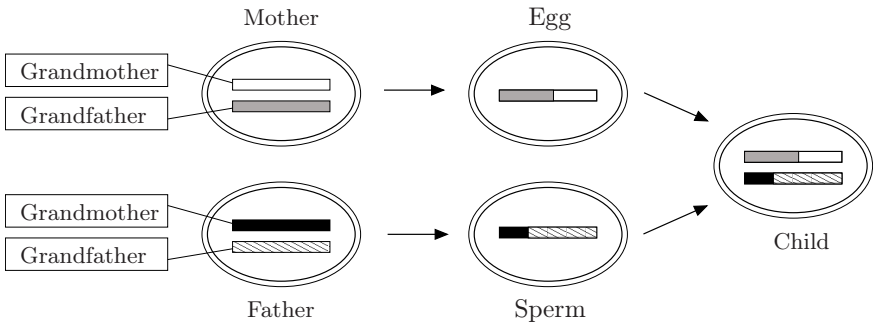


Figure 1.1. Simplified reproduction process involving recombination (exemplified by a single set of chromosomes only). Diploid maternal and paternal cells form haploid gamete cells (eggs and sperms) during meiosis. During this cell division, grandmaternal and grandpaternal chromosomes may recombine and exchange parts of their DNA sequences. When egg and sperm cells fuse afterwards, they form a diploid zygote that is composed of a mixture of maternal and paternal genetic material. Only single-crossover outcomes are shown.

During fertilisation, haploid egg and sperm cells join together and form a diploid *zygote*. In order to produce gamete cells, a special process of cell division (called *meiosis*) is required. During meiosis, the grandmaternal and grandpaternal chromosomes that belong together (*homologous chromosomes*) may interact and physically swap some of their genetic material; a process called *recombination*, see Figure 1.1 or [20, 49, 110, 118] for a detailed overview. After the physical swap, there are two possible recombination products that can arise: *crossover* and *noncrossover*. The first is the result of a reciprocal exchange of the chromosomes, the latter of a nonreciprocal one. In our model, we will neglect noncrossover outcomes and identify recombination with crossover recombination in the following. At a crossover point, two chromosomes cross and interchange their genetic sequences to either the right or the left hand side of this crossover point. A crossover may happen between any pair of nucleotides. As illustrated in Figure 1.2, even multiple crossovers can occur within one reproduction step between a single pair of chromosomes. Due to the observation that one crossover decreases the probability of a second crossover nearby (a phenomenon called *interference*), the majority of recombination events is assumed to be generated by a single crossover.

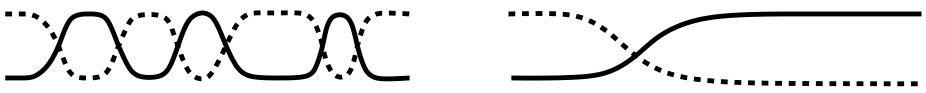


Figure 1.2. Multiple crossovers on the left hand side. A single crossover on the right hand side.

Our recombination model will be based on set-partitions. The partition framework has proofed very useful in the setting of recombination [6, 27, 28, 40, 89]. On the one hand, partitions occur naturally when describing an offspring sequence with regard to the particular parts that are inherited from the mother and the particular parts inherited from the father (see Figure 1.3). On the other hand, partitions may describe the dispersal of genetic material across the ancestors of an individual backward in time. In some parts of this thesis, we restrict ourselves to the special case of single-crossover recombination, for which

the partition notation is not compulsory, but in many cases more convenient. Restriction to single-crossover recombination corresponds to the assumption of complete interference [25].

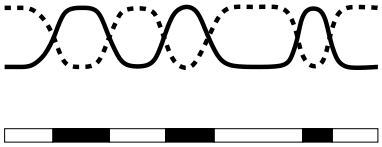


Figure 1.3. Correspondence between crossovers and partitions.

1.3.1 Multi-crossover recombination

Let a chromosome be described via a linear arrangement of n discrete positions called *sites*, which are collected in the set $S = \{1, 2, \ldots, n\}$. Each of those sites $i \in S$ may represent a nucleotide or a gene locus captured in a finite set $\mathbb{X}_i$. If sites are nucleotide sites, a natural choice for each $\mathbb{X}_i$ is the nucleotide alphabet $\{A, C, G, T\}$; if sites are gene loci, $\mathbb{X}_i$ is the set of alleles that can occur at locus i. We restrict ourselves here to finite sets $\mathbb{X}_i$, but generalisations to locally compact sets are available [6, 11, 12]. We describe individuals on the level of gametes and identify the genetic type of each individual with the sequence $x = (x_1, x_2, \ldots, x_n)$, $x_i \in \mathbb{X}_i$. If sites represent gene loci, x may be considered as the *haplotype* of an individual. The complete *type space* is given by $\mathbb{X} := \mathbb{X}_1 \times \mathbb{X}_1 \times \cdots \times \mathbb{X}_n$.

Throughout this thesis, we assume that an individual is created as a mixture of (at most) two parental individuals, but we do not keep track of which part is maternal and which is paternal. If not stated otherwise, any recombination event (including multi-crossover ones) can occur in a single reproduction step. Assume, for instance, that two individuals, say the first of type x and the second of type y, experience a double-crossover event; one crossover between i and $i+1$, and one between j and $j+1$, $1 \leqslant i < j < n$. As illustrated in Figure 1.4, the result is a mixed type offspring that inherits the type $(x_1, \ldots, x_i, y_{i+1}, \ldots, y_j, x_{j+1}, \ldots, x_n)$. It will turn out useful to describe such a recombination event via an exchange of the parental sequences according to the partition $\mathcal{A} = \{\{1, \ldots, i, j+1, \ldots, n\}, \{i+1, \ldots, j\}\}$. All partitions of S into at most two parts, namely all partitions in $\mathbb{P}_{\leqslant 2}(S)$, can be realised within one reproduction event.

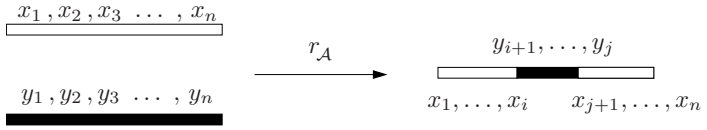


Figure 1.4. An ordered pair individuals, the first of type x and the second of type y, are chosen to reproduce according to the partition $\mathcal{A} = \{\{1, \ldots, i, j+1, \ldots, n\}, \{i+1, \ldots, j\}\}$, $1 \leqslant i < j < n$. The result is the mixed-type individual on the right hand side.

We formalise the reproduction process of individuals as follows: At each reproduction step, with probability $r_{\mathcal{A}}$, $\mathcal{A} \in \mathbb{P}_2(S)$, where $\mathcal{A} = \{A_1, A_2\}$, two parents recombine to form an offspring. The new individual copies the letters at all sites in A_1 from the first individual and the letters at all sites in A_2 from the second individual. The sum $\sum_{\mathcal{A} \in \mathbb{P}_2(S)} r_{\mathcal{A}} \leq 1$ is the probability that at least one crossover takes place during reproduction. With probability $r_{\mathbf{1}} = 1 - \sum_{\mathcal{A} \in \mathbb{P}_2(S)} r_{\mathcal{A}}$, there is no recombination and the offspring is an unaltered copy of

a single parent. The collection $\{r_{\mathcal{A}}\}_{\mathcal{A}\in\mathbb{P}_2(S)}$ is called *recombination distribution* [23, 25] or *linkage distribution* [53, 89].

From the perspective of the newly synthesised individual, only those parts of the parental types matter that are copied by the offspring. If a recombination event occurred according to the partition $\mathcal{A} = \{\{1,\dots,i,j+1,\dots,n\},\{i+1,\dots,j\}\}$, the offspring is of type x whenever the first parent is of type $(x_1,\dots,x_i,*,\dots,*,x_{j+1},\dots,x_n)$ and the second parent is of type $(*,\dots,*,x_{i+1},\dots,x_j,*,\dots,*)$. Here, a $*$ at site i may represent any element of $\mathbb{X}_i$ and refers to marginalisation. We generalise the idea of marginal types in the next section by defining so-called *recombination operators* or *recombinators*. They turned out to be useful to describe the dynamics under recombination in a compact way.

Recombination operators

Denote by $\mathcal{M}_+(\mathbb{X})$ the set of all positive measures on $\mathbb{X}$ (including the zero measure) and by $\mathcal{P}(\mathbb{X})$ be the set of probability measures on $\mathbb{X}$. If we define δ_x as the point measure on x (i.e. $\delta_x(y) = \delta_{x,y}$ for $x, y \in \mathbb{X}$), we can also write $\omega = \sum_{x\in\mathbb{X}} \omega(x)\,\delta_x$ for every $\omega \in \mathcal{M}_+(\mathbb{X})$. Let $\omega(\mathbb{A}) := \sum_{x\in\mathbb{A}} \omega(x)$ for $\mathbb{A} \subseteq \mathbb{X}$ and denote by $\|\cdot\|$ the norm (or total variation norm) of ω, which in our case is $\|\omega\| := \sum_{x\in\mathbb{X}} \omega(x) = \omega(\mathbb{X})$. Define the canonical projection operator $\pi_I\colon \mathbb{X} \to \bigtimes_{i\in I} \mathbb{X}_i =: \mathbb{X}_I$ by $\pi_I(x) = (x_i)_{i\in I}$ for every $I \subseteq S$ as usual. For $\omega \in \mathcal{M}_+(\mathbb{X})$, let $\pi_I.\omega := \omega \circ \pi_I^{-1}$ represent the marginal measure with respect to the sites in $I \subset S$, where π_I^{-1} denotes the preimage of π_I. The operation . (where the dot is on the line and should not be confused with a multiplication sign) is known as the *pushforward* of ω with respect to π_I. When the context is clear, we will write $\omega^I := \pi_I.\omega$. To be precise,

$$\omega^I(x_I) = \omega \circ \pi_I^{-1}(x_I) = \omega(\{x \in \mathbb{X} \mid \pi_I(x) = x_I\}), \quad x_I \in \mathbb{X}_I.$$

In particular, $\omega^S = \omega$. For a partition $\mathcal{A} = \{A_1,\dots,A_m\}$ of $\mathbb{P}(S)$ and a measure $\omega \in \mathcal{M}_+(\mathbb{X})$, we define the *nonnormalised recombinator* $\bar{R}_{\mathcal{A}}\colon \mathcal{M}_+(\mathbb{X}) \to \mathcal{M}_+(\mathbb{X})$ as

$$\bar{R}_{\mathcal{A}}(\omega) = \omega^{A_1} \otimes \cdots \otimes \omega^{A_m}, \tag{1.7}$$

where $\otimes$ indicates the tensor product. The ordering of the sites is specified by the set S. In words, $\bar{R}_{\mathcal{A}}$ turns ω into the product of its marginals with respect to the blocks in $\mathcal{A}$. $\bar{R}_{\mathcal{A}}$ is nonlinear for all partitions except $\mathcal{A} = \mathbf{1} = \{S\}$, which refers to no recombination. In particular, one has $\bar{R}_{\mathbf{1}}(\omega) = \omega$ and $\|\bar{R}_{\mathcal{A}}(\omega)\| = \|\omega\|^{|\mathcal{A}|}$, where $|\mathcal{A}|$ denotes the number of blocks in $\mathcal{A}$. We will throughout indicate nonnormalised mappings by an overbar. The corresponding normalised version

$$R_{\mathcal{A}}(\omega) := \frac{\bar{R}_{\mathcal{A}}(\omega)}{\|\bar{R}_{\mathcal{A}}(\omega)\|} = \frac{1}{\|\omega\|^{|\mathcal{A}|}}\,\bar{R}_{\mathcal{A}}(\omega), \quad \omega \in \mathcal{M}_+(\mathbb{X}) \setminus 0 \tag{1.8}$$

defines a probability measure on $\mathbb{X}$. For consistency, set $R_{\mathcal{A}}(0) := 0$. Consider the set $S = \{1,2,3,4\}$, a partition $\mathcal{A} = \{\{1,2,4\},\{3\}\}$ and a measure $\omega \in \mathcal{M}_+(\mathbb{X}) \setminus 0$. In this case, we can write out the recombinator for any $x \in \mathbb{X}$ as

$$(R_{\mathcal{A}}(\omega))(x) = \frac{1}{\|\omega\|^2}\,\omega^{\{1,2,4\}}(x_{\{1,2,4\}})\;\omega^{\{3\}}(x_{\{3\}}) = \frac{1}{\|\omega\|^2}\,\omega(x_1,x_2,*,x_4)\;\omega(*,*,x_3,*),$$

where a $*$ at site i refers to marginalisation. We will learn more about the probabilistic meaning of the recombinator in Section 4.1. Note here, that in former descriptions of the model [8, 9, 10, 12, 125], the recombinator was defined with a normalisation factor that differs from the one in (1.8) by a factor of $\|\omega\|$. Obviously, both recombinators agree on the set of probability measures. As we will see in Section 2.2.1, the operator in (1.8) seems to be better adapted for the stochastic model, whereas the one in [8, 9, 10, 12, 125] seems to be more natural in the deterministic situation. We point out the differences when necessary.

1.3.2 Single-crossover recombination

In the single-crossover case, in which at most one crossover is allowed per pair of chromosomes per reproduction step, any new synthesised individual inherits two *contiguous* segments, the leading one from the first parent and the trailing one from the second parent (see Figure 1.2). The corresponding partition is always an *ordered* partition into at most two parts, i.e. an element of $\mathbb{O}_{\leqslant 2}(S)$ (cf. Section 1.2.1). We may thus work with the multi-crossover model described above and set $r_{\mathcal{A}} = 0$ for all $\mathcal{A} \notin \mathbb{O}_{\leqslant 2}(S)$. It will, nonetheless, sometimes be more convenient to use a simplified notation based on sets of subsets of 'break points'.

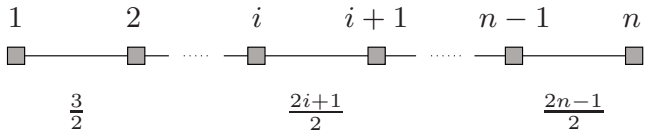


Figure 1.5. The chromosome as a linear arrangement of sites $S = \{1, 2, \ldots, n\}$. The elements in $L = \left\{\frac{3}{2}, \ldots, \frac{2n-1}{2}\right\}$ connect neighbouring sites.

As the set of 'break points' we choose the set of half integers $L = \left\{\frac{3}{2}, \frac{5}{2}, \ldots, \frac{2n-1}{2}\right\}$. The elements of L will be called *links* and be indicated by Greek letters in the following. As represented in Figure 1.5, each link $\alpha \in L$ connects the two neighbouring sites $\lfloor\alpha\rfloor$ and $\lceil\alpha\rceil$, where $\lfloor\alpha\rfloor$ ($\lceil\alpha\rceil$) denotes the largest integer below (the smallest integer above) α. Let $G = \{\alpha_1, \ldots, \alpha_{|G|}\}$ be a subset of L with $\alpha_1 < \alpha_2 < \cdots < \alpha_{|G|}$. Every *ordered* partition $\sigma \in \mathbb{O}_{\leqslant 2}(S)$ of the form $\sigma = \{\sigma_1, \ldots, \sigma_{|G|+1}\}$ with blocks

$$\sigma_1 = \{1, \ldots, \lfloor\alpha_1\rfloor\}, \quad \sigma_2 = \{\lceil\alpha_1\rceil, \ldots, \lfloor\alpha_2\rfloor\}, \quad \ldots, \quad \sigma_{|G|+1} = \{\lceil\alpha_{|G|}\rceil, \ldots, n\}, \tag{1.9}$$

has a one-to-one correspondence to the set G. In the single-crossover case, recombination events may therefore be described as crossover events with respect to certain links.

The reproduction process in terms of links may be formalised as follows: At each reproduction step, with probability r_α, $\alpha \in L$, two individuals, say the first of type x and the second of type y, are chosen to recombine at link α. If a crossover at link α occurs, all sites up to α will be passed on to the offspring from the first individual, all following sites are inherited by the second individual. The result is the mixed-type individual $(x_1, \ldots, x_{\lfloor\alpha\rfloor}, y_{\lceil\alpha\rceil}, \ldots, y_n)$. With probability $1 - \sum_{\alpha\in L} r_\alpha$, there is no crossover and the offspring is a complete copy of a single individual. The recombination distribution is $\{r_\alpha\}_{\alpha\in L}$.

In the single-crossover case, the notation may vary between the notation based on ordered partitions and the notation based on links.

2 Forward time: Dynamics under recombination

In this chapter, we investigate the dynamics of a population evolving under the evolutionary effect of recombination *forward* in time. Our aim is to state the genetic composition of a population at any time based on a given initial population. To this end, we make the following (highly idealised) assumptions to simplify matters and to ease calculations: First, we assume that the population is of constant size N over time. Secondly, we describe the dynamics of the population on the level of gametes, that is, we identify a population with the haploid egg and sperm cells that are produced in each generation. This haploid model is exact in some cases and approximates the diploid model well if population size is large [47, p. 130 & 227]. In the literature, it is common to start with a diploid population of N individuals, approximated by a population of $2N$ haploid individuals. Our results may therefore differ from known results by a factor of two. We further neglect the existence of mutation and selection events and assume that recombination is the only evolutionary force. Moreover, we pretend that both sexes are equal and that the population evolves according to the concept of *random mating*; namely that the individuals mate without any regard to ancestry, geographical or social structure, or any other preferences one could think of. Such a population is sometimes called *panmictic*.

In general, one can distinguish between stochastic and deterministic formulations, in discrete or continuous time. Even though, one or the other approach might be favoured by some groups, evolution is per se a random process so that sampling effects, known as random *genetic drift* or *resampling*, will affect the genetic composition of a population. This makes it unavoidable to study the stochastic perspective, especially for rather small populations. For large populations, on the other hand, the effect of chance events is small. It will need a number of generations that these effects contribute noticeably to the population. In many cases, this number of generations is of the same order as population size [28]. Hence, especially for short time-scales, the deterministic dynamics are assumed to approximate the dynamics of large populations quite well.

Some population geneticists prefer continuous-time models, some discrete-time models. Since 'real populations might exist somewhere in between these two extremes' [128, p. 54] and since 'it is, to a certain extent, a matter of taste whether to use discrete-time or continuous-time models' [23, p. 40], we will try to compare both perspectives throughout the complete thesis, see [4] for a general overview. In *discrete time*, generations do not overlap. More precisely, at each time step, the entire population dies out and is replaced by

its offspring generation. This corresponds to the assumption of equal lifetime expectancy and simultaneous death and birth times for all individuals in the population. It seems that the overwhelming part of the literature deals with nonoverlapping generations, which often allow direct comparison to experimental data (in the lab generation of some species can be kept discrete). In *continuous time*, only a single reproduction event takes place at each time point and generations overlap. The continuous-time models often allow for explicit expressions and thus became more and more attractive for mathematicians.

In the following, we first concentrate on the deterministic model and thereafter investigate the corresponding class of stochastic models in discrete and continuous time. To ease recognition, we will throughout the thesis attach to the discrete-time variables a hat and to the continuous-time ones a caron. If statements hold for both of them, we omit the additional indication.

2.1 Deterministic models

Consider a sufficiently large population (or in fact a population of infinite size) that is identified with a probability vector $p \in \mathcal{P}(\mathbb{X})$, where $p(x) := p(\{x\})$ denotes the proportion of individuals of type $x \in \mathbb{X}$. Let $\check{p}$ and $\hat{p}$ indicate the corresponding continuous-time and discrete-time versions. In discrete time and in the absence of recombination, genotype frequencies attain an equilibrium state after one generation of random mating according to the *Hardy-Weinberg law* [23, Chap. 1.2]. Under recombination, however, every individual is either an unaltered copy of an individual in the former generation or is composed of two recombined sequences. Allele frequencies are thereby conserved but genotype frequencies are not.

The first deterministic attempts to tackle the effects of recombination on gamete frequencies are discrete ones and go back to Jennings [78] (1917) and Robbins [114] (1918). From the beginning, the most challenging part was the nonlinearity of the system caused by the interaction of the parental individuals. For the special case of two diallelic loci, Robbins overcame the obstacles of nonlinearity by defining specific functions acting on gamete frequencies that linearise and diagonalise the system. This method turned out to be the conventional approach for decades. In 1944, Geiringer [53] was the first to establish a general recombination equation for multiple loci, multiple alleles and arbitrary recombination pattern. Translating her framework to the handy notation of partitions and recombinators (cf. (1.8)), the recombination equation reads

$$\hat{p}_{t+1} = \hat{p}_t + \sum_{\mathcal{A}\in\mathbb{P}_2(S)} r_{\mathcal{A}} \left(R_{\mathcal{A}} - \mathbb{1}\right)(\hat{p}_t), \quad t \in \mathbb{N}_0, \tag{2.1}$$

where $r_{\mathcal{A}}$ is the probability of a recombination event according to the partition $\mathcal{A}$, and $\mathbb{P}_2(S)$ is the set of partitions of $S = \{1, \ldots, n\}$ into two blocks. Geiringer gave a general but quite cumbersome procedure to linearise the dynamics. Bennett [14] streamlined this method for up to six sites in 1954. His linear transformation of type frequencies was obtained with the help of linear combinations of functions of allele frequencies, called *principal components*. The principal components can be seen as a particular choice to measure correlations of sites (called *linkage disequilibria* in biology). We will investigate different choices of linkage

disequilibria in Section 4.3. The principal components decay exponentially and depend on recombination probabilities for more than three sites. Bennett computed them up to six sites but did not generalise the method. This was accomplished more than 20 years later in two different directions. One combinatorial approach to obtain recursive expressions for the principal components was given by Dawson [27, 28]. A second approach was worked out by Ljubič [89, Chap. 6] using *genetic algebras*, see also [113] and references therein.

Another deterministic approach was investigated by E. Baake and M. Baake [12] in 2003 for the continuous-time analogue restricted to single crossovers. The model was later generalised to allow arbitrary crossovers and even arbitrary many parents [6]. Even though, from the biological perspective, a generalisation to more than two parents does not seem useful, one could think of a cultural inheritance process such as language, where multiple 'parents' contribute to the general linguistic usage of the offspring. In the following, we briefly summarise the results in the bi-parental case in continuous time followed by those in discrete time.

2.1.1 Continuous time

As stated in [6, Eq. (6)], the deterministic dynamics under general multi-crossover recombination in continuous time can be described via the system of nonlinear differential equations

$$\frac{\mathrm{d}}{\mathrm{d}t}\,\breve{p}_t = \sum_{\mathcal{A}\in\mathbb{P}_{\geqslant 2}(S)} \varrho_{\mathcal{A}}\,(R_{\mathcal{A}} - \mathbb{1})(\breve{p}_t), \quad t \geqslant 0, \tag{2.2}$$

where $\varrho_{\mathcal{A}} > 0$, $\mathcal{A} \in \mathbb{P}_2(S)$, is the *rate* at which two individuals recombine according to $\mathcal{A}$. The case $\mathcal{A} = \{S\}$, which corresponds to no recombination, does not have an effect since gain and loss are equal. It was shown in [5] that the solution to the Cauchy problem (or initial value problem) of (2.2) with initial value $\breve{p}_0 \in \mathcal{P}(\mathbb{X})$ is of the form

$$\breve{p}_t = \sum_{\mathcal{A}\in\mathbb{P}(S)} \breve{a}_t(\mathcal{A})\, R_{\mathcal{A}}(\breve{p}_0), \quad t \geqslant 0, \tag{2.3}$$

which shows that at time t, the population will be a mixture of various recombined populations sampled from the initial population. For the coefficient functions $\breve{a}_t(\mathcal{A})$, there is not yet an explicit expression in the general recombination case available. A recursion is given in [5, 6]. We will revisit the coefficient functions in Chapter 5 and relate them to certain objects of the corresponding stochastic model backward in time. For $t \to \infty$, $\breve{p}_t$ in (2.3) turns into the product of its marginals [6, p. 15]. The asymptotic behaviour is given by

$$\breve{p}_\infty = (\pi_1.\,\breve{p}_0) \otimes (\pi_2.\,\breve{p}_0) \otimes \ldots \otimes (\pi_n.\,\breve{p}_0), \tag{2.4}$$

which was already shown by Robbins for two diallelic loci in 1918.

For the special case of single-crossover recombination, where $\varrho_{\mathcal{A}} = 0$ if $\mathcal{A} \notin \mathbb{O}_2(S)$ (i.e. if $\mathcal{A}$ is not an ordered partition of S into two blocks), the coefficient functions can be expressed explicitly as

$$\breve{a}_t(\mathcal{A}) = \exp\Bigl(-\sum_{\substack{\mathcal{B}\in\mathbb{O}_2(S)\\ \mathcal{B}\not\succeq\mathcal{A}}} \varrho_{\mathcal{B}}\, t\Bigr) \prod_{\substack{\mathcal{B}\in\mathbb{O}_2(S)\\ \mathcal{B}\succcurlyeq\mathcal{A}}} (1 - \exp(-\varrho_{\mathcal{B}}\, t)) \quad \text{if } \mathcal{A} \in \mathbb{O}(S). \tag{2.5}$$

and $\breve{a}_t(\mathcal{A}) = 0$ if $\mathcal{A} \notin \mathbb{O}(S)$.

Due to the one-to-one correspondence between ordered partitions and subsets of links that separate the blocks of the partitions, each $\breve{a}_t(\mathcal{A})$, $\mathcal{A} \in \mathbb{O}(S)$, can be described in terms of links. If $G \subseteq L$ is a subset of links with $G = \{\alpha_1, \ldots, \alpha_{|G|}\}$, $\alpha_1 < \alpha_2 < \cdots < \alpha_{|G|}$, and $\mathcal{A} = \{\sigma_1, \ldots, \sigma_{|G|+1}\}$ is an ordered partition with blocks as in (1.9), the counterpart of (2.5) in the link notation is

$$\breve{a}_t(G) = \exp\Big(-\sum_{\alpha\in L\setminus G} \varrho_\alpha\, t\Big) \prod_{\alpha\in G} (1-\exp(-\varrho_\alpha\, t)), \quad G \subseteq L, \tag{2.6}$$

where ϱ_α is the rate for a crossover at link α. The coefficient functions have a probabilistic interpretation in terms of the corresponding stochastic process (Section 2.2.1). That is, $\breve{a}_t(G)$, $G \subseteq L$, is the probability that up to time t recombination affects exactly the links in G and none of the links in the complementary set $L\setminus G$. Since (2.2) describes a large system of coupled nonlinear differential equations, the existence of an explicit solution is surprising. Baake and Baake [12] emphasize that the astonishingly easy solution is due to the simple form of the transformation functions (what Bennett calls principal components). The transformation functions decouple the single-crossover version of (2.2) into a linear system with the usual exponential solution. In the general recombination case, there is again some underlying linearity. This is worked out in [5].

2.1.2 Discrete time

Due to the results in continuous time, one expects the solution of the discrete-time recombination equation from (2.1) with initial value $\hat{p}_0 \in \mathcal{P}(\mathbb{X})$ to be again of the form

$$\hat{p}_t = \sum_{\mathcal{A}\in\mathbb{P}(S)} \hat{a}_t(\mathcal{A})\, R_{\mathcal{A}}(\hat{p}_0), \tag{2.7}$$

where the $\hat{a}_t(\mathcal{A})$'s are nonnegative and need to be determined. Recursive formulations for the coefficient functions are given in [5, 27, 28, 89].

The transformation method used for the continuous-time case did not admit a closed solution in the discrete-time, single-crossover case [125]. Since in each time step, one crossover forbids further crossovers at any other links, additional dependencies arise that affect the joint distribution of sites for $|S| > 2$. Nonetheless, the method did yield additional insight. It is for instance shown, that the transformation functions used in continuous time linearise the dynamics (for up to three sites they also diagonalise it). The resulting transformed linear system has a subtriangular structure and can be solved by a simple recursion in a second step. This is a great improvement compared to previous solutions. Additionally, it was shown that the coefficient functions in the link notation follow the iteration

$$\hat{a}_{t+1}(G) = \Big(1-\sum_{\alpha\in L} r_\alpha\Big)\hat{a}_t(G) + \sum_{\alpha\in G} r_\alpha \Big(\sum_{H\subseteq L_{\geq\alpha}} \hat{a}_t(G_{<\alpha}\cup H)\Big)\Big(\sum_{K\subseteq L_{\leq\alpha}} \hat{a}_t(G_{>\alpha}\cup K)\Big), \tag{2.8}$$

with initial condition $\hat{a}_0(G) = \delta_{G,\varnothing}$, where

$$\begin{aligned} G_{<\alpha} &= \{\beta \in G \mid \beta < \alpha\}, \qquad & G_{>\alpha} &= \{\beta \in G \mid \beta > \alpha\}, \\ L_{\leq\alpha} &= \{\beta \in L \mid \beta \leq \alpha\}, \qquad & L_{\geq\alpha} &= \{\beta \in L \mid \beta \geq \alpha\} \end{aligned}$$

and where δ denotes the Kronecker function. A verbal description of this iteration was

already given by Geiringer [53]. As we see, the $\hat{a}_t(G)$'s evolve nonlinear, except for the cases in which $G_{<\alpha} = \varnothing$, or $G_{>\alpha} = \varnothing$, which corresponds to the situation in which at least one of the involved segments is not affected by a previous crossover. Explicitly, this corresponds to the links $\alpha \in \{\frac{3}{2}, \frac{2n-1}{2}\}$ and explains the existence of a closed solution for up to three sites. For $S = \{1,2,3\}$ and $L = \{\frac{3}{2}, \frac{5}{2}\}$, this solution reads

$$\begin{aligned}
\hat{a}_t(\varnothing) &= \left(1 - r_{\frac{3}{2}} - r_{\frac{5}{2}}\right)^t, \\
\hat{a}_t(\{\tfrac{3}{2}\}) &= (1 - r_{\frac{5}{2}})^t - \left(1 - r_{\frac{3}{2}} - r_{\frac{5}{2}}\right)^t, \\
\hat{a}_t(\{\tfrac{5}{2}\}) &= (1 - r_{\frac{3}{2}})^t - \left(1 - r_{\frac{3}{2}} - r_{\frac{5}{2}}\right)^t, \\
\hat{a}_t(\{\tfrac{3}{2}, \tfrac{5}{2}\}) &= 1 - (1 - r_{\frac{3}{2}})^t - (1 - r_{\frac{5}{2}})^t + \left(1 - r_{\frac{1}{2}} - r_{\frac{3}{2}}\right)^t.
\end{aligned} \tag{2.9}$$

We will revisit the coefficient functions for an arbitrary number of sites in Chapter 5 and relate them to certain probabilities of the corresponding stochastic process backward in time. The change from the forward to the backward point of view will not only admit a closed expression for the $\hat{a}_t(G)$'s for arbitrary many sites (which was already accomplished in [125]), but will further allow to give them a clear probabilistic interpretation. In particular, it will reveal that the explicit solution for an arbitrary number of sites is again an instance of a Möbius inversion.

2.2 Stochastic models

In line with modern population genetics, let us investigate changes in the genetic composition of a *finite* population due to stochastic fluctuations caused by random sampling (*resampling*). In the absence of recombination, the effects of resampling have been first studied by Fisher (1930) and Wright (1931) in discrete time, followed by Moran (1958) in continuous time. The *Wright-Fisher model* and the *Moran model* are the most commonly used models to describe the dynamics of finite populations. Different evolutionary forces such as migration, selection, mutation, varying population sizes and many others have been incorporated to these models, see for example [23, 47] for good overviews. We restrict ourselves to the effect of recombination and resampling alone and start with the continuous-time model, which is the *Moran model with recombination* studied in [8] in the single-crossover case and in [9] in the general recombination case. We proceed with the discrete-time counterpart, the *Wright-Fisher model with recombination*, investigated in [10] for single crossovers (see also [65, Chap. 5.4] or [60]).

Consider a population of constant size N. Let $t \in \mathbb{T}$, where $\mathbb{T}$ either represents $\mathbb{N}_0$ for the discrete-time model, or $\mathbb{R}_{\geq 0}$ for the continuous-time model. We identify the population at time t with a (random) counting measure $Z_t = \big(Z_t(x)\big)_{x\in\mathbb{X}}$ on $\mathbb{X}$, where $Z_t(x) := Z_t(\{x\})$ denotes the number of individuals of type $x \in \mathbb{X}$ at time t. Since our population has constant size N, we have $\|Z_t\| = N$ for all times, where $\|Z_t\| := \sum_{x\in\mathbb{X}} Z_t(x) = Z_t(\mathbb{X})$ is the norm (or total variation) of Z_t. We will define a Markov process $(Z_t)_{t\in\mathbb{T}}$, with values in

$$E := \{z \in \{0,\dots,N\}^{|\mathbb{X}|} \mid \|z\| = N\}, \tag{2.10}$$

where $|\mathbb{X}|$ is the number of elements in $\mathbb{X}$. More precisely, we will define a continuous-

time version $(\breve{Z}_t)_{t\geqslant 0}$ representing the Moran model and a discrete-time version $(\hat{Z}_t)_{t\in\mathbb{N}_0}$ representing the Wright-Fisher model. If statements hold for both of them, we will simply write Z instead of $\breve{Z}$ and $\hat{Z}$.

2.2.1 Moran model with recombination

Consider a population of N haploid individuals (gametes) that evolves as follows (see Figure 2.1). Each individual has an exponential lifespan with parameter 1 (this choice of the parameter is without loss of generality; it simply sets the time scale). When an individual dies, it is replaced by a new one as follows. First draw a partition $\mathcal{A}$ according to the recombination distribution $\{r_{\mathcal{A}}\}_{\mathcal{A}\in\mathbb{P}_{\leqslant 2}(S)}$. Then draw $|\mathcal{A}|$ parents from the population (the parents may include the individual that is about to die), uniformly and with replacement, where $|\mathcal{A}|$ is the number of parts in $\mathcal{A}$. If $\mathcal{A}=\{A_1,A_2\}$, the offspring inherits all sites in A_1 from the first and all sites in A_2 from the second parent as described in Section 1.3.1. If $|\mathcal{A}|=1$ (and thus $\mathcal{A}=\{S\}$), the offspring is a full copy of a single parent (again chosen uniformly from all individuals); this is called a *(pure) resampling* event. All events are independent of each other.

It may seem biologically more realistic to draw two parents *without* replacement. However, assuming sampling *with* replacement entails significant simplifications and yields the same process as sampling without replacement with a slight change in the recombination distribution. More precisely, since drawing the same individual twice means that the offspring is a full copy of this single parent, our process agrees (in distribution) with the analogous process without replacement if $r_{\mathcal{A}}$ is replaced by $r_{\mathcal{A}}(N-1)/N$ for all $\mathcal{A}\in\mathbb{P}_2(S)$, and $r_{\{S\}}$ is set accordingly.

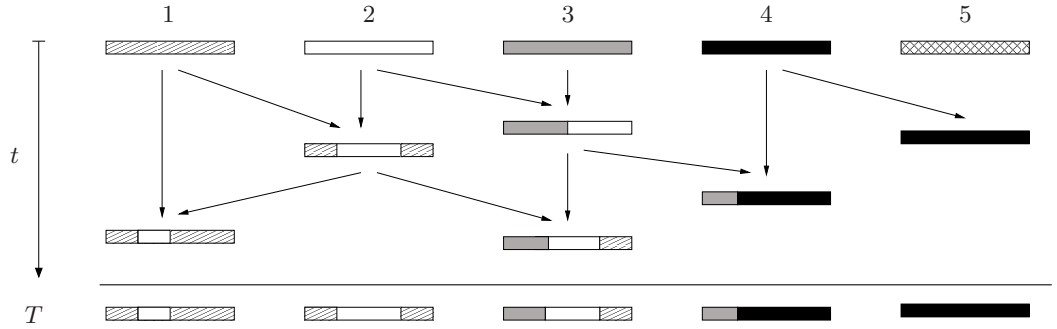


Figure 2.1. One possible realisation of the Moran model with recombination forward in time with $N=5$. For example, in the first event, individual 3 dies and is replaced by a recombined copy of individuals 2 and 3.

Since all individuals die at rate 1, the population loses type-y individuals at rate $\breve{Z}_t(y)$. Each loss is replaced by a new individual, which is sampled uniformly from $R_{\mathcal{A}}(\breve{Z}_t)$ with probability $r_{\mathcal{A}}$, $\mathcal{A}\in\mathbb{P}_{\leqslant 2}(S)$. Therefore, when $\breve{Z}_t=z$, the transition to $z+\delta_x-\delta_y$ occurs with rate

$$\lambda(z;\,y,\,x) := \sum_{\mathcal{A}\in\mathbb{P}_{\leqslant 2}(S)} r_{\mathcal{A}}\big(R_{\mathcal{A}}(z)\big)(x)\,z(y). \tag{2.11}$$

The summand for $\mathcal{A}=\mathbf{1}$ corresponds to pure resampling, whereas all other summands involve recombination. Note that λ includes 'silent transitions' ($x=y$).

Definition 2.1 (Moran model with recombination)**.** The Moran model with recombination is the continuous-time Markov chain $(\breve{Z}_t)_{t\geqslant 0}$ with state space E from (2.10) and generator matrix Λ with nondiagonal elements

$$\Lambda(z,\, z+w) = \sum_{\substack{x,y\in\mathbb{X}\\ \delta_x-\delta_y=w}} \lambda(z;\, y,\, x), \quad w\neq 0,$$

for $z\in E,\ w\in E-z$ (where $E-z := \{v \mid z+v\in E\}$) and $\Lambda(z,\, z) = -\sum\limits_{\substack{v\in E-z\\ v\neq 0}} \Lambda(z, z+v)$.

The model may alternatively be formulated in terms of reproducing individuals rather than dying individuals, as follows. Each individual reproduces at rate 1 and picks a partition $\mathcal{A}\in\mathbb{P}_{\leqslant 2}(S)$ according to the recombination distribution. If $\mathcal{A}\in\mathbb{P}_2(S)$, the reproducing individual contributes the sites in one of the blocks in $\mathcal{A}$ and picks a random partner that contributes the sites in the other block to the offspring. If $\mathcal{A}=\mathbf{1}$, the reproducing individual contributes all sites. The offspring pieced together in this way replaces a uniformly chosen individual from the population. In this formulation, which was used in former descriptions of the model [8, 12] and which is closer to the spirit of the deterministic recombination model, an offspring of type x is created at rate $Nr_{\mathcal{A}}(R_{\mathcal{A}}(\breve{Z}_t))(x)$ and replaces an individual of type y with probability $\breve{Z}_t(y)/N$. This explains the different normalisation factor of the recombinator mentioned in Section 1.3.1. The resulting transition rates, however, are again those in (2.11).

Remark 2.1. Our Moran model differs from the one described in [8] in two ways. In [8], a *decoupled* formulation of recombination and resampling is given, under which individuals experience the two forces independently of each other. If a resampling event happens, an individual reproduces, inherits its type to the offspring and replaces a random individual in the population (possibly its own parent). If recombination occurs, the reproducing individual chooses a random partner (maybe himself) as well as a partition $\mathcal{A}\in\mathbb{P}_2(S)$ from the recombination distribution. In contrast to our model, the two parents now interchange their genetic material reciprocally to create two mixed type individuals; one inherits the leading part from the first and the trailing part from the second individual, the other one the respective counterparts. The two mixed-type offspring individuals replace their own parents, which decouples the reshuffling of genetic material from the resampling process. ♢

2.2.2 Wright-Fisher model with recombination

The Wright-Fisher model with recombination assumes discrete and nonoverlapping generations. In each generation, all present individuals die out and are replaced by a set of N new individuals sampled from the parental generation according to the following rules: Each individual, independently of all others, picks a partition $\mathcal{A}\in\mathbb{P}_{\leqslant 2}(S)$ according to the recombination distribution $\{r_{\mathcal{A}}\}_{\mathcal{A}\in\mathbb{P}_{\leqslant 2}(S)}$. If $\mathcal{A}=\mathbf{1}$, a single parent is chosen uniformly from the parental generation and the offspring is an unaltered copy of this parent. If $\mathcal{A}\neq\mathbf{1}$, two parents of the former generation are chosen uniformly with replacement to recombine according to $\mathcal{A}$. The offspring is the mixed-type individual described in Section 1.3.1.

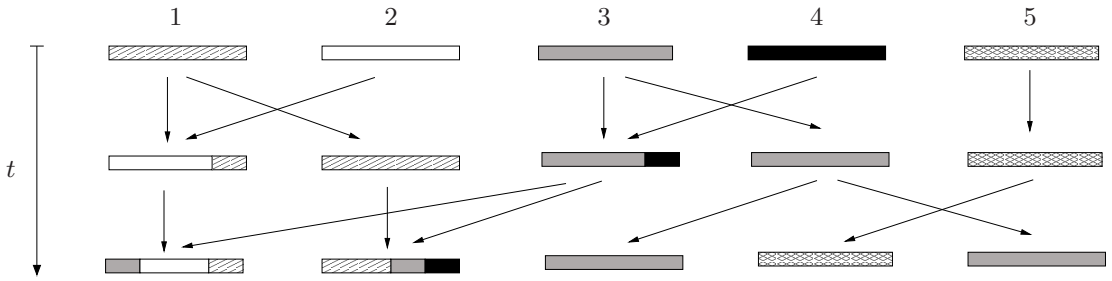


Figure 2.2. One possible realisation of the Wright-Fisher model with recombination.

Within this construction, an individual in generation $t \in \mathbb{N}$ is of type x whenever it first draws $\mathcal{A} = \mathbf{1}$ and then chooses an individual of type x from generation $t-1$ (which happens with probability $r_{\mathbf{1}}\, \widehat{Z}_{t-1}(x)/N = r_{\mathbf{1}}\, (R_{\mathbf{1}}(\widehat{Z}_{t-1}))(x)$) or it draws a partition $\mathcal{A} = \{A_1, A_2\}$ of S and then selects an ordered pair of parents from generation $t-1$ such that the first parent is of type x at all sites in A_1 and the second at all sites in A_2. The latter case happens with probability $r_{\mathcal{A}}\, (R_{\mathcal{A}}(\widehat{Z}_{t-1}))(x)$. Altogether, the probability for an individual to be of type x in generation t is given by $\sum_{\mathcal{A}\in\mathbb{P}_{\leqslant 2}(S)} r_{\mathcal{A}}\, (R_{\mathcal{A}}(\widehat{Z}_{t-1}))(x)$. Since every individual chooses its parent(s) independently from all other individuals with replacement, the N individuals in generation t are obtained by multinomial sampling from generation $t-1$.

Definition 2.2. The Wright-Fisher model with recombination is the discrete-time Markov chain $(\widehat{Z}_t)_{t\in\mathbb{N}_0}$ with state space E from (2.10) and

$$\widehat{Z}_{t+1} \sim \mathrm{Mult}\Big(N, \sum_{\mathcal{A}\in\mathbb{P}_{\leqslant 2}(S)} r_{\mathcal{A}}\, R_{\mathcal{A}}(\widehat{Z}_t)\Big), \quad t \in \mathbb{N}_0.$$

Recall here, that $\widehat{Z}_t = (\widehat{Z}_t(x))_{x\in\mathbb{X}}$. We will see next that different assumptions with respect to time measurement or sampling arrangements loose their effect under a certain time and space scaling when $N \to \infty$; a phenomenon called *universality*. This is well known in the one-locus case, but rarely made explicit if recombination is involved.

2.2.3 Limit processes

For the Wright-Fisher model, explicit formulas for quantities of interest are often impossible to find. Even for the Moran model, where tractable analytic results are available in many cases, the results are rather complex and cumbersome. One therefore aims to find mathematically more tractable processes that approximate the original models in suitable parameter regimes. In the literature, one finds two important classes of such processes. The first is the *deterministic limit*, which emerges when population size tends to infinity without rescaling evolutionary parameters or time. As a consequence, random fluctuations vanish, and expected type frequencies often converge to the solution of the corresponding deterministic process. The second one, the *diffusion limit*, refers to the simultaneous time and space scaling that results when $N \to \infty$ and when time is rescaled by population size. If recombination is present, recombination probabilities are assumed to satisfy $2Nr_{\mathcal{A}} \to \rho_{\mathcal{A}}$, where $\rho_{\mathcal{A}}$ is constant for all $\mathcal{A} \in \mathbb{P}_2(S)$. The diffusion limit maintains random fluctuations but

simultaneously simplifies calculations. It serves as an appropriate limit if population size is large and evolutionary parameters are of the same order as $\frac{1}{N}$ [127]. In contrast, the deterministic limit is well adapted if evolutionary forces are so strong that the additional effect of resampling is negligible. In between the standard diffusion limit and the deterministic limit, there is also a class of intermediate diffusions intended for moderate or rather large recombination rates, see for example [46, 73].

Crossover probabilities between neighbouring base pairs are small enough to satisfy the approximation condition for the diffusion limit. In humans, they are assumed to be of the order 10^{-8} [80]. However, these probabilities increase with increasing distance of base pairs. The largest human chromosome (chromosome 1), for instance, consists of 2.5×10^8 base pairs [51]. It thus depends on the length of the considered chromosome region whether or not recombination probabilities are sufficiently small to ensure proper approximation to the diffusion limit or whether or not they are sufficiently large to ensure proper approximation to the deterministic limit.

Deterministic limit

Consider the family of processes $\big(Z^{(N)}\big)_{N=1,2,\dots}$, with $Z^{(N)} = (Z_t^{(N)})_{t\in\mathbb{T}}$ and where the upper index indicates dependence on population size. For $N \to \infty$ and without any rescaling of the recombination distribution or of time, a *dynamical law of large numbers* applies. For the continuous-time model this reads:

Theorem 2.1. *Let $\big(\check{Z}^{(N)}\big)_{N=1,2,\dots}$ be the family of Moran models as in Definition 2.1, and assume that $\lim_{N\to\infty} \check{Z}_0^{(N)}/N = \check{p}_0$. Then, for every $t \geqslant 0$, one has*

$$\lim_{N\to\infty} \sup_{s\leqslant t} \left| \frac{\check{Z}_s^{(N)}}{N} - \check{p}_s \right| = 0 \text{ with probability 1,} \tag{2.12}$$

where $\check{p}_t$ is the solution to the Cauchy problem of the deterministic recombination equation in continuous time with initial value $\check{p}_0$ given in (2.3).

Note that the probability $r_{\mathcal{A}}$ in the Moran model is multiplied by the unit rate at which each individual reproduces and this way turns into the recombination *rate* $\varrho_{\mathcal{A}}$.

Proof. The proof is an analogue of the proof of Proposition 1 in [8]. As in [8], we want to rely on the law of large numbers from [44, Thm. 11.2.1]. To this end, we first need to ensure that $\big(\check{Z}^{(N)}\big)_{N=1,2,\dots}$ is a *density-dependent family* corresponding to nonnegative functions q_w defined on a subset of $\mathbb{R}_{\geqslant 0}^{|\mathbb{X}|}$. To be precise, we need to ensure that $\check{Z}^{(N)}$ has transition intensities $\Lambda(z, z+w) = N q_w(\frac{z}{N})$, where Λ is the generator matrix of $\check{Z}_t^{(N)}$. These conditions are obviously satisfied if we define

$$q_w(\nu) := \sum_{\substack{x,y\in\mathbb{X}\\ \delta_x-\delta_y=w}} \sum_{\mathcal{A}\in\mathbb{P}_{\leqslant 2}(S)} r_{\mathcal{A}}(R_{\mathcal{A}}(\nu))(x)\,\nu(y), \quad \nu \in \mathcal{P}(\mathbb{X}).$$

Now, let $\breve{p}_t$ be the solution of the deterministic recombination equation with initial value $\breve{p}_0$ stated in (2.3). To conclude the convergence (2.12), it suffices to show that $\breve{p}_t$ solves the differential equation $\frac{\mathrm{d}}{\mathrm{d}t}\breve{p}_t = \sum_w w\, q_w(\breve{p}_t)$ with initial value $\breve{p}_0$. For a fixed type $x \in \mathbb{X}$, in a single transition step, we can either gain an individual of type x, loose one, or the proportion of x individuals does not change. Hence $q_w(\nu(x)) = 0$ if $w \notin \{-1, 0, 1\}$, and we obtain

$$\begin{aligned}\sum_w w\, q_w(\nu(x)) &= \sum_{\mathcal{A}\in\mathbb{P}_{\leqslant 2}(S)} r_{\mathcal{A}}\Bigg(\sum_{\substack{x,y\in\mathbb{X}\\ \delta_x-\delta_y=1}} \big(R_{\mathcal{A}}(\nu)\big)(x)\,\nu(y) - \sum_{\substack{x,y\in\mathbb{X}\\ \delta_x-\delta_y=-1}} \big(R_{\mathcal{A}}(\nu)\big)(x)\,\nu(y)\Bigg)\\ &= \sum_{\mathcal{A}\in\mathbb{P}_{\leqslant 2}(S)} r_{\mathcal{A}}\left[\big(R_{\mathcal{A}}(\nu)\big)(x)\,(1-\nu(x)) - \big(1-\big(R_{\mathcal{A}}(\nu)\big)(x)\big)\,\nu(x)\right]\\ &= \sum_{\mathcal{A}\in\mathbb{P}_{\leqslant 2}(S)} r_{\mathcal{A}}(R_{\mathcal{A}} - \mathbb{1})\big(\nu(x)\big).\end{aligned}$$

We conclude that if $\breve{p}_t$ is the solution of the deterministic recombination equation with initial value $\breve{p}_0$, then $\breve{p}_t$ also solves $\frac{\mathrm{d}}{\mathrm{d}t}\breve{p}_t = \sum_w w\, q_w(\breve{p}_t)$ with initial value $\breve{p}_0$. The claim follows from Theorem 11.2.1 in [44]. □

The convergence in (2.12) is also true for the decoupled Moran model with recombination described in [8] (cf. Remark 2.1). Simulations in [8] show that the expected type frequencies are well approximated by the deterministic solution, even for moderate population sizes ($N = 10^5$).

In discrete time, one has:

Theorem 2.2. *Let $\big(\widehat{Z}^{(N)}\big)_{N=1,2,\ldots}$ be the family of processes corresponding to the Wright-Fisher model as in Definition 2.2. Assume that $\lim_{N\to\infty} \widehat{Z}_0^{(N)}/N = \widehat{p}_0$. Then, for every $t \in \mathbb{N}_0$, one has*

$$\lim_{N\to\infty} \frac{\widehat{Z}_t^{(N)}}{N} = \widehat{p}_t \text{ in probability,} \tag{2.13}$$

where $\widehat{p}_t$ is the solution of the deterministic recombination equation in discrete time with initial value $\widehat{p}_0$ given in (2.7).

Theorem 2.2 is the generalisation of the single-crossover statement in [10, Prop. 1].

Proof. As in [10, Prop. 1], we use induction over t. The claim holds for $t = 0$ by assumption. If the convergence in (2.13) holds for a fixed $t \in \mathbb{N}_0$, then

$$\lim_{N\to\infty} \frac{\widehat{Z}_{t+1}}{N} = \sum_{\mathcal{A}\in\mathbb{P}_{\leqslant 2}(S)} r_{\mathcal{A}}\, R_{\mathcal{A}}(\widehat{p}_t) \text{ in probability}$$

due to Definition 2.2 and the law of large numbers. Since $r_{\mathbf{1}} = 1 - \sum_{\mathcal{A}\in\mathbb{P}_2(S)} r_{\mathcal{A}}$, we obtain that

$$\sum_{\mathcal{A}\in\mathbb{P}_{\leqslant 2}(S)} r_{\mathcal{A}}\, R_{\mathcal{A}}(\widehat{p}_t) = \widehat{p}_t + \sum_{\mathcal{A}\in\mathbb{P}_2(S)} r_{\mathcal{A}}\, R_{\mathcal{A}}(\widehat{p}_t),$$

and the claim follows from (2.1). □

The convergence results in Theorem 2.1 and Theorem 2.2 are true for any finite time t, but in general not for $t \to \infty$. Since there is no mutation in our set-up, the stochastic process is an absorbing Markov chain in which one type will ultimately go to fixation in the long run (we will consider fixation probabilities in Section 4.4). In contrast, the deterministic counterpart never looses any type. The asymptotic behaviour is a product of marginals of the initial population, see (2.4).

Diffusion limit

We now turn to the diffusion limit. As in the deterministic limit, we again let $N \to \infty$, but at the same time speed up time by N with the effect that random fluctuations are still observable. This standard space and time scaling is the basis for most investigations in population genetics. The theory coexisted from the very beginning. Dominating figures are Fisher and Wright. We will only briefly recall the basic definition for diffusions. A general overview can be found for instance in [105, Chap. 7] or [36]. For the specific perspective in population genetics, see [37, Chap. 7,8], [41] or [47, Chap. 4] and references therein.

A (time-homogeneous) Itô *diffusion* is a stochastic process $X_t(\omega) : [0,\infty)\times\Omega \to \mathbb{R}^n$ satisfying for all $t \geqslant s$ a stochastic differential equation of the form

$$\mathrm{d}X_t = b(t, X_t)\,\mathrm{d}t + \sigma(t, X_t)\,\mathrm{d}B_t, \quad X_s = v,$$

where B_t is a m-dimensional Brownian motion, $b : [0,\infty) \times \mathbb{R}^n \to \mathbb{R}^n$ is the *drift coefficient* and $\sigma : [0,\infty) \times \mathbb{R}^n \to \mathbb{R}^{n\times m}$ is the *diffusion coefficient* satisfying

$$|b(v) - b(w)| + |\sigma(v) - \sigma(w)| \leqslant D|v - w|, \quad v, w \in \mathbb{R}^n,$$

for some constant D, where $|\sigma^2| = \sum_{ij} |\sigma_{ij}|^2$ [105, Def. 7.1.1]. If X_t is a diffusion in $\mathbb{R}^n$, then the (infinitesimal) generator $\mathcal{L}$ of X_t is defined as

$$\mathcal{L}f(p) = \lim_{t\downarrow 0} \frac{\mathbf{E}[f(X_t) \mid X_0 = v] - f(v)}{t}, \quad v \in \mathbb{R}^n. \tag{2.14}$$

The *domain* $\mathcal{D}(\mathcal{L})$ of $\mathcal{L}$ is the set of functions $f : \mathbb{R}^n \to \mathbb{R}$ for which the limit (2.14) exists for all $v \in \mathbb{R}^n$ [105, Def. 7.3.1].

In population genetics, one observes that many finite population processes converge to a diffusion process if time is sped up by population size and if population size tends to infinity. Having our population processes in mind, we define the following diffusion process with recombination:

Definition 2.3. The *Wright-Fisher diffusion with recombination* is a process $X = (X_t)_{t\geqslant 0}$ with state space

$$E' := \big\{p = (p_x)_{x\in\mathbb{X}} \in [0,1]^{|\mathbb{X}|} \mid \|p\| = 1\big\}, \tag{2.15}$$

generator

$$\mathcal{L}f(p) := \frac{1}{2}\sum_{x,y\in\mathbb{X}} p_x(\delta_{xy} - p_y)\frac{\partial^2}{\partial p_x \partial p_y} f(p) + \frac{1}{2}\sum_{x\in\mathbb{X}}\sum_{\mathcal{A}\in\mathbb{P}_2(S)} \rho_{\mathcal{A}}(R_{\mathcal{A}} - \mathbb{1})(p_x)\frac{\partial}{\partial p_x} f(p) \tag{2.16}$$

and domain $C^2(E')$, where we abbreviated $p_x := p(x)$ for any type $x \in \mathbb{X}$ and where $C^2(E')$ is the set of all twice differentiable functions on E'.

We will see below that, under the appropriate scaling, this is indeed the right limit process for the population processes we defined in Section 2.2. Note that the representation of the generator in (2.16) differs from other two-locus representations, such as the one from Durrett [37] or Ethier and Griffiths [43], by a factor of two.

Remark 2.2. There is a minor difference in time scaling depending on whether one starts with the discrete-time Wright-Fisher model or with the continuous-time Moran model. The coalescence process (the ancestral process backward in time, see Chap. 3) corresponding to the Moran model runs at twice the speed compared to the Wright-Fisher model. This is a well-known property in the case without recombination. The common way to overcome this problem is to speed up time in the Wright-Fisher model with N and in the Moran model with $\frac{N}{2}$. We will transfer this different scalings to the recombination parameters and assume that $2Nr_{\mathcal{A}} \to \rho_{\mathcal{A}}$ in the Wright-Fisher model, and that $Nr_{\mathcal{A}} \to \rho_{\mathcal{A}}$ in the Moran model, where $\mathcal{A} \in \mathbb{P}_2(S)$, and $\rho_{\mathcal{A}}$ is a constant. ♢

Starting from the Moran model, we obtain the following convergence result in the general multi-locus, multi-allele and multi-crossover setting.

Theorem 2.3. *Let $(\check{Z}_t^{(N)})_{t\geqslant 0}$ be the Moran model from Definition 2.1, and assume that $Nr_{\mathcal{A}} \to \rho_{\mathcal{A}}$ for all $\mathcal{A} \in \mathbb{P}_2(S)$, where $\rho_{\mathcal{A}}$ is a constant. As $N \to \infty$, we have the following convergence in distribution:*

$$\Big(\frac{1}{N}\check{Z}_{Nt/2}^{(N)}\Big)_{t\geqslant 0} \longrightarrow (X_t)_{t\geqslant 0},$$

where $(X_t)_{t\geqslant 0}$ is the Wright-Fisher diffusion with recombination from Definition 2.3.

In the case of two loci and two alleles, starting from the Wright-Fisher model, the convergence result in Theorem 2.3 goes back to Ohta and Kimura [103, 104]; see also: [37, Chap. 8.2] for a modern exposition. Two loci with an arbitrary (but finite) number of alleles are treated in [77]. Griffiths et al. [59] give an expression for the generator of the diffusion process in the multi-locus, multi-allele, single-crossover case (including mutation) but do not prove convergence. To the best of our knowledge, there is no proof for the convergence of the rescaled Moran model with recombination to the Wright-Fisher diffusion with recombination for more than two loci. We thus include the proof here.

Proof. Consider the Moran model $\check{Z}^{(N)} = (\check{Z}_t^{(N)})_{t\geqslant 0}$ from Definition 2.1. Let z be a realisation of $\check{Z}^{(N)}$ and $\check{Y}^{(N)}$ be the rescaled process $\check{Y}^{(N)} = (\check{Y}_t^{(N)})_{t\geqslant 0}$ with $\check{Y}_t^{(N)} = \frac{1}{N}\check{Z}_{Nt/2}^{(N)}$. $\check{Y}^{(N)}$ is a Markov process with state space

$$E^{(N)} := \Big\{q \in \Big\{0, \frac{1}{N}, \frac{2}{N}, \dots, 1\Big\}^{|\mathbb{X}|} \Big| \, \|q\| = 1\Big\}. \tag{2.17}$$

Due to the rescaling of time, the generator of $\check{Y}^{(N)}$ takes the form

$$(\check{\mathcal{L}}^{(N)} f)\Big(\frac{z}{N}\Big) = \frac{N}{2} \sum_{\substack{x,y\in\mathbb{X}\\ x\neq y}} \lambda\Big(\frac{z}{N}; y, x\Big)\Big[f\Big(\frac{z}{N} + \frac{\delta_x}{N} - \frac{\delta_y}{N}\Big) - f\Big(\frac{z}{N}\Big)\Big], \quad z \in E,$$

where $\lambda(z; y, x)$ is defined as in (2.11) and δ_x is the point measure on x.

Let $f \in C^3([0,1])$, and use the Taylor expansion of f around $\frac{z}{N} + \frac{\delta_x}{N} - \frac{\delta_y}{N}$ up to second order to obtain for any $q = (q_x)_{x\in\mathbb{X}} \in E^{(N)}$:

$$\begin{aligned}(\check{\mathcal{L}}^{(N)}f)(q) = \frac{N}{2}\sum_{\substack{x,y\in\mathbb{X}\\ x\neq y}}\sum_{\mathcal{A}\in\mathbb{P}_{\leqslant 2}(S)} r_{\mathcal{A}}\, R_{\mathcal{A}}(q_x)\, q_y \Bigg[\frac{\partial}{\partial q_x}f(q) - \frac{\partial}{\partial q_y}f(q) \\ + \frac{1}{2N}\left(\frac{\partial^2}{\partial^2 q_x}f(q) - 2\frac{\partial^2}{\partial q_x \partial q_y}f(q) + \frac{\partial^2}{\partial^2 q_y}f(q)\right) + NB_3^{(N)}(q)\Bigg],\end{aligned} \tag{2.18}$$

where $B_3^{(N)}(q)$ is an error term of the form

$$B_3^{(N)}(q) = \frac{1}{N^3}\sum_{\substack{k=(k_x,k_y)\\ k_x+k_y=3}} (-1)^{k_y} A_k\Big(q + \frac{\delta_x}{N} - \frac{\delta_y}{N}\Big), \tag{2.19}$$

with $\sup_{q\in E^{(N)}} |A_k(q)| \leqslant C$, for some constant C. We can now separate the sums in (2.18) and use the identities $\sum_{y\neq x} q_y = 1 - q_x$ and $\sum_{y\neq x} R_{\mathcal{A}}(q_y) = 1 - R_{\mathcal{A}}(q_x)$ for fixed $x \in \mathbb{X}$ and $\mathcal{A} \in \mathbb{P}_{\leqslant 2}(S)$. Rearranging yields

$$\begin{aligned}(\check{\mathcal{L}}^{(N)}f)(q) &= \frac{N}{2}\sum_{\mathcal{A}\in\mathbb{P}_{\leqslant 2}(S)} r_{\mathcal{A}}\Bigg[\sum_{x\in\mathbb{X}}\Bigg(\Big((1-q_x)R_{\mathcal{A}}(q_x) + (R_{\mathcal{A}}(q_x)-1)q_x\Big)\frac{\partial}{\partial q_x}f(q) \\ &\qquad + \frac{1}{2N}\Big((1-q_x)R_{\mathcal{A}}(q_x) + (1-R_{\mathcal{A}}(q_x))\,q_x\Big)\frac{\partial^2}{\partial^2 q_x}f(q)\Bigg) \\ &\qquad - \frac{1}{N}\sum_{\substack{x,y\in\mathbb{X}\\ x\neq y}} R_{\mathcal{A}}(q_x)\, q_y\, \frac{\partial^2}{\partial q_x \partial q_y}f(q) + NB_3^{(N)}(q)\Bigg] \\ &= \frac{N}{2}\sum_{\mathcal{A}\in\mathbb{P}_{\leqslant 2}(S)} r_{\mathcal{A}}\Bigg[\sum_{x\in\mathbb{X}}\Big(R_{\mathcal{A}}(q_x) - q_x\Big)\frac{\partial}{\partial q_x}f(q) \\ &\qquad - \frac{1}{N}\sum_{x,y\in\mathbb{X}}\Big(R_{\mathcal{A}}(q_x)\,q_y - \frac{\delta_{x,y}}{2}(R_{\mathcal{A}}(q_x)+q_y)\Big)\frac{\partial^2}{\partial q_x\partial q_y}f(q) + NB_3^{(N)}(q)\Bigg].\end{aligned}$$

Separating the $\mathcal{A} = \mathbf{1}$ term leads to

$$\begin{aligned}(\check{\mathcal{L}}^{(N)}f)(q) &= \frac{r_{\mathbf{1}}}{2}\sum_{x,y\in\mathbb{X}}\Big(q_x(\delta_{x,y}-q_y)\frac{\partial^2}{\partial q_x\partial q_y}f(q) - NB_3^{(N)}(q)\Big) \\ &\quad + \frac{N}{2}\sum_{\mathcal{A}\in\mathbb{P}_2(S)} r_{\mathcal{A}}\Bigg[\sum_{x\in\mathbb{X}}(R_{\mathcal{A}}(q_x)-q_x)\frac{\partial}{\partial q_x}f(q) \\ &\qquad - \frac{1}{N}\sum_{x,y\in\mathbb{X}}\Big(R_{\mathcal{A}}(q_x)\,q_y - \frac{\delta_{x,y}}{2}(R_{\mathcal{A}}(q_x)+q_y)\Big)\frac{\partial^2}{\partial q_x\partial q_y}f(q) + NB_3^{(N)}(q)\Bigg],\end{aligned}$$

where we used that $R_{\mathbf{1}}(q) = q$ and $\|q\| = 1$.

If $N \to \infty$ and $Nr_{\mathcal{A}} \to \rho_{\mathcal{A}}$ for all $\mathcal{A} \in \mathbb{P}_2(S)$, we obtain together with (2.19) that

$$\lim_{N\to\infty} \sup_{q\in E^{(N)}} |\check{\mathcal{L}}^{(N)} f(q) - \mathcal{L}f(q)| = 0$$

for every $f \in C^3([0,1])$. Note that $r_{\mathbf{1}} \to 1$ under the prescribed scaling. We conclude the convergence of the rescaled Moran model to the Wright-Fisher diffusion with recombination by [44, Thm. 1.6.1 & Thm. 4.2.11 & Thm. 8.2.1]. □

Remark 2.3. In the absence of recombination ($\rho_{\mathcal{A}} = 0$ for all $\mathcal{A} \in \mathbb{P}_2(S)$) and restricted to two types, one rediscovers from Theorem 2.3 the convergence of the two-type Moran model (without recombination) to the two-type Wright-Fisher diffusion (without recombination). The Wright-Fisher diffusion (without recombination) $X = (X_t)_{t\geqslant 0}$ is a process on $[0,1]$ with generator

$$\mathcal{L}f(p) = \frac{1}{2}\, p(1-p) f''(p), \quad p \in [0,1]. \tag{2.20}$$

The Wright-Fisher diffusion (without recombination) is arguably the most famous diffusion process in population genetics. The convergence to the Wright-Fisher diffusion also holds for the time-scaled, two-type Wright-Fisher model (without recombination). The convergence is then with respect to the Skorokhod topology of $\mathbb{D}([0,\infty),[0,1])$, the set of all cádlág functions on $[0,\infty)$ with values in $[0,1]$ (see for instance [44, Chap. 3]). ♢

In discrete time, if we speed up time by N and if we assume $2Nr_{\mathcal{A}} \to \rho_{\mathcal{A}}$ for every partition $\mathcal{A} \in \mathbb{P}_2(S)$, it is also very natural to expect that the Wright-Fisher model with recombination converges to the Wright-Fisher diffusion with recombination with respect to the Skorokhod topology of $\mathbb{D}([0,\infty),[0,1])$. The precise proof is beyond the scope of this thesis, but let us give at least an heuristic argument here.

Let f be twice differentiable and $(X_t)_{t\geqslant 0}$ be a multi-dimensional Itô diffusion. The generator takes the form

$$\mathcal{L}f(v) = \frac{1}{2}\sum_{i,j} a_{ij}(v) \frac{\partial^2}{\partial v_i \partial v_j} f(v) + \sum_i b_i(v) \frac{\partial}{\partial v_i} f(v), \quad v \in \mathbb{R}^n, \tag{2.21}$$

where

$$b(v) = \lim_{h\downarrow 0} \mathbf{E}[X_{t+h} - X_t \,|\, X_t = v], \qquad a(v) = \lim_{h\downarrow 0} \mathbf{E}[(X_{t+h} - X_t)^2 \,|\, X_t = v]$$

represent the infinitesimal mean and the infinitesimal covariance matrix [105, Chap. 7.3]. Let $\big(\widehat{Z}_t^{(N)}\big)_{t\in\mathbb{N}_0}$ denote the Wright-Fisher model with recombination from Definition 2.2. If the current state is $\widehat{Z}_t^{(N)} = z$, then $\widehat{Z}_{t+1}^{(N)}$ follows a multinomial distribution with parameter N and $\sum_{\mathcal{A}\in\mathbb{P}_{\leqslant 2}(S)} r_{\mathcal{A}}\, R_{\mathcal{A}}(z)$. According to (2.21), the generator of the time-scaled process $\hat{Y}^{(N)} = \big(\hat{Y}_t^{(N)}\big)_{t\geqslant 0}$, with $\hat{Y}_t^{(N)} = \frac{1}{N}\widehat{Z}_{\lfloor Nt\rfloor}^{(N)}$, is of the form

$$\hat{\mathcal{L}}^{(N)} f\Big(\frac{z}{N}\Big) = \frac{1}{2}\sum_{x,y\in\mathbb{X}} a_{x,y}\Big(\frac{z}{N}\Big) \frac{\partial^2}{\partial z_x \partial z_y} f\Big(\frac{z}{N}\Big) + \sum_{x\in\mathbb{X}} b_x\Big(\frac{z}{N}\Big) \frac{\partial}{\partial z_x} f\Big(\frac{z}{N}\Big), \quad z \in E,$$

with

$$b\Big(\frac{z}{N}\Big) = N\,\mathbf{E}\bigg[\frac{\widehat{Z}^{(N)}_{t+\Delta}}{N} - \frac{z}{N}\,\bigg|\,\frac{\widehat{Z}^{(N)}_{t}}{N} = \frac{z}{N}\bigg],$$
$$a_{xx}\Big(\frac{z}{N}\Big) = N\,\mathrm{Cov}\bigg[\frac{\widehat{Z}^{(N)}_{t+\Delta}(x)}{N}, \frac{\widehat{Z}^{(N)}_{t+\Delta}(x)}{N}\,\bigg|\,\frac{\widehat{Z}^{(N)}_{t}}{N} = \frac{z}{N}\bigg],$$
$$a_{xy}\Big(\frac{z}{N}\Big) = N\,\mathrm{Var}\bigg[\frac{\widehat{Z}^{(N)}_{t+\Delta}(x)}{N}, \frac{\widehat{Z}^{(N)}_{t+\Delta}(y)}{N}\,\bigg|\,\frac{\widehat{Z}^{(N)}_{t}}{N} = \frac{z}{N}\bigg], \quad x \neq y,$$

where $\Delta := \frac{1}{N}$ and where we abbreviated $z_x := z(x)$. Using the well-known expressions for the expectation, variance and covariance of the multinomial distribution, a straightforward calculation yields for any $q \in E^{(N)}$ and $x, y \in \mathbb{X}$ that

$$b(q) = \sum_{\mathcal{A}\in\mathbb{P}_2(S)} N r_{\mathcal{A}}\,(R_{\mathcal{A}} - \mathbb{1})(q), \qquad a_{xx}(q) = q_x(1-q_x) + \sum_{\mathcal{A}\in\mathbb{P}_2(S)} r_{\mathcal{A}}\,R_{\mathcal{A}}(q_x) + \mathcal{O}\Big(\frac{1}{N}\Big),$$
$$a_{x,y}(q) = -q_x q_y + \sum_{\mathcal{A}\in\mathbb{P}_2(S)} r_{\mathcal{A}}\,R_{\mathcal{A}}(q_x) + \mathcal{O}\Big(\frac{1}{N}\Big), \quad x \neq y \tag{2.22}$$

where we again used the abbreviation $q_x := q(x)$ for all $x \in \mathbb{X}$. Direct comparison of the coefficients in (2.22) with the generator of the Wright-Fisher diffusion with recombination from (2.16) shows that if $2Nr_{\mathcal{A}} \to \rho_{\mathcal{A}}$ for all $\mathcal{A} \in \mathbb{P}_2(S)$ and $N \to \infty$, the convergence of $(\hat{Y}^{(N)}_t)_{t\geqslant 0}$ to the Wright-Fisher diffusion with recombination seems reasonable.

3 Backward time: Ancestral recombination process

Evolution takes place on a long time scale. In the majority of cases, evolutionary changes are not observable within a researchers lifetime. For population genetics, it is thus natural to shift perspective from the forward, prospective, view to a backward, retrospective, view and use the availability of data today to infer what happened in the past. The retrospective view comes with the great advantage that only those genes or individuals matter that contribute to today's sample. The idea of looking backward in time goes back to Malécot in 1948 and ever since turned into an indispensable building block of population genetics. Starting with a sample at present, one first constructs the ancestral process, or *coalescent process*, that describes the dispersal of genetic material of a sample at present to the ancestors backward in time. For an introduction into the topic, we refer the reader to [65, 128]. A detailed and comprehensive investigation of the mathematical theory can be found in [15].

In this chapter, we investigate the ancestral process of a sample of individuals by arguing on the grounds of the underlying Moran model or Wright-Fisher model with recombination. We start with an introduction to ancestral processes without recombination. Thereafter, we summarise known approaches in the case with recombination such as the *ancestral recombination graph* (ARG) in its different versions. We then define a marginal version of the ancestral process with recombination, study the process in detail and investigate some scaling limits in the end. In Chapter 4, we will see that the marginalised process (in continuous time) is indeed the right dual process for the Moran model with recombination.

Imagine we let a population process $(Z_t)_{t\in\mathbb{T}}$ *without* recombination (Wright-Fisher or Moran model) run until some time t *forward* in time; here $\mathbb{T}$ again either represents $\mathbb{N}_0$ for the discrete-time model or $\mathbb{R}_{\geqslant 0}$ for the continuous-time model. At time $t' < t$, we draw a sample of individuals $x_1, \dots, x_m$, $m \leqslant N$, from $Z_{t'}$ without replacement. Given $(Z_s)_{0\leqslant s\leqslant t'}$, we can construct the *ancestral partitioning process* $(\Pi_t^{N,m})_{t\in\mathbb{T}}$ *backward* in time, which is a process on the set of partitions of $\{1, \dots, m\}$. Two elements i and j, $i, j \in \{1, \dots, m\}$, are in the same block at time t if and only if x_i and x_j have a common ancestor at time $-t$. The graphical picture of the ancestral partitioning process is a tree (called *genealogical tree*) that starts with m lines. Lines merge when individuals find their common ancestor, see Figure 3.1.

In the finite discrete-time case, it might happen that more than two individuals find their common ancestor in the previous generation. From the forward perspective, this means that one individual gave birth to more than two individuals that are ancestors of the current

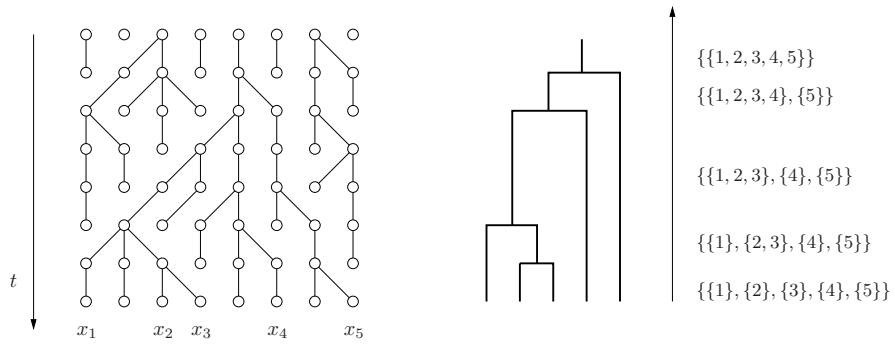


Figure 3.1. Left: Population evolving under the Wright-Fisher model without recombination (no types are shown). At present, a sample $\{x_1, \dots, x_5\}$ of individuals is taken. Right: Corresponding genealogical tree and ancestral partitioning process for the sampled individuals.

sample. The limit process that emerges if $N \to \infty$ and if time is rescaled by N (discrete model) or $\frac{N}{2}$ (continuous model) is the *Kingman's m-coalescent* $(\Pi_t^m)_{t \geqslant 0}$, introduced by Kingman [81, 82] in 1982. The Kingman's m-coalescent is a continuous-time Markov chain with state space $\mathbb{P}(\{1, \dots, m\})$ and initial value $\Pi_0^m = \{\{1\}, \dots, \{m\}\}$. If the current state is σ, any ordered pair of blocks in σ merges at rate 1, see for example [15, Chap. 2.1].

3.1 Ancestral process with recombination

From the backward perspective, recombination events refer to a splitting of genetic material to two ancestors. Resampling events refer to the situation that two sequences (or parts of sequences) find their common ancestor. Due to the branching events generated by recombination, there is no single genealogical tree that describes the ancestral process with recombination (ARP). In fact, different sites may have different ancestries. The complete graphical picture is therefore a graph rather than a tree, see Figure 3.2. This adds a challenging layer to usual coalescence analysis.

Each single site is nonetheless still inherited from a single parent. Any genealogy with respect to a single site can thus still be represented by a genealogical tree (called *local tree*). Local trees for different sites share edges in the graph and may coincide with local trees for other sites depending on the ordering of the recombination events. Obviously, the family of local trees is not independent.

Most coalescent processes with recombination assume the diffusion limit ($N \to \infty$, $t \to Nt$, $2Nr_{\mathcal{A}} \to \rho_{\mathcal{A}}$ const., see Sect. 2.2.3). They are best known under the keyword *ancestral recombination graph* (ARG). For overviews see [65, Chap. 5], [37, Chap. 3.4] or [128, Chap. 7.2]. We, on the other hand, decided to start with the finite model first, which allows to consider the well-known scaling limits in an efficient way in the end. In the diffusion limit, recombination and coalescence act in isolation. In the *finite* ancestral recombination process (ARP), however, additional mixed recombination-coalescence events will arise (cf. Fig. 3.2). From the forward perspective this means that at least one of the parents that contributed parts to the recombined offspring also contributed genetic material to some other individual that

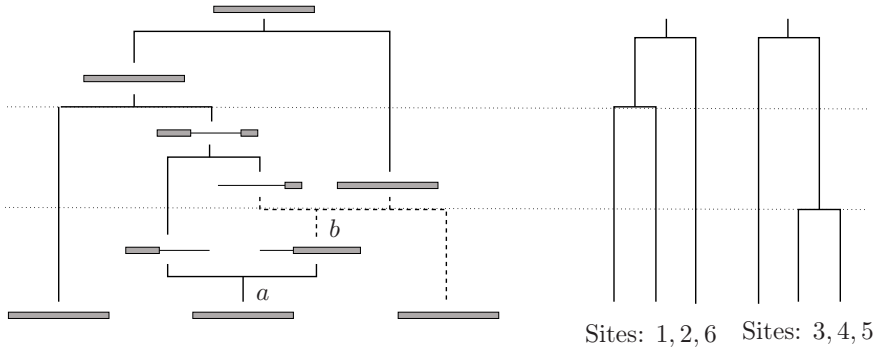


Figure 3.2. Left: A realisation of the full ancestral recombination process, starting from $m = 3$ individuals with six sites; ancestral material is shaded, nonancestral material is indicated by thin horizontal lines. The first recombination event is a crossover between site 2 and 3. The mixed recombination-coalescence event indicated by dashed lines can only appear in the finite ancestral recombination process. In the diffusion limit, and thus in the ARG, recombination and coalescence act in isolation. Right: the family of local trees. The local trees for the sites 1,2,6 and the local trees for the sites 3,4,5 coincide.

is an ancestor of the sample. Since the probability of such an event is of order $\frac{1}{N^2}$, these mixed-type events vanish in the diffusion limit.

Before we start with the finite model, let us summarise the main achievements obtained in the diffusion limit.

3.1.1 Ancestral recombination graph

In 1983, shortly after the discovery of the Kingman coalescent, Hudson incorporated single-crossover recombination into the ancestral process. He gave an efficient algorithm to construct the genealogy of a sample of individuals evolving under recombination if time and space are rescaled according to the standard diffusion limit; first in the two-locus case [68], later generalised to a multi-locus model with selection [70, 79]. Griffiths and Marjoram picked up his idea and elaborated the corresponding graphical picture, the *ancestral recombination graph* (ARG) [60]. Today, the ARG is the standard genealogical approach for models with recombination, but many different notions of 'ARG' are in use. Some are two-locus versions [58, 68], some multi-locus ones [17, 79], and the majority is based on a continuous-sequence assumption ($n \to \infty$, see, e.g. [37, Chap. 3.4], [30, 60, 61, 79, 95, 131]). In any case, let us stick to the usual convention here that the ARG is based on the diffusion limit.

Translating the continuous-sequence algorithm of Hudson to an n-locus, multi-crossover algorithm leads to the following description of the process: Start with m sequences and follow their ancestry backward. If there are currently k sequences, the time to a coalescence event is exponentially distributed with parameter $\frac{k(k-1)}{2}$; the time to a recombination event is exponentially distributed with parameter $k\rho$, where $\rho = \sum_{\mathcal{A}\in\mathbb{P}_2(S)} \rho_{\mathcal{A}}$ is the total (population-scaled) recombination rate. If a coalescence event happens, choose two sequences among the current sequences to merge and decrease the number of sequences by one. If a recombination event appears, draw a random sequence and a partition $\mathcal{A}$ from the recombination

distribution, split the sequence into two and distribute the ancestral material between the new sequences according to $\mathcal{A}$. The number of sequences is increased by one. Stop if there is only one sequence left. At that time, all parts of the sampled individuals found their most recent common ancestor (MRCA). Since the birth rate with respect to the number of lines is linear and the death rate is quadratic, the MRCA always exists. Over the decades, various properties of the ARG, such as the waiting time to the last MRCA of the sample [60, 61, 133], the number of recombination events [43, 60, 61, 69], the number of segregating sites [68], the distribution of the size of the ARG [42] or the existence of a sampling formula have been studied [17, 75, 76, 77]. Dominating figures are Griffiths, Hudson, Kaplan, Wiuf and Hein.

In the original, very straightforward algorithm of the ARG described above, certain time consuming silent events are included, namely those recombination events that happen outside of ancestral or trapped material. *Trapped material* is nonancestral material enclosed between two parts of ancestral material. Events happening in nonancestral material that is not trapped, neither affect the partitioning of ancestral material, nor the family of local trees. Moreover, if an event in such a regions occurs, the respective lineage will split into two lineages of which one does not share any genetic material with the sample. In order to keep the number of lines and events as small as possible, these silent events may thus be removed by increasing the memory capacity of the algorithm. Such modified algorithms, where all nonancestral lines are excluded and where for every sequence the information about the (continuous) region spanned by ancestral material is stored, belong to the class of *reduced* ARGs and are for example investigated in [95, 131]. A latest example in this class of modified algorithms is the sequential coalescent with recombination model (SCRM). The SCRM improved the algorithm studied in [95, 132] with regard to accuracy and efficiency [122].

3.2 Marginal ancestral recombination process

Let us now come back to the *finite* ancestral recombination process. The state space of the ARP is enormous, even for small sample sizes. If we want to describe the full ancestral partitioning process under recombination for a sample of m individuals, we need to trace back the ancestral material on all sequences at all times, starting with the initial state $\bigtimes_{i=1}^{m} S$, where $S = \{1, \dots, n\}$ is the set of considered sites. If there are currently k sequences, the state of the process is a product of k subsets of $\{1, \dots, n\}$. Writing down all transitions is beyond the scope of this thesis. We will, in contrast, investigate a simplified version of the ancestral process that only aims at reduced information of the full process. Namely, at every time point, we only consider *one* set S of sites, distributed along the individuals, where each site is considered in one individual only. To this end, let $\mathcal{A} = \{A_1, A_2, \dots, A_m\}$ be a partition of S with $m \leqslant \min\{n, N\}$. Start with a sample of m individuals from the present population, and follow back the ancestry of the sites in A_1 in the first individual, in A_2 in the second individual, ..., in A_m in the m'th individual, without considering any other sites and any other individuals, as shown in Figure 3.3. The result may be viewed as a *marginalised* version of the ancestral recombination process. In the diffusion limit, this marginalised version will turn into a marginalised version of the reduced ARG that starts with a sample of size m.

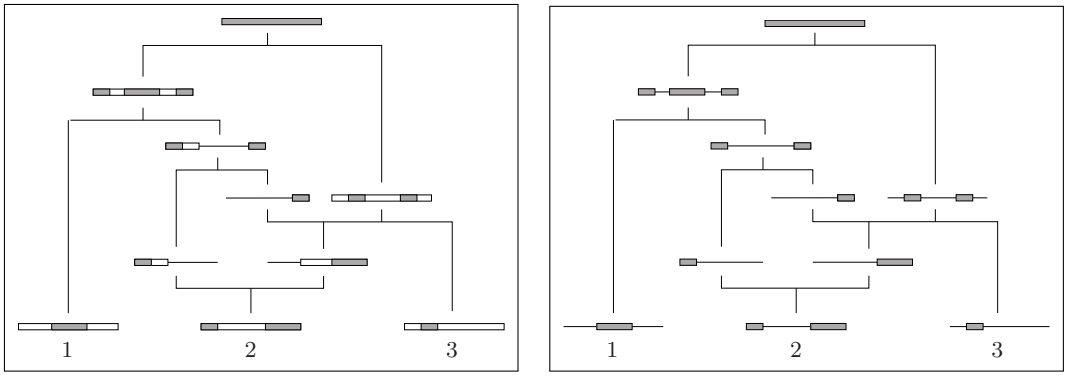


Figure 3.3. The marginalised version corresponding to the ARP in Figure 3.2, in which we follow back the ancestry of the blocks of $\mathcal{A} = \{\{1, 5, 6\}, \{2\}, \{3, 4\}\}$ (shaded), that is, block A_i is sampled in individual number i, $1 \leqslant i \leqslant 3$. Material that is ancestral to the sampled individuals, but not to the blocks considered, is shown as open rectangles (left). But since this is not traced back, it can be treated in the same way as material nonancestral to the sampled individuals (right). Consequently, the sample will finally consist of the blocks of the partition only.

A configuration of types at present can be obtained in a three-step procedure (see Figure 3.4). First, we run a partitioning process $(\Sigma_t)_{t\in\mathbb{T}}$ on $\mathbb{P}(S)$ backward in time, starting at a given initial partition Σ_0 with $|\Sigma_0| = m$. The process $(\Sigma_t)_{t\in\mathbb{T}}$ describes the partitioning of sites into parental individuals at time t (independent of the types) and will be considered in detail in the next section. In the second step, a letter is assigned to each site of S at time t in the following way. For every part of Σ_t, pick an individual from the initial population (without replacement) and copy its letters to the sites in the block considered. For illustration, also assign a colour to each block, thus indicating different parental individuals. In the last step, the letters and colours are propagated downward (i.e. forward in time) according to the realisation of $(\Sigma_t)_{t\in\mathbb{T}}$ laid down in the first step. A similar construction was used in the ancestral process by Baake and von Wangenheim [10] restricted to a sample of size 1 (i.e. start with $\Sigma_0 = \mathbf{1} = \{S\}$) and in the deterministic limit.

Obviously, the marginal approach admits only reduced access to information of interest. This might be problematical for some quantities, such as (co)variances for instance. Nevertheless, we will see in Section 4.3 and Section 4.4 that the marginal ansatz is rich enough to study many interesting objects, such as fixation probabilities or the time evolution of expected correlations of sites (called linkage disequilibria).

3.2.1 The partitioning process

The partitioning process $(\Sigma_t)_{t\in\mathbb{T}}$ is a Markov chain on $\mathbb{P}(S)$ which describes the dispersal of sites $S = \{1, \ldots, n\}$ to ancestral individuals backward in time. Sites within one block correspond to the same individual, whereas different blocks refer to different individuals which are not further specified or labelled. Clearly, $|\Sigma_t|$ is the number of ancestral individuals at time t. Since there is a one-to-one relationship between the individuals and the blocks of the partition, we may identify individuals with the ancestral material they carry. The process $(\Sigma_t)_{t\in\mathbb{T}}$ consists of a mixture of splitting (S) and coalescence (C) events.

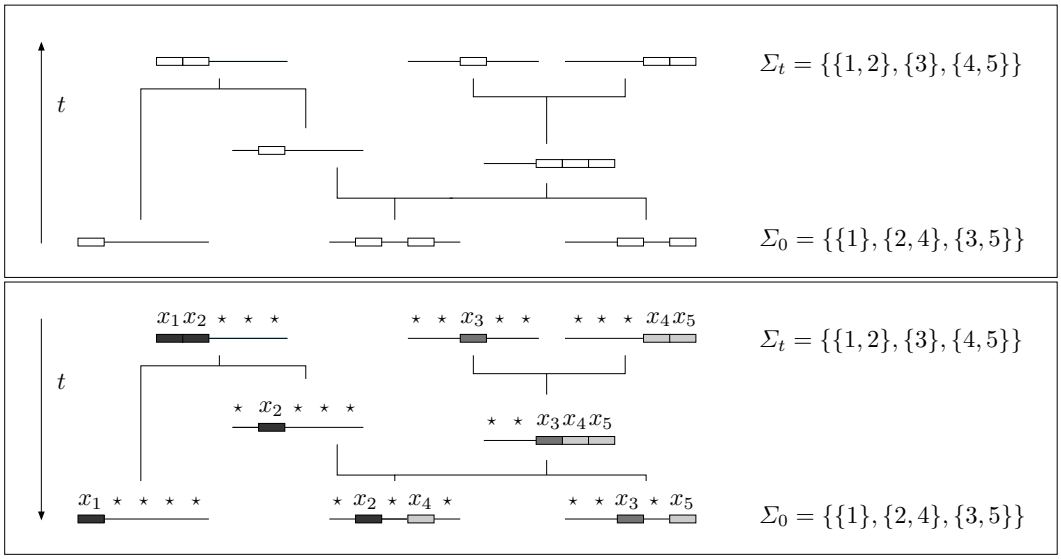


Figure 3.4. Construction of one possible ancestry of a collection of sites that correspond to the initial partition $\Sigma_0 = \{\{1\},\{2,4\},\{3,5\}\}$. The upper panel shows the partitioning process (backward in time). In the lower panel, letters and colours are assigned to each block of Σ_t and propagated downward (forward in time).

In a previous paper [40], we described the process for single crossovers only. We present here a generalised partitioning process that allows multi-crossover recombination.

Assume that $\Sigma_t = \mathcal{A}$ and consider a block $U \in \mathcal{A}$. If U is ordered in S, i.e. if U is of the form $U = \{x \in S : \min(U) \leqslant x \leqslant \max(U)\}$, every recombination event within U will split up the ancestral material. If U is unordered in S, this means that there is trapped material enclosed between ancestral regions and that more than one recombination event may lead to the same splitting result. In the multi-crossover case, moreover, not every recombination event will lead to a decomposition of ancestral material. To see this, let U be an unordered subset of S, $|S| > 2$, with $U = \{1, n\}$. Any recombination event with an even number of crossovers will not break up the connection between the sites 1 and n, whereas every event with an odd number of crossovers does. In the following, we will work with *marginal recombination probabilities* on subsystems that capture all mentioned case distinctions. Let us define

$$r_{\mathcal{B}}^{U} := \sum_{\substack{\mathcal{A}\in\mathbb{P}_{\leqslant 2}(S)\\ \mathcal{A}|_U=\mathcal{B}}} r_{\mathcal{A}}^{S}, \quad \mathcal{B} \in \mathbb{P}_{\leqslant 2}(U), \tag{3.1}$$

where $r_{\mathcal{A}}^{S} = r_{\mathcal{A}}$ and $\mathcal{A}|_U$ is the partition in $\mathbb{P}(U)$ that consists precisely of all nonempty sets of the form $A_i \cap U$, see Section 1.2.1. Obviously $\sum_{\mathcal{A}\in\mathbb{P}_{\leqslant 2}(U)} r_{\mathcal{A}}^{U} = 1$, and the only recombination parameter for $|U| = 1$ is $r_{\mathbf{1}}^{U} = 1$. Technically, the superscript can be dispensed with since $U = \cup_{i=1}^{|\mathcal{B}|} B_i$ if $\mathcal{B} \in \mathbb{P}(U)$.

Example 3.1. Consider $S = \{1,2,3,4\}$ and $U = \{1,4\}$. The probability that U remains unchanged is

$$r_{\mathbf{1}}^{U} = r^{S}_{\{\{1,3,4\},\{2\}\}} + r^{S}_{\{\{1,2,4\},\{3\}\}} + r^{S}_{\{\{1,4\},\{2,3\}\}} + r^{S}_{\{\{1,2,3,4\}\}}.$$

The probability that U is split up into the two subsets $\{1\}$ and $\{4\}$ is

$$r^U_{\{\{1\},\{4\}\}} = r^S_{\{\{1\},\{2,3,4\}\}} + r^S_{\{\{1,2,3\},\{4\}\}} + r^S_{\{\{1,2\},\{3,4\}\}} + r^S_{\{\{1,3\},\{2,4\}\}}. \qquad \diamondsuit$$

Before we proceed, let us recall that the index set of a partition $\mathcal{A} = \{A_1, \dots, A_m\} \in \mathbb{P}(U)$, $U \subseteq S$, is denoted by $M := M(\mathcal{A}) = \{1, 2, \dots, m\}$ and that $\mathcal{A}_J = \{A_j\}_{j\in J}$ and $A_J = \cup_{j\in J} A_j$ for $J \subseteq M$. We will first describe the continuous-time partitioning process $(\breve{\Sigma}_t)_{t\geqslant 0}$ and thereafter the discrete-time counterpart $(\widehat{\Sigma}_t)_{t\in\mathbb{N}_0}$. If statements hold for both of them, we simply write Σ.

The partitioning process in continuous time. Let $(\breve{\Sigma}_t)_{t\geqslant 0}$ be the continuous-time partitioning process that starts with the initial partition $\breve{\Sigma}_0$. Suppose that the current state is $\breve{\Sigma}_t = \mathcal{A} = \{A_1, \dots, A_m\}$, and denote by Δ the waiting time to the next event. Δ is exponentially distributed with parameter m since each block corresponds to an individual and each individual is independently affected at rate 1. When the bell rings, choose a block uniformly. If A_j is picked, then $\breve{\Sigma}_{t+\Delta}$ is obtained as follows (see Figure 3.5 for an example). In the splitting step, block A_j turns into an intermediate state $\mathcal{J}$ with probability $r^{A_j}_{\mathcal{J}}$, $\mathcal{J} \in \mathbb{P}_{\leqslant 2}(A_j)$.

- (S_1) With probability $r^{A_j}_{\mathbf{1}}$, the block A_j remains unchanged. The resulting intermediate state (of this block) is $\mathcal{J} = \mathbf{1}|_{A_j}$. Note that $r^{A_j}_{\mathbf{1}}$ takes into account *all* recombination probabilities such that A_j remains intact (cf. (3.1)).
- (S_2) With probability $r^{A_j}_{\mathcal{J}}$, $\mathcal{J} \in \mathbb{P}_2(A_j)$, block A_j splits into two parts. The resulting intermediate state is $\mathcal{J} = \{A_{j_1}, A_{j_2}\}$.

Now, each block of $\mathcal{J}$ chooses out of N parents, uniformly and with replacement. Among these, there are $m-1$ parents that carry one block of $\mathcal{A}_{M\setminus j} = \mathcal{A} \setminus A_j$ each; the remaining $N-(m-1)$ parents are *empty*, that is, they do not carry ancestral material. Coalescence happens if the choosing block picks a parent that carries ancestral material; otherwise, the choosing block becomes an ancestral block of its own, which is available for coalescence from then onwards. The possible outcomes are certain coarsenings of $\mathcal{A}_{M\setminus j} \cup \mathcal{J}$.

If $\mathcal{J} = \{A_j\}$ (case (S_1)), then either

- ($C_{1,1}$) With probability $\frac{N-(m-1)}{N}$, block A_j does not coalesce with any block of $\mathcal{A}_{M\setminus j}$. As a result, $\breve{\Sigma}_{t+\Delta} = \breve{\Sigma}_t = \mathcal{A}$.
- ($C_{1,2}$) With probability $\frac{1}{N}$, block A_j coalesces with block A_k, $k \in M \setminus j$. This results in $\breve{\Sigma}_{t+\Delta} = \mathcal{A}_{M\setminus\{j,k\}} \cup A_{\{j,k\}}$.

If $\mathcal{J} = \{A_{j_1}, A_{j_2}\}$ (case (S_2)), we get the following possibilities:

- ($C_{2,1}$) With probability $\frac{(N-(m-1))(N-m)}{N^2}$, no block of $\mathcal{J}$ coalesces with a block of $\mathcal{A}_{M\setminus j}$, so $\breve{\Sigma}_{t+\Delta} = \mathcal{A}_{M\setminus j} \cup \mathcal{J}$.

($C_{2,2}$) With probability $\frac{N-(m-1)}{N^2}$, one block of $\mathcal{J}$ coalesces with block A_k, $k \in M \setminus j$, while the other block of $\mathcal{J}$ chooses an empty individual. This ends up in the state $\breve{\Sigma}_{t+\Delta} = \mathcal{A}_{M\setminus\{j,k\}} \cup \{A_{\{j_1,k\}}, A_{j_2}\}$ or $\breve{\Sigma}_{t+\Delta} = \mathcal{A}_{M\setminus\{j,k\}} \cup \{A_{\{j_2,k\}}, A_{j_1}\}$. That is, from $\breve{\Sigma}_t$ to $\breve{\Sigma}_{t+\Delta}$, either block A_{j_1} or A_{j_2} is moved from A_j to A_k.

($C_{2,3}$) With probability $\frac{N-(m-1)}{N^2}$, the blocks A_{j_1} and A_{j_2} coalesce with each other, but choose an empty individual, which gives $\breve{\Sigma}_{t+\Delta} = \mathcal{A}$.

($C_{2,4}$) With probability $\frac{1}{N^2}$, the block A_{j_1} coalesces with A_k and A_{j_2} coalesces with A_ℓ, $k, \ell \in M \setminus j$. This yields either $\breve{\Sigma}_{t+\Delta} = \mathcal{A}_{M\setminus\{j,k,\ell\}} \cup \{A_{\{j_1,k\}}, A_{\{j_2,\ell\}}\}$ if $k \neq \ell$, or $\breve{\Sigma}_{t+\Delta} = \mathcal{A}_{M\setminus\{j,k\}} \cup A_{\{j,k\}}$ if $k = \ell$.

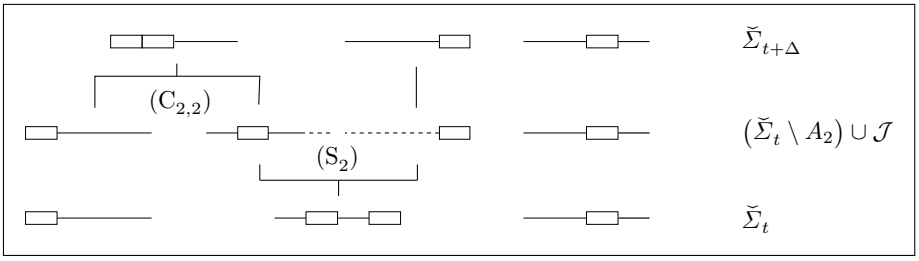


Figure 3.5. One step of the partitioning process with current state $\breve{\Sigma}_t = \{A_1, A_2, A_3\} = \{\{1\}, \{2,4\}, \{3\}\}$. In this example, A_2 is chosen and splits into $\mathcal{J} = \{\{2\}, \{4\}\}$. In the following step ($C_{2,2}$), the leading part coalesces with A_1, whereas the trailing part remains separate, so that we end up in $\breve{\Sigma}_{t+\Delta} = \{\{1,2\}, \{3\}, \{4\}\}$.

Summarising, we see that a transition from $\mathcal{A}$ to $\mathcal{B}$, via partitioning of block A_j into $\mathcal{J}$, $j \in M$, $\mathcal{J} \in \mathbb{P}_{\leqslant 2}(A_j)$, is possible whenever $\mathcal{B} \succcurlyeq \mathcal{A}_{M\setminus j} \cup \mathcal{J}$ and $\mathcal{B}|_{A_{M\setminus j}} = \mathcal{A}_{M\setminus j}$, or, equivalently, whenever

$$\mathcal{B}|_{A_j} \succcurlyeq \mathcal{J} \quad \text{and} \quad \mathcal{B}|_{A_{M\setminus j}} = \mathcal{A}_{M\setminus j}\,.$$

Each block of $\mathcal{J}$ coalesces into every block currently available with probability $\frac{1}{N}$ and remains separate with probability $\frac{N-k}{N}$ if there are currently k blocks available; in the latter case, the block considered becomes number $k+1$. We can therefore summarise the rate of the said transition as

$$\vartheta_{j,\mathcal{J};\mathcal{A},\mathcal{B}} = \begin{cases} r_{\mathcal{J}}^{A_j} \frac{1}{N^{|\mathcal{J}|}} \frac{(N-(m-1))!}{(N-|\mathcal{B}|)!}, & \text{if } \mathcal{B}|_{A_j} \succcurlyeq \mathcal{J},\ \mathcal{B}|_{A_{M\setminus j}} = \mathcal{A}_{M\setminus j}, \\ 0, & \text{otherwise.} \end{cases} \tag{3.2}$$

Note that this includes silent events where $\mathcal{B} = \mathcal{A}$. Thus, the partitioning process $(\breve{\Sigma}_t)_{t\geqslant 0}$ is a continuous-time Markov chain on $\mathbb{P}(S)$ characterised by the generator $\breve{\Theta} := (\breve{\Theta}_{\mathcal{A}\mathcal{B}})_{\mathcal{A},\mathcal{B}\in\mathbb{P}(S)}$

with nondiagonal elements

$$\begin{aligned}\breve{\Theta}_{\mathcal{A}\mathcal{B}} &= \sum_{j\in M}\sum_{\mathcal{J}\in\mathbb{P}_{\leqslant 2}(A_j)} \vartheta_{j,\mathcal{J};\mathcal{A},\mathcal{B}} \\ &= \begin{cases} r_{\mathcal{J}}^{A_j}\,\frac{1}{N^2}\,\frac{(N-(m-1))!}{(N-|\mathcal{B}|)!}, & \text{if } \mathcal{B}|_{A_j}=\mathcal{J}, \mathcal{B}|_{A_{M\setminus j}}=\mathcal{A}_{M\setminus j}, \\ & \qquad \text{for some } j\in M,\ \mathcal{J}\in\mathbb{P}_2(A_j), \\ \frac{2}{N^2}+\frac{N-1}{N^2}\big(r_{\mathbf{1}}^{A_j}+r_{\mathbf{1}}^{A_k}\big), & \text{if } \mathcal{B}=\mathcal{A}_{M\setminus\{j,k\}}\cup A_{\{j,k\}} \text{ for some } j\neq k\in M, \\ 0, & \text{for all other } \mathcal{B}\neq\mathcal{A}, \end{cases}\end{aligned} \tag{3.3}$$

and $\breve{\Theta}_{\mathcal{A}\mathcal{A}} = -\sum_{\mathcal{B}\in\mathbb{P}(S)\setminus\mathcal{A}}\breve{\Theta}_{\mathcal{A}\mathcal{B}}$. For $\mathcal{J}\in\mathbb{P}_2(A_j)$, we have distinguished between $\mathcal{B}|_{A_j}=\mathcal{J}$ and $\mathcal{B}|_{A_j}=\mathbf{1}|_{A_j}\succ\mathcal{J}$. The latter corresponds to $k=\ell$ in ($\mathrm{C}_{2,4}$) and leads to the same transition as a *pure coalescence event* in ($\mathrm{C}_{1,2}$). The total coalescence rate of j and k is

$$\frac{1}{N}\big(r_{\mathbf{1}}^{A_j}+r_{\mathbf{1}}^{A_k}\big)+\frac{1}{N^2}\Bigg(\sum_{\mathcal{J}\in\mathbb{P}_2(A_j)} r_{\mathcal{J}}^{A_j}+\sum_{\mathcal{K}\in\mathbb{P}_2(A_k)} r_{\mathcal{K}}^{A_k}\Bigg)=\frac{2}{N^2}+\frac{N-1}{N^2}\,\big(r_{\mathbf{1}}^{A_j}+r_{\mathbf{1}}^{A_k}\big)$$

since $\sum_{\mathcal{J}\in\mathbb{P}_2(U)} r_{\mathcal{J}}^{U} = 1-r_{\mathbf{1}}^{U}$, $U\subseteq S$. Note that transitions to partitions $\mathcal{B}$ with $|\mathcal{B}|>N$ are impossible, as it must be.

For three sites, using the abbreviations $r_1 := r_{\{\{1\},\{2,3\}\}}$, $r_2 := r_{\{\{1,2\},\{3\}\}}$, $r_{12} := r_{\{\{1,3\},\{2\}\}}$ and the following ordering of partitions of $\mathbb{P}(S)$

$$\{\{1,2,3\}\}\quad \{\{1\},\{2,3\}\}\quad \{\{1,2\},\{3\}\}\quad \{\{1,3\},\{2\}\}\quad \{\{1\},\{2\},\{3\}\},$$

the generator of the continuous-time partitioning process reads

$$\breve{\Theta}=\begin{pmatrix} -\frac{N-1}{N}(r_1+r_2+r_{12}) & \frac{N-1}{N}r_1 & \frac{N-1}{N}r_2 & \frac{N-1}{N}r_{12} & 0 \\ \frac{2}{N}-\frac{N-1}{N^2}r_2 & -\frac{2}{N}-\frac{(N-1)^2}{N^2}r_2 & \frac{N-1}{N^2}r_2 & \frac{N-1}{N^2}r_2 & \frac{(N-1)(N-2)}{N^2}r_2 \\ \frac{2}{N}-\frac{N-1}{N^2}r_1 & \frac{N-1}{N^2}r_1 & -\frac{2}{N}-\frac{(N-1)^2}{N^2}r_1 & \frac{N-1}{N^2}r_1 & \frac{(N-1)(N-2)}{N^2}r_1 \\ \frac{2}{N}-\frac{N-1}{N^2}(r_1+r_2) & \frac{N-1}{N^2}(r_1+r_2) & \frac{N-1}{N^2}(r_1+r_2) & -\frac{2}{N}-\frac{(N-1)^2}{N^2}(r_1+r_2) & \frac{(N-1)(N-2)}{N^2}(r_1+r_2) \\ 0 & \frac{2}{N} & \frac{2}{N} & \frac{2}{N} & -\frac{6}{N} \end{pmatrix}. \tag{3.4}$$

In the single-crossover case ($r_{12}=0$), the forth entry in the first line vanishes, and the diagonal element is updated accordingly.

Remark 3.1. The single-crossover version of $\breve{\Theta}$ coincides with the generator Θ worked out by Bobrowski and co-workers in [21] and [22] with a very different approach, forward in time. For $n\leqslant 3$, they state the generator matrices explicitly, and the identity with the single-crossover version of (3.4) is easily checked by elementwise comparison. For $n>3$, they provide an algorithm which runs through all individuals and all sites and builds up the matrix Θ as the sum $\Theta=\sum_{\mathcal{A}\in\mathbb{O}_{\leqslant 2}(S)} r_{\mathcal{A}}\,\Theta_{\mathcal{A}}$ incrementally, where $\Theta_{\mathcal{A}}$ describes the transitions obtained from recombination with respect to $\mathcal{A}$ (in the single-crossover case, $r_{\mathcal{A}}=0$ if $\mathcal{A}\notin\mathbb{O}_{\leqslant 2}(S)$). The algorithm does not distinguish between transitions induced by recombination events within ancestral (or trapped) material and recombination events that

are invisible in the genealogical perspective. Instead, a case distinction is performed that is based on whether or not one or both segments coalesce with individuals that do or do not carry ancestral material. A detailed investigation of this approach, which involves expanding the cases into 11 subcases and rearranging these according to the emerging partitions of the complete ancestral material, leads precisely to our cases $(C_{2,1})$ to $(C_{2,4})$ (here, both emerging segments contain ancestral material) and $(C_{1,1})$ and $(C_{1,2})$ (here one segment is empty). Since this approach somehow disguises or mixes the various partitions of ancestral material that may arise due to a transition, it does not lead to a closed expression for Θ. The complexity of the algorithm is of the order $n^4 B_n + B_n^2$, where B_n is the n-th Bell number (see Sect. 1.2.1). In contrast, our approach yields the matrix elements explicitly for arbitrary n and gives them a natural and plausible meaning in terms of the partitioning process in backward time. ♢

The partitioning process in discrete time. The partitioning process $(\widehat{\Sigma}_t)_{t\in\mathbb{N}_0}$ in discrete time is a Markov chain on $\mathbb{P}(S)$. In contrast to the continuous-time case, in which at every time point only a single block is chosen, in discrete time *every* block performs splitting and coalescence independently of all other blocks at every point in time.

To be precise, suppose that the current state is $\widehat{\Sigma}_t = \mathcal{A} = \{A_1, \dots, A_m\}$. In the splitting step, every block $A_j \in \mathcal{A}$, $j \in M$, independently of all others, turns into an intermediate state $A_j = \mathcal{J}_j$, $\mathcal{J}_j \in \mathbb{P}_{\leqslant 2}(S)$, with probability $r_{\mathcal{J}_j}^{A_j}$. As a result, we can obtain any partition $\mathcal{B} = \{\mathcal{J}_1, \dots, \mathcal{J}_m\}$ with probability $\prod_{j=1}^{m} r_{\mathcal{J}_j}^{A_j}$, where $\mathcal{J}_j \in \mathbb{P}_{\leqslant 2}(A_j)$ for all $j \in M$.

In the subsequent coalescence step, each block in $\mathcal{B}$ chooses out of N parents uniformly and with replacement according to the following rules: The first block $B_1 \in \mathcal{B}$ chooses an arbitrary parent; it becomes the first block carrying ancestral material. The second block $B_2 \in \mathcal{B}$ either chooses the same parent as B_1 (and merges with B_1) with probability $\frac{1}{N}$, or it chooses a different parent (and remains separate) with probability $\frac{N-1}{N}$. If there are currently k individuals carrying ancestral material, the block B_i, $k \leqslant i \leqslant |\mathcal{B}|$, chooses a particular parent that has previously been chosen by (at least) one block B_j, $j < i$, with probability $\frac{1}{N}$ (resulting in a merger of the involved blocks), or chooses a parent that has not been selected before (and remains separate) with probability $\frac{N-k}{N}$.

Let $\mathcal{B} = \{\mathcal{J}_1, \dots \mathcal{J}_m\}$, $\mathcal{J}_j \in \mathbb{P}_{\leqslant 2}(A_j)$, $j \in M$, be the intermediate state of the partitioning process obtained after performing the splitting step. The possible outcomes after the coalescence step are then certain coarsenings of $\mathcal{B}$. To be precise, a particular coarsening $\mathcal{C}$ with $|\mathcal{C}|$ blocks ($|\mathcal{C}| < |\mathcal{B}|$) can be obtained, if in the coalescence step $|\mathcal{C}| - 1$ blocks remain separate and $|\mathcal{B}| - |\mathcal{C}|$ blocks coalesce. This happens with probability

$$\frac{1}{N^{|\mathcal{B}|-|\mathcal{C}|}} \frac{(N-1)!}{N^{|\mathcal{C}|-1}(N-|\mathcal{C}|)!} = \frac{1}{N^{|\mathcal{B}|-1}} \frac{(N-1)!}{(N-|\mathcal{C}|)!}\,. \tag{3.5}$$

In the following, we will only be interested in the limiting behaviour of the discrete-time partitioning process. We therefore do not aim to give an explicit expression for the Markov matrix $\hat{\Theta} := (\hat{\Theta}_{\mathcal{A}\mathcal{B}})_{\mathcal{A},\mathcal{B}\in\mathbb{P}(S)}$ of $(\widehat{\Sigma}_t)_{t\in\mathbb{N}_0}$ for an arbitrary number of sites.

In the three-site case, $\hat{\Theta}$ reads

$$\begin{pmatrix}
1-\frac{N-1}{N}(r_1+r_2+r_{12}) & \frac{N-1}{N}r_1 & \frac{N-1}{N}r_2 & \frac{N-1}{N}r_{12} & 0 \\
\frac{1}{N}-\frac{N-1}{N^2}r_2 & -\frac{N-1}{N}-\frac{(N-1)^2}{N^2}r_2 & \frac{N-1}{N^2}r_2 & \frac{N-1}{N^2}r_2 & \frac{(N-1)(N-2)}{N^2}r_2 \\
\frac{1}{N}-\frac{N-1}{N^2}r_1 & \frac{N-1}{N^2}r_1 & -\frac{N-1}{N}-\frac{(N-1)^2}{N^2}r_1 & \frac{N-1}{N^2}r_1 & \frac{(N-1)(N-2)}{N^2}r_1 \\
\frac{1}{N}-\frac{N-1}{N^2}(r_1+r_2) & \frac{N-1}{N^2}(r_1+r_2) & \frac{N-1}{N^2}(r_1+r_2) & -\frac{N-1}{N}-\frac{(N-1)^2}{N^2}(r_1+r_2) & \frac{(N-1)(N-2)}{N^2}(r_1+r_2) \\
\frac{1}{N^2} & \frac{(N-1)}{N^2} & \frac{(N-1)}{N^2} & \frac{(N-1)}{N^2} & -\frac{(N-1)(N-2)}{N^2}
\end{pmatrix}.$$

The ordering of the partitions, as well as the abbreviations r_1, r_2, r_{12} are as in (3.4). As one can see, apart from the last row, the generator does not differ too much from its continuous-time counterpart (cf. (3.4)). From four sites onwards, there are considerable differences, namely when two or more blocks in one partition may split up within one time step. In the single-crossover case ($r_{12} = 0$), the forth entry in the first line of $\hat{\Theta}$ vanishes.

3.2.2 Limit processes

Consider the family of partitioning processes $\left(\Sigma^{(N)}\right)_{N=1,2,\ldots}$, with $\Sigma^{(N)} = \left(\Sigma_t^{(N)}\right)_{t\in\mathbb{T}}$, where we make the dependence on population size explicit through the upper index. We now examine how $\left(\Sigma_t^{(N)}\right)_{t\in\mathbb{T}}$ behaves in the two limiting cases mentioned in Section 2.2.3, namely in the deterministic limit and the diffusion limit. In Section 4.4, we will also consider the stationary distribution ($t \to \infty$) of $\left(\Sigma_t^{(N)}\right)_{t\in\mathbb{T}}$ for up to three sites. We start with the deterministic limit, first in continuous and thereafter in discrete time.

Deterministic limit

Recall that in the deterministic limit, we let $N \to \infty$ without rescaling the recombination probabilities or time. In this limit, only the pure splitting events survive, more precisely:

Proposition 3.1 (Deterministic limit, continuous time)**.** *In the deterministic limit, the sequence of continuous-time partitioning processes $\left(\breve{\Sigma}_t^{(N)}\right)_{t\geqslant 0}$ with initial states $\breve{\Sigma}_0^{(N)} = \sigma$ converges in distribution to the process $(\breve{\Sigma}'_t)_{t\geqslant 0}$ with initial state $\breve{\Sigma}'_0 = \sigma$ and generator $\breve{\Theta}'$ defined by its nondiagonal elements*

$$\breve{\Theta}'_{\mathcal{A}\mathcal{B}} = \begin{cases} r_{\mathcal{J}}^{A_j}, & \text{if } \mathcal{B} = \mathcal{A}_{M\setminus j} \cup \mathcal{J} \text{ for some } j \in M \text{ and } \mathcal{J} \in \mathbb{P}_2(A_j), \\ 0, & \text{for all other } \mathcal{B} \neq \mathcal{A}. \end{cases}$$

Hence, $(\breve{\Sigma}'_t)_{t\geqslant 0}$ is a process of progressive refinements, that is, $\breve{\Sigma}'_\tau \preccurlyeq \breve{\Sigma}'_t$ for all $\tau > t$.

Proof. Inspecting the N-dependence of the elements of $\check{\Theta}^{(N)} = \check{\Theta}$ in (3.3) gives the following order of magnitude for the nondiagonal elements:

$$\check{\Theta}^{(N)}_{\mathcal{AB}} = \begin{cases} \frac{1}{N^{m+1-|\mathcal{B}|}}\, r_{\mathcal{J}}^{A_j}(1+\mathcal{O}(\frac{1}{N})), & \text{if } \mathcal{B}|_{A_j} = \mathcal{J},\ \mathcal{B}|_{A_{M\setminus j}} = \mathcal{A}_{M\setminus j} \text{ for } j \in M, \mathcal{J} \in \mathbb{P}_2(A_j), \\ \frac{1}{N}(r_{\mathbf{1}}^{A_j} + r_{\mathbf{1}}^{A_k}) + \mathcal{O}(\frac{1}{N^2}), & \text{if } \mathcal{B} = \mathcal{A}_{M\setminus\{j,k\}} \cup A_{\{j,k\}} \text{ for some } j \neq k \in M, \\ 0, & \text{for all other } \mathcal{B} \neq \mathcal{A}. \end{cases} \tag{3.6}$$

Obviously, $\check{\Theta}^{(N)} = \check{\Theta}' + \mathcal{O}(\frac{1}{N})$, which proves convergence of the sequence of generators of $(\check{\Sigma}_t^{(N)})_{t\geqslant 0}$ to that of $(\check{\Sigma}'_t)_{t\geqslant 0}$. This entails convergence of the corresponding sequence of semigroups. With the help of Theorems 4.2.11 and 4.9.10 of [44], this guarantees convergence of $(\check{\Sigma}_t^{(N)})_{t\geqslant 0}$ to $(\check{\Sigma}'_t)_{t\geqslant 0}$ in distribution. □

The discrete-time counterpart of Proposition 3.1 reads:

Proposition 3.2 (Deterministic limit, discrete time)**.** *If $N \to \infty$, the sequence of discrete-time partitioning processes $(\widehat{\Sigma}_t^{(N)})_{t\in\mathbb{N}_0}$ with initial states $\widehat{\Sigma}_0^{(N)} = \sigma$ converges in distribution to the process $(\widehat{\Sigma}'_t)_{t\in\mathbb{N}_0}$ with initial state $\widehat{\Sigma}'_0 = \sigma$ and Markov matrix $\widehat{\Theta}'$ defined by*

$$\widehat{\Theta}'_{\mathcal{AB}} = \begin{cases} \prod\limits_{j\in M} r_{\mathcal{J}_j}^{A_j}, & \text{if } \mathcal{B} = \{\mathcal{J}_1, \ldots, \mathcal{J}_m\} \text{ and } \mathcal{J}_j \in \mathbb{P}_{\leqslant 2}(A_j) \text{ for all } j \in M, \\ 0, & \text{for all other } \mathcal{B} \neq \mathcal{A}. \end{cases}$$

Hence, $(\widehat{\Sigma}'_t)_{t\in\mathbb{N}_0}$ is again a process of progressive refinements.

Proof. As described in Section 3.2.1, $(\widehat{\Sigma}_t^{(N)})_{t\in\mathbb{N}_0}$ consists of a mixture of splitting and coalescence events performed one after the other. For a fixed $t \in \mathbb{N}_0$, suppose that the current state is $\widehat{\Sigma}_t = \mathcal{A}$, $|\mathcal{A}| = m$. In the splitting step, with probability $\prod_{j\in M} r_{\mathcal{J}_j}^{A_j}$ (independently of N), we may obtain any refined partition $\mathcal{C} = \{\mathcal{J}_1, \ldots, \mathcal{J}_m\}$, where $\mathcal{J}_j \in \mathbb{P}_{\leqslant 2}(A_j)$ for all $j \in M$. Let the subsequent coalescence step be described by the Markov chain $(\widehat{\Sigma}_t^{(N,C)})_{t\in\mathbb{N}_0}$ on $\mathbb{P}(S)$. From (3.5), we can read off that the entries of the Markov matrix $\widehat{\Theta}^{(N,C)} := (\widehat{\Theta}^{(N,C)}_{\mathcal{AB}})_{\mathcal{A},\mathcal{B}\in\mathbb{P}(S)}$ of $(\widehat{\Sigma}_t^{(N,C)})_{t\in\mathbb{N}_0}$ are given by

$$\widehat{\Theta}^{(N,C)}_{\mathcal{AB}} = \begin{cases} \frac{1}{N^{|\mathcal{A}|-|\mathcal{B}|}}\, \frac{(N-1)!}{N^{|\mathcal{B}|}(N-|\mathcal{B}|)!}, & \text{if } \mathcal{B} \succcurlyeq \mathcal{A}, \\ 0, & \text{otherwise.} \end{cases}$$

To show the claim, it therefore suffices to show that the sequence $(\widehat{\Sigma}_t^{(N,C)})_{t\in\mathbb{N}_0}$ with initial states $\widehat{\Sigma}_0^{(N,C)} = \sigma^{(C)}$ converges in distribution to the process $(\widehat{\Sigma}_t^{(C)})_{t\in\mathbb{N}_0}$ with initial state $\widehat{\Sigma}_0^{(C)} = \sigma^{(C)}$ and Markov matrix $\widehat{\Theta}^{(C)} := (\widehat{\Theta}^{(C)}_{\mathcal{AB}})_{\mathcal{A},\mathcal{B}\in\mathbb{P}(S)}$ defined by $\widehat{\Theta}^{(C)}_{\mathcal{AB}} = \delta_{\mathcal{AB}}$ for all $\mathcal{A}, \mathcal{B} \in \mathbb{P}(S)$. Since $\frac{1}{N^{|\mathcal{A}|-|\mathcal{B}|}}\, \frac{(N-1)!}{N^{|\mathcal{B}|}(N-|\mathcal{B}|)!} = \mathcal{O}(N^{|\mathcal{B}|-|\mathcal{A}|})$, we conclude that $\widehat{\Theta}^{(N,C)} = \widehat{\Theta}^{(C)} + \mathcal{O}(\frac{1}{N})$,

which proofs convergence of $(\widehat{\Sigma}_t^{(N,C)})_{t\in\mathbb{N}_0}$ to $(\widehat{\Sigma}_t^{(C)})_{t\in\mathbb{N}_0}$ in distribution with the help of Corollary 1.6 in [44]. Finally, $\widehat{\Theta}'$ is a Markov matrix since $\widehat{\Theta}'_{\mathcal{AB}} > 0$ for all $\mathcal{A},\mathcal{B}\in\mathbb{P}(S)$ and $\sum_{\mathcal{B}\in\mathbb{P}(S)}\widehat{\Theta}'_{\mathcal{AB}} = \prod_{i=1}^{|\mathcal{B}|}\sum_{\mathcal{B}_i\in\mathbb{P}_{\leqslant 2}(A_i)} r_{\mathcal{B}_i}^{A_i} = 1$, $\mathcal{A}\in\mathbb{P}(S)$. □

As we see, there are no coalescence events in the deterministic limit. Ancestral material that has been separated once will never come together again in one individual (both, $\check{\Theta}'$ and $\widehat{\Theta}'$ are triangular matrices). The absorbing state of $(\Sigma'_t)_{t\in\mathbb{T}}$ is $\mathbf{1} = \{\{1\},\{2\},\dots,\{n\}\}$. The convergence rate to the stationary state as well as the quasi-stationary behaviour of $(\widehat{\Sigma}'_t)_{t\in\mathbb{N}_0}$ conditioned on the event that $\widehat{\Sigma}'$ has not hit the limiting distribution is investigated in [93, Thm. 5.5, Corol. 5.7 & Corol. 5.8]. When starting with a single individual, i.e. with initial state $\Sigma'_0 = \{S\}$, the genealogy of this single individual may be represented by a binary tree whose nodes indicate successive splitting into smaller segments; for other initial conditions, one gets a corresponding collection (i.e. a forest) of binary trees. We investigate these trees in detail in Chapter 5.

In the special case of single-crossover recombination, splitting events will always lead to an ordered partition of sites. Recall here, that the set of ordered partitions of sites is $\mathbb{O}(S)$. If $\Sigma'_0\in\mathbb{O}(S)$ (in particular if $\Sigma'_0 = \mathbf{1}$), then all blocks in Σ'_0 are ordered in S and all blocks of Σ'_t will be ordered in S for all times.

Fact 3.1. If $\Sigma'_0\in\mathbb{O}(S)$, the single-crossover version of $(\Sigma'_t)_{t\in\mathbb{T}}$ is a Markov chain on $\mathbb{O}(S)$. □

Diffusion limit

We now turn to the diffusion limit. Recall that we speed up time in the continuous model by $\frac{N}{2}$ and assume that $Nr_{\mathcal{A}}\to\rho_{\mathcal{A}}$ for all $\mathcal{A}\in\mathbb{P}_2(S)$, where $\rho_{\mathcal{A}}$ is a constant (see Section 2.2.3). As already pointed out in the beginning of this chapter, the ARG describes the diffusion limit of the *full* ancestral recombination process. If we now restrict attention to the ancestry of n sites partitioned between m individuals, we obtain a *marginal version* of the *reduced* ARG (transitions that do not affect the genealogy are not taken into account, see p. 32), which may be formulated as follows.

Definition 3.1 (Marginalised n-locus ARG)**.** Start with the set of n sites distributed across $m\leqslant n$ individuals (or lines) according to a partition Σ''_0 with m parts. Throughout the process, every line is identified with the ancestral material it carries. If it currently carries ancestral sites $U\subseteq S$, it splits into $\mathcal{J}\in\mathbb{P}_2(U)$ at rate $\rho^U_{\mathcal{J}}$, where $\rho^U_{\mathcal{J}}$, $\mathcal{J}\in\mathbb{P}_2(U)$, is defined as in (3.1) but with r replaced by the population-scaled recombination rate ρ. Every ordered pair of lines coalesces at rate 1 and so do the ancestral sites they carry. The marginalised ARG is a partition-valued process $(\Sigma''_t)_{t\geqslant 0}$ defined by the generator Θ'' with nondiagonal elements

$$\Theta''_{\mathcal{AB}} = \begin{cases} \rho_{\mathcal{J}}^{A_j}/2, & \text{if } \mathcal{B} = \mathcal{A}_{M\setminus j}\cup\mathcal{J} \text{ for some } j\in M, \mathcal{J}\in\mathbb{P}_2(A_j),\\ 1, & \text{if } \mathcal{B}\succ\mathcal{A} \text{ and } |\mathcal{B}| = |\mathcal{A}|-1,\\ 0, & \text{for all other } \mathcal{B}\neq\mathcal{A}. \end{cases}$$

Note that this formulation of the marginal ARG differs from the version in our previous paper [40, Def. 2] by a factor of two caused by a different scaling assumption. We chose to speed up time here by $\frac{N}{2}$ (rather then by N) to point out the universality between the continuous-time and the discrete-time model (see Remark 2.2).

Proposition 3.3 (Diffusion limit, continuous time)**.** *In the diffusion limit, the sequence of continuous-time partitioning processes* $(\check{\Sigma}^{(N)}_{Nt/2})_{t\geqslant 0}$ *with initial states* $\check{\Sigma}^{(N)}_0 = \sigma$ *converges in distribution to the process* $(\Sigma''_t)_{t\geqslant 0}$ *with initial state* $\Sigma''_0 = \sigma$ *and generator* Θ'' *from Definition 3.1.*

Proof. Due to the rescaling of time, the generator of $(\check{\Sigma}^{(N)}_{Nt/2})_{t\geqslant 0}$ has nondiagonal elements $\frac{N}{2}\check{\Theta}^{(N)}_{\mathcal{A}\mathcal{B}}$. Referring back to (3.6), we obtain $\lim_{N\to\infty}\frac{N}{2}\check{\Theta}^{(N)}_{\mathcal{A}\mathcal{B}} = \Theta''_{\mathcal{A}\mathcal{B}}$ since $r^U_{\mathbf{1}} \to 1$ and $Nr^U_{\mathcal{J}} \to \rho^U_{\mathcal{J}}$ for all $\mathcal{J} \in \mathbb{P}_2(U)$. With the same argument as in the proof of Proposition 3.1, one obtains convergence in distribution as claimed. □

As already pointed out in the beginning of this chapter, only pure splitting events and pure coalescence events survive in the diffusion limit. The 'mixed transitions', which involve both splitting and coalescence (i.e. the dashed lines in Figure 3.2), vanish under the rescaling, see also [65, Fig. 5.11].

In analogy with the corresponding forward model, we also expect the convergence of the rescaled sequence of discrete-time partitioning processes $(\widehat{\Sigma}^{(N)}_{Nt})_{t\in\mathbb{N}_0}$ to $(\Sigma''_t)_{t\geqslant 0}$ in the diffusion limit under the Skorokhod topology $\mathbb{D}([0,\infty), \mathbb{P}(\{1,\dots,n\}))$, provided the initial states converge appropriately. As in the forward model, we do not aim at a rigorous proof but only want to provide the intuition behind.

We saw in Section 3.2.1, that the nonrescaled process $(\widehat{\Sigma}^{(N)}_t)_{t\in\mathbb{N}_0}$ contains splitting and coalescence events. Suppose that we speed up time by N and that $\widehat{\Sigma}_{Nt} = \mathcal{A} = \{A_1,\dots,A_m\}$ at some fixed time t. In the splitting step, when time is sped up by N, every refinement $\mathcal{B}$ of $\mathcal{A}$ with $\mathcal{B} = \{\mathcal{J}_1,\dots,\mathcal{J}_m\}$, $\mathcal{J}_j \in \mathbb{P}_{\leqslant 2}(A_j)$, $j \in M$, occurs at rate $N\prod_{j\in M} r^{A_j}_{\mathcal{J}_j}$, where $r^{A_j}_{\mathbf{1}} = 1 - \sum_{\mathcal{B}\in\mathbb{P}_2(A_j)} r^{A_j}_{\mathcal{B}}$. If now $2Nr_{\mathcal{A}} \to \rho_{\mathcal{A}}$ for all $\mathcal{A} \in \mathbb{P}_2(S)$, all transitions that involve more than one (real) splitting event are of the order $\frac{1}{N}$. They will vanish when $N \to \infty$. In the diffusion limit, therefore, *at most* one block, say A_j, $j \in M$, will split into two parts in the splitting step. The block A_j either splits into $\mathcal{J} = \{A_1, A_2\}$ with rate $Nr^{A_j}_{\mathcal{J}}$, or remains unchanged ($\mathcal{J} = \{A_j\}$) with rate $Nr^{A_j}_{\mathbf{1}}$. If A_j remains unchanged, it then coalesces with any other block A_k, $k \neq j$, with probability $\frac{1}{N}$. In the diffusion limit ($r^U_{\mathbf{1}} \to 1$), this results in a total coalescence rate of 1 per ordered pair of blocks (lineages). If $\mathcal{J} = \{A_1, A_2\}$, the coalescence steps $(\mathrm{C}_{2,1}) - (\mathrm{C}_{2,4})$ of the continuous-time partitioning process are performed, from which only $(\mathrm{C}_{2,1})$ (which is of probability $1 + \mathcal{O}(\frac{1}{N})$) survives as $N \to \infty$. When $2Nr_{\mathcal{A}} \to \rho_{\mathcal{A}}$ and $N \to \infty$, the only splitting event that is left is therefore the transition from $\mathcal{A} \to \mathcal{A}_{M\setminus j} \cup \mathcal{J}$, $j \in M$, $\mathcal{J} \in \mathbb{O}_2(A_j)$, which occurs at rate $\rho^{A_j}_{\mathcal{J}}/2$.

4 Duality: Looking forward and backward

In the past decades, enormous progress has been made in understanding the forward time population processes by studying the corresponding processes backward in time. In this chapter, we will work out a duality relation between the Moran model forward in time and the partitioning process backward in time, both with respect to the continuous-time versions and without any scaling. Duality is a general and powerful tool to infer information about one process by studying another, the dual process. The latter may, in an optimal case, have a much smaller state space than the original one, and closed expressions are often obtained more easily. Duality results are essential in interacting particle systems in physics and in population genetics.

We will briefly explain the general duality concept and give some known examples from population genetics. We then define *sampling functions*, which are closely related to the recombinators we have already met. The collection of sampling functions will serve as a duality function in Section 4.2, where the duality between the continuous-time Moran model $(\check{Z}_t)_{t\geqslant 0}$ and the continuous-time partitioning process $(\check{\Sigma}_t)_{t\geqslant 0}$ is proved. Applications of the duality relation will be considered at the end oft this chapter (Sect. 4.3 & Sect. 4.4). First, with respect to multi-locus correlation functions, known as *linkage disequilibria*, which measure the deviation of allelic frequencies from independence. Thereafter, we look at the asymptotic behaviour of $(\check{Z}_t)_{t\geqslant 0}$ and $(\check{\Sigma}_t)_{t\geqslant 0}$ as $t \to \infty$, which leads to a one-to-one correspondence between the fixation probabilities of the continuous-time Moran model and the stationary distribution of the continuous-time partitioning process for an arbitrary number of sites. Since we will throughout the entire chapter deal with the continuous-time versions only, we omit the additional indication for continuous time.

For the general principle, let $X = (X_t)_{t\geqslant 0}$ and $Y = (Y_t)_{t\geqslant 0}$ be two Markov processes with state spaces E and F. Define by $M_b(E \times F)$ the set of all bounded measurable functions on $E \times F$. The following definition of duality with respect to a function goes back to Liggett [88]; see also the review by Jansen and Kurt [71].

Definition 4.1 (Duality)**.** Two Markov processes X and Y, with laws φ and ψ, respectively, are said to be *dual* with respect to a function $H \in M_b(E \times F)$ if, for all $v \in E$, $w \in F$ and $t \geqslant 0$,

$$\mathbf{E}_\varphi \left[H(X_t, w) \mid X_0 = v \right] = \mathbf{E}_\psi \left[H(v, Y_t) \mid Y_0 = w \right]. \tag{4.1}$$

For the special case that E and F are finite, every function $H \in M_b(E \times F)$ may be

represented by a matrix with bounded entries $H(p,q)$, $p \in E$, $q \in F$. If, further, X and Y are time-homogeneous with generator matrices Λ and Θ respectively, the expectations in (4.1) may be written in terms of the corresponding semigroups, i.e.,

$$\begin{aligned} \mathbf{E}_\varphi \left[H(X_t, w) \mid X_0 = v \right] &= \sum_{p \in E} (\mathrm{e}^{t\Lambda})_{vp} H(p,w), \\ \mathbf{E}_\psi \left[H(v, Y_t) \mid Y_0 = w \right] &= \sum_{q \in F} (\mathrm{e}^{t\Theta})_{wq} H(v,q). \end{aligned} \tag{4.2}$$

Since the duality equation (4.1) is automatically satisfied at $t = 0$, it is sufficient to check the identity of the derivatives at $t = 0$. That is, Equation (4.1) holds for all times if and only if

$$\begin{aligned} \frac{\mathrm{d}}{\mathrm{d}t} \mathbf{E}_\varphi \left[H(X_t, w) \mid X_0 = v \right] \big|_{t=0} &= \sum_{p \in E} \Lambda_{vp} \, H(p,w) \\ &= \sum_{q \in F} H(v,q) \, \Theta_{wq} = \frac{\mathrm{d}}{\mathrm{d}t} \, \mathbf{E}_\psi \left[H(v, Y_t) \mid Y_0 = w \right] \big|_{t=0} \end{aligned} \tag{4.3}$$

for all $v \in E$, $w \in F$. As a short-hand of (4.3), one can write $\Lambda H = H\Theta^T$, where the superscript T denotes transpose.

Within population genetics, dual processes naturally come along with the interchange of forward and backward perspective. Pioneering work goes back to Donnelly and Kurtz [32, 33] as well as to Krone and Neuhauser [83], both in the diffusion limit. Finite population dualities are for instance investigated in [99]. The most famous example of duality in population genetics is probably the moment duality between the Wright-Fisher diffusion without recombination (cf. Eq. (2.20)) and the *block-counting process* of the Kingman coalescent (see p. 30). If $(X_t)_{t \geqslant 0}$ is the Wright-Fisher diffusion without recombination with law φ and $|\Pi_t|$ is the number of lines of the Kingman coalescent with law ψ, then, for all $p \in E'$ from (2.15) and all $n \geqslant 1$, the precise duality statement reads

$$\mathbf{E}_\varphi[(X_t)^n \,|\, X_0 = p] = \mathbf{E}_\psi[\, p^{|\Pi_t|} \,|\, |\Pi_0| = n]. \tag{4.4}$$

For a fixed type $x \in \mathbb{X}$, the left hand side is the probability to sample n individuals of type x at time t conditional on starting with some initial value p. For the right hand side, imagine that we run the coalescence process starting with n lines at time 0. If we sample a certain number of individuals at time t (which is time 0 in the Wright-Fisher diffusion), each of these lines corresponds to an individual of type x with probability p. Since these types will — in the absence of mutation — be inherited forward according to the realisation of the Kingman coalescent, the right hand side is the probability to sample n individuals of type x at time 0 in the coalescent process or at time t in the Wright-Fisher diffusion. For more details or the precise proof, see [32, 100]. Dual processes are not necessarily unique. The Kingman coalescent, for example, is also dual to the Fleming-Viot process [33, 45].

Dual processes corresponding to processes describing the dynamics of evolutionary forces such as mutation or selection are extensively studied. For recombination, however, there are only few examples available. One of these examples is the duality between the Wright-Fisher diffusion with recombination (Def. 2.3) and the block-counting process of the *reduced*

ARG. As described in Section 3.1.1, the block-counting process of the *original* ARG is the birth-death process for which the number of lines is increased by one at rate $k\rho/2$ and decreased by one at rate $\frac{k(k-1)}{2}$, where k is the number of current lines and ρ is the total population-scaled recombination rate. All transitions are irrespective of whether the recombination event affects the genealogy of the sampled individuals or not. In the reduced ARG (p. 32), all transitions that do not affect the genealogy are excluded. In contrast to the original ARG, all lineages in the reduced ARG therefore carry ancestral material. For the corresponding block-counting process in the two-site case, let A_t (B_t) be the number of lineages with ancestral material at site 1 (2) and nonancestral material at site 2 (1) at time t, and let C_t be the number of lineages with ancestral material at both sites. The reduced block-counting process of the ARG is then a jump process in $(\mathbb{Z}_+)^3 \setminus \{(0,0,0)\}$. If the current state is $(A_t, B_t, C_t) = (a, b, c)$, the following transitions are possible:

$$(a,b,c) \to \begin{cases} (a+1, b+1, c-1), & \text{at rate } c\rho/2, \\ (a-1, b-1, c+1), & \text{at rate } ab, \\ (a-1, b, c), & \text{at rate } ac + a(a-1)/2, \\ (a, b-1, c), & \text{at rate } bc + b(b-1)/2, \\ (a, b, c-1), & \text{at rate } c(c-1)/2. \end{cases}$$

Let $(X_t)_{t\geqslant 0}$ be the two-locus Wright-Fisher diffusion with recombination from Theorem 2.3 and law φ, and let $\big((A_t, B_t, C_t)\big)_{t\geqslant 0}$ be the reduced block-counting process with law ψ. The generalisation of the moment duality in (4.4) with recombination is

$$\begin{aligned} &\mathbf{E}_\varphi\Big[\big((\pi_{\{1\}}.X_t)(x)\big)^k \cdot \big((\pi_{\{2\}}.X_t)(x)\big)^l \cdot \big(X_t(x)\big)^m \,\Big|\, X_0 = p\Big] \\ &\qquad = \mathbf{E}_\psi\Big[\big((\pi_{\{1\}}.p)(x)\big)^{A_t} \cdot \big((\pi_{\{2\}}.p)(x)\big)^{B_t} \cdot \big(p(c)\big)^{C_t} \,\Big|\, (A_0, B_0, C_0) = (k,l,m)\Big], \end{aligned}$$

which, in the two-allele case, was first stated in [43, Eq. 2.14] and later worked out using different arguments in [91]. The generalisation to multiple loci, multiple alleles and parent-independent mutation is given in [59, Sect. 6]. In [73] one also finds a modified version of the reduced two-locus block-counting process that may serve as a dual process to the Wright-Fisher diffusion with recombination in the moderate diffusion limit ($t \to N^\beta t$ and $rN^\beta \to \rho$, $\beta \in (0,1)$, r crossover probability between the two loci). Further duality results including recombination are for example the duality between the Ξ-coalescent with recombination and the diffusion limit of the Moran model with skewed offspring distribution [19]. Ethier and Kurtz [45] further elaborated a remarkably universal duality relation that holds between the Fleming-Viot process with general mutation, selection and recombination operators, and its respective block-counting process.

For the *finite* duality relation we have in mind, recall that in the Moran model with recombination, an individual dies at rate 1 and is replaced by either one randomly chosen individual or by a mixture of two individuals that are sampled from the previous population *with* replacement. The procedure is systematically described via the recombination operators (recombinators), which we defined in Section 1.3.1. On the contrary, within the partitioning process (Section 3.2.1), different blocks refer to different individuals, i.e. to

sampling *without* replacement. To describe the second sampling method in a compact way, we complement our set of recombinators by closely related *sampling functions*. The collection of sampling functions will serve as a duality function in Section 4.2.

4.1 Sampling operators

Recall that $\mathcal{M}_+(\mathbb{X})$ is the space of all finite measures on the type space $\mathbb{X} = \mathbb{X}_1 \times \cdots \times \mathbb{X}_n$, and $E = \{z \in \{0, \ldots, N\}^{|\mathbb{X}|} \mid \|z\| = N\}$ is the state space of the Moran model $(Z_t)_{t \geqslant 0}$ from Definition 2.1. In the very beginning, we have already met the nonnormalised recombination operator $\bar{R}_{\mathcal{A}} \colon \mathcal{M}_+(\mathbb{X}) \to \mathcal{M}_+(\mathbb{X})$, defined for any $\mathcal{A} = \{A_1, A_2, \ldots, A_m\} \in \mathbb{P}(S)$ as

$$\bar{R}_{\mathcal{A}}(\omega) = \omega^{A_1} \otimes \cdots \otimes \omega^{A_m},$$

see Section 1.3.1. The normalised counterpart $R_{\mathcal{A}}(\omega) = 1/||\omega||^{|\mathcal{A}|}\, \bar{R}_{\mathcal{A}}(\omega)$ defines a probability measure on $\mathbb{X}$. Let us now give a probabilistic interpretation for the case that the recombinator $R_{\mathcal{A}}$ acts on a certain population described by a counting measure $z \in E$. For the moment, attach labels $1, 2, \ldots, N$ to the N individuals in the population, and let these individuals have (random) types $X_t^1, X_t^2, \ldots, X_t^N \in \mathbb{X}$ at time t. The type distribution then is $Z_t = \sum_{k=1}^N \delta_{X_t^k}$. For $U \subseteq S$ and $k \in \{1, \ldots, N\}$, let $X_{t,U}^k := \pi_U(X_t^k)$, and consider the following procedure. Let a partition $\mathcal{A} = \{A_1, \ldots, A_m\}$ of S together with a collection of labels $\ell = (\ell_1, \ldots, \ell_m) \in \{1, \ldots, N\}^m$ associated with the blocks be given, i.e., $(\mathcal{A}, \ell) := \{(A_1, \ell_1), \ldots, (A_m, \ell_m)\}$. Then, piece together a sequence by taking the sites in A_1 from individual ℓ_1, the sites in A_2 from individual ℓ_2, … the sites in A_m from individual ℓ_m. The resulting sequence is $X_{t,\mathcal{A}}^{\ell} := (X_{t,A_1}^{\ell_1}, \ldots, X_{t,A_m}^{\ell_m})$. We are now interested in the event

$$\{X_{t,\mathcal{A}} = x\} := \dot{\bigcup_{\ell \in \{1,\ldots,N\}^m}} \{X_{t,\mathcal{A}}^{\ell} = x\} \tag{4.5}$$

and the corresponding counting measure $|\{X_{t,\mathcal{A}} = x\}|$. Equation (4.5) is also taken as the definition of the random variable $X_{t,\mathcal{A}}$. Clearly, this counts how often one obtains sequence x when performing the above procedure on a population Z_t in which combinations of individuals are included. Let us emphasise that individuals are combined *with replacement*, that is, two or more blocks may come from the same individual. Therefore, the event $\{X_{t,\mathcal{A}} = x\}$ may also be understood as the union of the independent events $\{X_{t,A_j} = x_{A_j}\}$, $j \in M$, where

$$\{X_{t,A_j} = x_{A_j}\} := \dot{\bigcup_{\ell_j \in \{1,\ldots,N\}}} \{X_{t,A_j}^{\ell_j} = x_{A_j}\}. \tag{4.6}$$

Therefore,

$$|\{X_{t,\mathcal{A}} = x\}| = \prod_{j \in M} |\{X_{t,A_j} = x_{A_j}\}| = \prod_{j \in M} Z_t^{A_j}(x_{A_j}) = (\bar{R}_{\mathcal{A}}(Z_t))(x). \tag{4.7}$$

Clearly, $R_{\mathcal{A}}(Z_t)$, the corresponding normalised version, is the type distribution that results when a sequence is created by taking the letters for the blocks in $\mathcal{A}$ from individuals drawn uniformly and with replacement from the population Z_t. So

$$(R_{\mathcal{A}}(z))(x) = \mathbf{P}\big[X_{t,\mathcal{A}} = x \mid Z_t = z\big],$$

where $\mathbf{P}$ denotes probability. The left-hand side depends on time only through the value z of Z_t.

Sampling function. For $\mathcal{A} \in \mathbb{P}(S)$ and $\omega \in \mathcal{M}_+(\mathbb{X}) \setminus 0$, we now define our *sampling function*

$$\bar{H}_{\mathcal{A}}(\omega) := \sum_{\mathcal{B} \succcurlyeq \mathcal{A}} \mu(\mathcal{A}, \mathcal{B})\, \bar{R}_{\mathcal{B}}(\omega), \tag{4.8}$$

where μ is the Möbius function of the poset $(\mathbb{P}(S), \preccurlyeq)$. Recall here, that for any two partitions $\mathcal{A}, \mathcal{B} \in \mathbb{P}(S)$ with $\mathcal{A} \preccurlyeq \mathcal{B}$, the Möbius function is given by

$$\mu(\mathcal{A}, \mathcal{B}) = \prod_{i=1}^{|\mathcal{B}|} (-1)^{n_i - 1} (n_i - 1)!\,, \tag{4.9}$$

where n_i is the number of blocks of $\mathcal{A}$ within block B_i (cf. Ex. 1.3). $\bar{H}_{\mathcal{A}}(\omega)$ is not a positive measure in general, but it will turn out as positive for the important case where $\omega \in E$ with $\|\omega\| \geqslant |\mathcal{A}|$, see Lemma 4.1. We will therefore postpone the normalisation step. In any case, Möbius inversion (Theorem 1.1) immediately yields the inverse of (4.8):

Fact 4.1. For every $\mathcal{A} \in \mathbb{P}(S)$,

$$\bar{R}_{\mathcal{A}}(\omega) = \sum_{\mathcal{B} \succcurlyeq \mathcal{A}} \bar{H}_{\mathcal{B}}(\omega). \qquad \square$$

We can now give $\bar{H}_{\mathcal{A}}$ a meaning by reconsidering the procedure that led to (4.7) but, this time, individuals are not replaced. That is, for $|\mathcal{A}| \leqslant N$, we now look at the events

$$\{\widetilde{X}_{t,\mathcal{A}} = x\} := \dot{\bigcup_{\substack{\ell \in \{1,\dots,N\}^m \\ \ell_i \neq \ell_j \,\forall\, i \neq j}}} \{X^{\ell}_{t,\mathcal{A}} = x\} \tag{4.10}$$

and the corresponding counting measure $|\{\widetilde{X}_{t,\mathcal{A}} = x\}|$. Since individuals are not replaced, the events $\{\widetilde{X}_{t,A_j} = x_{A_j}\}$, $j \in M$ (defined as in (4.6) with X replaced by $\widetilde{X}$) are now *dependent*; an expression for $|\{\widetilde{X}_{t,\mathcal{A}} = x\}|$ analogous to (4.7) is therefore not immediate. Instead, we resort to an inclusion-exclusion argument and prove

Proposition 4.1. *For $\mathcal{A} \in \mathbb{P}(S)$ with $|\mathcal{A}| \leqslant N$ and $Z_t \in E$, we have*

$$|\{\widetilde{X}_{t,\mathcal{A}} = x\}| = (\bar{H}_{\mathcal{A}}(Z_t))(x).$$

Proof. Fix a given partition $\mathcal{A} \in \mathbb{P}(S)$ with $|\mathcal{A}| = m \leqslant N$. For every $\ell \in \{1, 2, \dots, N\}^m$, the pair $(\mathcal{A}, \ell)$ uniquely defines a pair $(\mathcal{B}, \tilde{\ell})$, where $\tilde{\ell} \in \{\ell \in \{1, 2, \dots, N\}^{|\mathcal{B}|} : \ell_j \neq \ell_k \,\forall\, j \neq k\}$ and $\mathcal{B} \succcurlyeq \mathcal{A}$, as follows. Join all blocks of $\mathcal{A}$ that have the same label, and attach that label to the new block. The result is $(\mathcal{B}, \tilde{\ell})$. The other way round, every $(\mathcal{B}, \tilde{\ell})$ with $\mathcal{B} \succcurlyeq \mathcal{A}$ and $\tilde{\ell} \in \{\ell \in \{1, 2, \dots, N\}^{|\mathcal{B}|} : \ell_j \neq \ell_k \,\forall\, j \neq k\}$ uniquely defines the labelling ℓ of the blocks of $\mathcal{A}$ (keep in mind that $\mathcal{A}$ is fixed): block $A_k \in \mathcal{A}$ receives the label of that block $B_j \in \mathcal{B}$ in which it is contained. We can therefore identify the set $\{(\mathcal{A}, \ell) : \ell \in \{1, 2, \dots, N\}^m\}$ with

the set $\bigcup_{\mathcal{B}\succcurlyeq\mathcal{A}}\{(\mathcal{B},\tilde{\ell}) : \tilde{\ell}\in\{\ell\in\{1,2,\dots,N\}^{|\mathcal{B}|} : \ell_j\neq\ell_k\,\forall\, j\neq k\}\}$. With (4.7) and (4.10) in mind, we can therefore *decompose* the event $\{X_{t,\mathcal{A}}=x\}=\dot{\bigcup}_{\mathcal{B}\succcurlyeq\mathcal{A}}\{\widetilde{X}_{t,\mathcal{B}}=x\}$, which entails

$$|\{X_{t,\mathcal{A}}=x\}|=\sum_{\mathcal{B}\succcurlyeq\mathcal{A}}|\{\widetilde{X}_{t,\mathcal{B}}=x\}|.$$

By (4.7), the left-hand side equals $(\overline{R}_{\mathcal{A}}(Z_t))(x)$. Due to the Möbius inversion principle (applied backward), $|\{\widetilde{X}_{t,\mathcal{B}}=x\}|$ on the right-hand side must equal $(\overline{H}_{\mathcal{B}}(Z_t))(x)$, as claimed. □

Lemma 4.1. *For $\mathcal{A}\in\mathbb{P}(S)$ with $|\mathcal{A}|=m\leqslant N$ and $z\in E$, $\overline{H}_{\mathcal{A}}(z)$ is a positive measure with*

$$\|\overline{H}_{\mathcal{A}}(z)\|=N\,(N-1)\cdots(N-m+1)>0.$$

Proof. Since, under the given assumptions, $(\overline{H}_{\mathcal{A}}(z))(x)=|\{\widetilde{X}_{t,\mathcal{A}}=x\mid Z_t=z\}|\geqslant 0$ for all x by Proposition 4.1, it is a positive measure, and its norm can be evaluated via

$$\|\overline{H}_{\mathcal{A}}(z)\|=\sum_{x\in\mathbb{X}}|\{\widetilde{X}_{t,\mathcal{A}}=x\mid Z_t=z\}|.$$

By means of (4.10), this equals the number of possibilities of how to choose m labelled individuals out of N individuals *without* replacement, where the order is respected; clearly, this is $N\,(N-1)\cdots(N-m+1)$, which is positive since $m\leqslant N$. □

Under the assumptions of Proposition 4.1, we can therefore define the normalised version of $\overline{H}_{\mathcal{A}}(z)$:

$$H_{\mathcal{A}}(z):=\frac{\overline{H}_{\mathcal{A}}(z)}{\|\overline{H}_{\mathcal{A}}(z)\|}=\frac{(N-m)!}{N!}\;\overline{H}_{\mathcal{A}}(z). \tag{4.11}$$

$H_{\mathcal{A}}(z)$ is the type distribution that results when a sequence is created by taking the letters for the blocks as encoded by $\mathcal{A}$ from individuals drawn uniformly and *without replacement* from the population z, hence

$$\big(H_{\mathcal{A}}(z)\big)(x)=\mathbf{P}\big[\widetilde{X}_{t,\mathcal{A}}=x\mid Z_t=z\big].$$

The situation described here is exactly what happens when a sample is taken in our marginal ancestral recombination process: either the initial sample (according to Σ_0, from the present population Z_t) or the ancestral one (according to Σ_t, from the initial population Z_0) — hence our name *sampling function*. In this light, Fact 4.1 expresses counting with replacement in terms of counting without replacement, provided ω is a counting measure.

It is also instructive to express the normalised sampling functions in terms of the normalised recombinators. For $z\in E$ and $|\mathcal{A}|\leqslant N$, this gives, via $\overline{R}_{\mathcal{A}}=N^{|\mathcal{A}|}R_{\mathcal{A}}(z)$,

$$H_{\mathcal{A}}(z)=\sum_{\mathcal{B}\succcurlyeq\mathcal{A}}\frac{(N-|\mathcal{A}|)!\,N^{|\mathcal{B}|}}{N!}\,\mu(\mathcal{A},\mathcal{B})\,R_{\mathcal{B}}(z). \tag{4.12}$$

Note that $\frac{(N-|\mathcal{A}|)!\,N^{|\mathcal{B}|}}{N!}=\mathcal{O}(N^{|\mathcal{B}|-|\mathcal{A}|})$. This illustrates how the inclusion of coarser partitions yields higher-order correction terms. The other way round, using $R_{\mathcal{A}}(z)=\frac{1}{N^{|\mathcal{A}|}}\overline{R}_{\mathcal{A}}$, Fact 4.1 and (4.11), one gets

$$R_{\mathcal{A}}(z)=\sum_{\mathcal{B}\succcurlyeq\mathcal{A}}\frac{N!}{N^{|\mathcal{A}|}(N-|\mathcal{B}|)!}\,H_{\mathcal{B}}(z). \tag{4.13}$$

Restriction to subsystems. Let ω be a measure in $\mathcal{M}_+(\mathbb{X})$ and $U \subseteq S$. Recall, that we write the restriction of ω to a subspace $\mathbb{X}_U = \bigtimes_{i\in U} \mathbb{X}_i$ of $\mathbb{X}$ as $\omega^U = \pi_U.\omega = \omega \circ \pi_U^{-1}$ and that $\mathcal{A}|_U$, $\mathcal{A} \in \mathbb{P}(S)$, is the partition that consists of the blocks $A_i \cap U$, see Section 1.2. For $V \subseteq U \subseteq S$, let π_V^U be the canonical projection operator acting on $\mathbb{X}_U$.

Clearly, we can also define recombinators and sampling functions for any nonempty subset $U \subseteq S$ and any partition $\mathcal{A} \in \mathbb{P}(U)$ as $\bar{R}_{\mathcal{A}}^{U}(\omega^U)$ and $\bar{H}_{\mathcal{A}}^{U}(\omega^U)$, in perfect analogy with $\bar{R}_{\mathcal{A}}^{S}(\omega)$ and $\bar{H}_{\mathcal{A}}^{S}(\omega)$ for $\mathcal{A} \in \mathbb{P}(S)$, which is $\bar{R}_{\mathcal{A}}(\omega)$ or $\bar{H}_{\mathcal{A}}(\omega)$ respectively; and likewise for $R_{\mathcal{A}}^{U}$ and $H_{\mathcal{A}}^{U}$. For clarity, we sometimes denote the subsystem by a superscript. However, as in the case of the marginal recombination probabilities, the superscript can be dispensed with since $U = \cup_{j=1}^{|\mathcal{A}|} A_j$ if $\mathcal{A} \in \mathbb{P}(U)$. The interpretation in terms of sampling, as well as Fact 4.1, carry over.

Let us collect some basic properties of recombinators:

Fact 4.2. For $\mathcal{A}, \mathcal{B} \in \mathbb{P}(S)$ and $U, V \subseteq S$ with $S = U \dot\cup V$ one has

(A) $R_{\mathcal{A}} R_{\mathcal{B}} = R_{\mathcal{A}\wedge\mathcal{B}}$.

(B) $\pi_U.R_{\mathcal{A}}^{S}(\omega) = R_{\mathcal{A}|_U}^{U}(\omega^U)$.

(C) If in addition $\mathcal{A} \preccurlyeq \{U, V\}$, then $\bar{R}_{\mathcal{A}}^{S} = \bar{R}_{\mathcal{A}|_U}^{U} \otimes \bar{R}_{\mathcal{A}|_V}^{V}$. Explicitly, this reads
$\bar{R}_{\mathcal{A}}^{S}(\omega) = \big(\bar{R}_{\mathcal{A}|_U}^{U} \otimes \bar{R}_{\mathcal{A}|_V}^{V}\big)(\omega) = \big(\bar{R}_{\mathcal{A}|_U}^{U}(\omega^U)\big) \otimes \big(\bar{R}_{\mathcal{A}|_V}^{V}(\omega^V)\big)$.

Here and in what follows, we may omit the argument when the meaning is clear. Property (A) is Proposition 2, and property (B) is Lemma 1 of [6] (they both remain true in our normalisation, see Section 1.3.1). Property (C) is an obvious generalisation of Proposition 2 of [125]. It is easily seen by using first property (A), then (1.7), then (B) and finally (1.7) once more to give

$$\begin{aligned}\bar{R}_{\mathcal{A}}^{S}(\omega) &= \bar{R}_{\{U,V\}}^{S}\big(\bar{R}_{\mathcal{A}}^{S}(\omega)\big) = \Big(\big(\pi_U.\bar{R}_{\mathcal{A}}^{S}\big) \otimes \big(\pi_V.\bar{R}_{\mathcal{A}}^{S}\big)\Big)(\omega)\\ &= \big(\bar{R}_{\mathcal{A}|_U}^{U}(\omega^U)\big) \otimes \big(\bar{R}_{\mathcal{A}|_V}^{V}(\omega^V)\big) = \big(\bar{R}_{\mathcal{A}|_U}^{U} \otimes \bar{R}_{\mathcal{A}|_V}^{V}\big)(\omega).\end{aligned}$$

Let us further investigate a connection between recombination and sampling that will be important in what follows.

Lemma 4.2. *Let $S = U \dot\cup V$ for two nonempty subsets $U, V \subseteq S$. For two partitions $\mathcal{A} \in \mathbb{P}(U)$, $\mathcal{B} \in \mathbb{P}(V)$, the recombinator and the sampling operator satisfy*

$$\bar{R}_{\mathcal{A}}^{U} \otimes \bar{H}_{\mathcal{B}}^{V} = \sum_{\substack{\mathcal{C} \succcurlyeq \mathcal{A}\cup\mathcal{B}\\ \mathcal{C}|_V = \mathcal{B}}} \bar{H}_{\mathcal{C}}^{S}.$$

Proof. Using (4.8) followed by Fact 4.2 (C) and Fact 4.1 we get

$$\bar{R}_{\mathcal{A}}^{U} \otimes \bar{H}_{\mathcal{B}}^{V} = \bar{R}_{\mathcal{A}}^{U} \otimes \Big(\sum_{\mathcal{D}\succcurlyeq\mathcal{B}} \mu(\mathcal{B},\mathcal{D})\, \bar{R}_{\mathcal{D}}^{V}\Big) = \sum_{\mathcal{D}\succcurlyeq\mathcal{B}} \mu(\mathcal{B},\mathcal{D})\, \bar{R}_{\mathcal{D}\cup\mathcal{A}}^{S} = \sum_{\mathcal{D}\succcurlyeq\mathcal{B}} \mu(\mathcal{B},\mathcal{D}) \sum_{\mathcal{E}\succcurlyeq\mathcal{D}\cup\mathcal{A}} \bar{H}_{\mathcal{E}}^{S}.$$

Changing the summation order and applying (1.5) finally leads to

$$\bar{R}_{\mathcal{A}}^{U} \otimes \bar{H}_{\mathcal{B}}^{V} = \sum_{\mathcal{C} \dot{\succcurlyeq} \mathcal{A} \cup \mathcal{B}} \bar{H}_{\mathcal{C}}^{S} \sum_{\mathcal{B} \preccurlyeq \mathcal{D} \preccurlyeq \mathcal{C}|_V} \mu(\mathcal{B}, \mathcal{D}) = \sum_{\substack{\mathcal{C} \dot{\succcurlyeq} \mathcal{A} \cup \mathcal{B} \\ \mathcal{C}|_V = \mathcal{B}}} \bar{H}_{\mathcal{C}}^{S}. \qquad \square$$

Remark 4.1. In a perfectly analogous way, one can show

$$\bar{H}_{\mathcal{A}}^{U} \otimes \bar{H}_{\mathcal{B}}^{V} = \sum_{\substack{\mathcal{C} \dot{\succcurlyeq} \mathcal{A} \cup \mathcal{B} \\ \mathcal{C}|_U = \mathcal{A}, \mathcal{C}|_V = \mathcal{B}}} \bar{H}_{\mathcal{C}}^{S}.$$

This illustrates once more that, unlike the $\bar{R}_{\mathcal{A}}$, the $\bar{H}_{\mathcal{A}}$ do *not* have a product structure; this reflects the dependence inherent to drawing without replacement. $\diamondsuit$

4.2 Duality

We will now present a duality result that justifies our construction of a marginalised sample at present via the partitioning process and sampling from the initial population (cf. Fig. 3.4).

Theorem 4.1. *The Moran model $(Z_t)_{t \geqslant 0}$ with generator Λ and law φ and the continuous-time partitioning process $(\Sigma_t)_{t \geqslant 0}$ with generator Θ and law ψ, are dual with respect to the sampling function H defined in* (4.11)*. Explicitly,*

$$\mathbf{E}_{\varphi}\big[H_{\mathcal{A}}(Z_t) \mid Z_0 = z\big] = \mathbf{E}_{\psi}\big[H_{\Sigma_t}(z) \mid \Sigma_0 = \mathcal{A}\big] \tag{4.14}$$

for all $\mathcal{A} \in \mathbb{P}(S)$ and $z \in E$.

Before we embark on the proof, let us briefly comment on the meaning of this duality result.

Remark 4.2. Equation (4.14) is the formal equivalent of the construction in Figure 3.4. To see this, recall the random variables $\widetilde{X}_{t,\mathcal{A}}$ from (4.10). With their help, the left-hand side of (4.14) may be reformulated as a probability distribution,

$$\mathbf{E}_{\varphi}\big[H_{\mathcal{A}}(Z_t) \mid Z_0 = z\big] = \mathbf{E}_{\varphi}\big[\mathbf{P}\big[\widetilde{X}_{t,\mathcal{A}} = \cdot\big] \mid Z_t, Z_0 = z\big] = \mathbf{P}_{\varphi}\big[\widetilde{X}_{t,\mathcal{A}} = \cdot \mid Z_0 = z\big],$$

since the expectation is over all realisations of Z_t. Thus, we let run the Moran model Z and look at the type distribution of Z at time t with respect to the partition $\mathcal{A}$. The right-hand side is the probability distribution considered in [22]. Likewise, the right-hand side of (4.14) is equal to

$$\mathbf{E}_{\psi}\big[H_{\Sigma_t}(z) \mid \Sigma_0 = \mathcal{A}\big] = \mathbf{E}_{\psi}\big[\mathbf{P}\big[\widetilde{X}_{0,\Sigma_t} = \cdot\big] \mid \Sigma_t, \Sigma_0 = \mathcal{A}\big] = \mathbf{P}_{\psi}\big[\widetilde{X}_{0,\Sigma_t} = \cdot \mid \Sigma_0 = \mathcal{A}\big]$$

since the expectation is over all realisations of Σ_t. The right-hand side is the distribution of types when sampling from the initial population according to the partition Σ_t. It is understood that the initial population consists of the types $X_0^1, \ldots, X_0^N$ with $\sum_{k=1}^{N} \delta_{X_0^k} = z$. Recall that time runs forward in Z_t, X_t^k and $\widetilde{X}_{t,\mathcal{A}}$, but backward in Σ_t. $\diamondsuit$

In order to avoid case distinctions in the calculations in the remainder of this section, let us agree on the following conventions concerning (partitions of) empty sets. Namely, we set $\mathcal{A}_\varnothing := \varnothing$, $\overline{H}_\varnothing(z^\varnothing) = \bar{R}_\varnothing(z^\varnothing) := z^\varnothing = \|z\| = N$ and $\mu(\varnothing, \varnothing) := 1$. We now collect some auxiliary results in the following Lemma.

Lemma 4.3. *Consider a counting measure $z \in E$, a partition $\mathcal{A} \in \mathbb{P}(S)$ with $|\mathcal{A}| = m \leqslant N$ and corresponding index set $M = \{1, \dots, m\}$, and a partition $\mathcal{B} \in \mathbb{P}(S)$. Then, the following statements hold:*

(A) $\displaystyle\sum_{x\in\mathbb{X}} \big(R_{\mathcal{B}}(z)\big)(x)\,\big[\overline{H}_{\mathcal{A}}(z+\delta_x) - \overline{H}_{\mathcal{A}}(z)\big] = \sum_{j\in M} \Big(\overline{H}_{\mathcal{A}_{M\setminus j}} \otimes R_{\mathcal{B}|_{A_j}}\Big)(z).$

(B) $\displaystyle\sum_{x\in\mathbb{X}} z(x)\,\big[\overline{H}_{\mathcal{A}}(z-\delta_x) - \overline{H}_{\mathcal{A}}(z)\big] = -m\,\overline{H}_{\mathcal{A}}(z).$

Before we prove the lemma, let us give some explanations.

Remark 4.3. Evaluating statement (A) for a given type $y \in \mathbb{X}$ yields the equivalent formulation

$$\Big(\sum_{x\in\mathbb{X}} \big(R_{\mathcal{B}}(z)\big)(x)\,\overline{H}_{\mathcal{A}}(z+\delta_x)\Big)(y) = \big(\overline{H}_{\mathcal{A}}(z)\big)(y) + \sum_{j\in M} \Big(\Big(\overline{H}_{\mathcal{A}_{M\setminus j}} \otimes R_{\mathcal{B}|_{A_j}}\Big)(z)\Big)(y).$$

Let us read the left-hand side as the expected number of y individuals when drawing the parts of $\mathcal{A}$ without replacement from the population z to which one individual with type distribution $R_{\mathcal{B}}(z)$ has been added. The statement then says that this can be achieved either by drawing *all* parts of $\mathcal{A}$ from z without replacement, *or* by drawing *all but one* of them from z without replacement and the parts of $\mathcal{B}$ induced by the remaining block independently of each other and of all other blocks. With the above conventions, the right-hand side of identity (A) furthermore simplifies to

$$\sum_{j\in M} \left(\overline{H}_{\mathcal{A}_{M\setminus j}} \otimes R_{\mathcal{B}|_{A_j}}\right)(z) = N R_{\mathcal{B}}(z) \quad \text{if } \mathcal{A} = \mathbf{1}.$$

Likewise, evaluating statement (B) for some type $y \in \mathbb{X}$ gives

$$\Big(\sum_{x\in\mathbb{X}} \frac{z(x)}{N}\overline{H}_{\mathcal{A}}(z-\delta_x)\Big)(y) = \frac{N-m}{N}\,\big(\overline{H}_{\mathcal{A}}(z)\big)(y).$$

The left-hand side is always well-defined since $z - \delta_x < 0$ can only occur with $z(x) = 0$, in which case the term vanishes. This left-hand side yields the expected number of y individuals when drawing the parts of $\mathcal{A}$ from the population z *after* removal of one randomly sampled individual. The statement then tells us that this is the same as *first* drawing the parts of $\mathcal{A}$ from *all* of z and then deciding whether none of the m affected individuals has been removed, which is the case with probability $\frac{N-m}{N}$. ♢

Proof of Lemma 4.3. We first observe that

$$\sum_{x\in\mathbb{X}} \big(R_{\mathcal{B}}(z)\big)(x)(\delta_x^U) = \pi_U.\sum_{x\in\mathbb{X}} \big(R_{\mathcal{B}}(z)\big)(x)\,\delta_x = \pi_U.\big(R_{\mathcal{B}}(z)\big) = R_{\mathcal{B}|_U}(z^U) \tag{4.15}$$

by Fact 4.2. We next evaluate $\sum_{x\in\mathbb{X}} \big(R_{\mathcal{B}}(z)\big)(x)\,\big[\bar{R}_{\mathcal{A}}(z+\delta_x) - \bar{R}_{\mathcal{A}}(z)\big]$ by expanding $\bar{R}_{\mathcal{A}}$ to

separate the action on z from that on δ_x, summing against $R_{\mathcal{B}}(z)$ (using (4.15)), applying Fact 4.1 and changing summation:

$$\begin{aligned}
&\sum_{x\in\mathbb{X}} (R_{\mathcal{B}}(z))(x)\,[\bar{R}_{\mathcal{A}}(z+\delta_x)-\bar{R}_{\mathcal{A}}(z)] \\
&\qquad = \sum_{x\in\mathbb{X}} (R_{\mathcal{B}}(z))(x) \sum_{\varnothing\neq J\subseteq M} \left(\bar{R}_{\mathcal{A}_{M\setminus J}}(z^{A_{M\setminus J}})\right)\otimes(\delta_x^{A_J}) \\
&\qquad = \sum_{\varnothing\neq J\subseteq M} \left(\bar{R}_{\mathcal{A}_{M\setminus J}}\otimes R_{\mathcal{B}|_{A_J}}\right)(z) = \sum_{\varnothing\neq J\subseteq M}\ \sum_{\mathcal{C}\succcurlyeq\mathcal{A}_{M\setminus J}} \left(\bar{H}_{\mathcal{C}}\otimes R_{\mathcal{B}|_{A_J}}\right)(z) \\
&\qquad = \sum_{\mathcal{D}\succcurlyeq\mathcal{A}}\ \sum_{j=1}^{|\mathcal{D}|} \left(\bar{H}_{\mathcal{D}\setminus D_j}\otimes R_{\mathcal{B}|_{D_j}}\right)(z),
\end{aligned}$$

where, in the last step, every A_J reappears as one D_j. Using this together with (4.8) and (1.5), we obtain

$$\begin{aligned}
&\sum_{x\in\mathbb{X}} (R_{\mathcal{B}}(z))(x)\,[\bar{H}_{\mathcal{A}}(z+\delta_x)-\bar{H}_{\mathcal{A}}(z)] = \sum_{\mathcal{C}\succcurlyeq\mathcal{A}} \mu(\mathcal{A},\mathcal{C}) \sum_{x\in\mathbb{X}} (R_{\mathcal{B}}(z))(x)\,[\bar{R}_{\mathcal{C}}(z+\delta_x)-\bar{R}_{\mathcal{C}}(z)] \\
&\quad = \sum_{\mathcal{C}\succcurlyeq\mathcal{A}} \mu(\mathcal{A},\mathcal{C}) \sum_{\mathcal{D}\succcurlyeq\mathcal{C}}\ \sum_{j=1}^{|\mathcal{D}|} \left(\bar{H}_{\mathcal{D}\setminus D_j}\otimes R_{\mathcal{B}|_{D_j}}\right)(z) \\
&\quad = \sum_{\mathcal{D}\succcurlyeq\mathcal{A}}\ \sum_{j=1}^{|\mathcal{D}|} \left(\bar{H}_{\mathcal{D}\setminus D_j}\otimes R_{\mathcal{B}|_{D_j}}\right)(z) \sum_{\mathcal{A}\preccurlyeq\mathcal{C}\preccurlyeq\mathcal{D}} \mu(\mathcal{A},\mathcal{C}) = \sum_{j\in M} \left(\bar{H}_{\mathcal{A}_{M\setminus j}}\otimes R_{\mathcal{B}|_{A_j}}\right)(z),
\end{aligned}$$

which is statement (A). In an analogous way, we can prove statement (B):

$$\begin{aligned}
&\sum_{x\in\mathbb{X}} z(x)\,[\bar{R}_{\mathcal{A}}(z-\delta_x)-\bar{R}_{\mathcal{A}}(z)] \\
&\qquad = \sum_{\varnothing\neq J\subseteq M} (-1)^{|J|} \sum_{x\in\mathbb{X}} z(x) \left(\bar{R}_{\mathcal{A}_{M\setminus J}}(z^{A_{M\setminus J}})\right)\otimes\left(\bar{R}_{\mathbf{1}}^{A_J}(\delta_x^{A_J})\right) \\
&\qquad = \sum_{\varnothing\neq J\subseteq M} (-1)^{|J|} \left(\bar{R}_{\mathcal{A}_{M\setminus J}}\otimes \bar{R}_{\mathbf{1}}^{A_J}\right)(z) = \sum_{\varnothing\neq J\subseteq M} (-1)^{|J|} \left(\bar{R}_{\mathcal{A}_{M\setminus J}\cup A_J}\right)(z) \\
&\qquad = \sum_{\varnothing\neq J\subseteq M} (-1)^{|J|} \sum_{\mathcal{B}\succcurlyeq\mathcal{A}_{M\setminus J}\cup A_J} \bar{H}_{\mathcal{B}}(z) = \sum_{\mathcal{C}\succcurlyeq\mathcal{A}} \bar{H}_{\mathcal{C}}(z) \sum_{j=1}^{|\mathcal{C}|}\ \sum_{\varnothing\neq K\subseteq C_j} (-1)^{|K|} \\
&\qquad = \sum_{\mathcal{C}\succcurlyeq\mathcal{A}} \bar{H}_{\mathcal{C}}(z) \sum_{j=1}^{|\mathcal{C}|} \left[(1-1)^{|C_j|}-1\right] = -\sum_{\mathcal{C}\succcurlyeq\mathcal{A}} |\mathcal{C}|\,\bar{H}_{\mathcal{C}}(z),
\end{aligned}$$

where, in the second-last step, every A_J reappears as a C_j. We therefore get

$$\begin{aligned}
\sum_{x\in\mathbb{X}} z(x)\,[\bar{H}_{\mathcal{A}}(z-\delta_x)-\bar{H}_{\mathcal{A}}(z)] &= \sum_{\mathcal{B}\succcurlyeq\mathcal{A}} \mu(\mathcal{A},\mathcal{B}) \sum_{x\in\mathbb{X}} z(x)\,[\bar{R}_{\mathcal{B}}(z-\delta_x)-\bar{R}_{\mathcal{B}}(z)] \\
&= -\sum_{\mathcal{B}\succcurlyeq\mathcal{A}} \mu(\mathcal{A},\mathcal{B}) \sum_{\mathcal{C}\succcurlyeq\mathcal{B}} |\mathcal{C}|\,\bar{H}_{\mathcal{C}}(z) = -\sum_{\mathcal{C}\succcurlyeq\mathcal{A}} |\mathcal{C}|\,\bar{H}_{\mathcal{C}}(z) \sum_{\mathcal{A}\preccurlyeq\mathcal{B}\preccurlyeq\mathcal{C}} \mu(\mathcal{A},\mathcal{B}) \\
&= -|\mathcal{A}|\,\bar{H}_{\mathcal{A}}(z),
\end{aligned}$$

as claimed. □

We can now proceed as follows.

Proof of Theorem 4.1. We start with the continuous-time partitioning process. We first observe that

$$\sum_{\substack{\mathcal{B}\succcurlyeq\mathcal{A}_{M\setminus j}\cup\mathcal{J}\\ \mathcal{B}|_{A_{M\setminus j}}=\mathcal{A}_{M\setminus j}}} \frac{(N-(m-1))!}{(N-|\mathcal{B}|)!} = N^{|\mathcal{J}|}, \quad j\in M,\ |\mathcal{J}|\leqslant 2. \tag{4.16}$$

This is easily verified by direct calculation; namely, for $|\mathcal{J}|=1$, the sum on the left-hand side equals $(N-(m-1))+(m-1)=N$; for $|\mathcal{J}|=2$, it evaluates to

$$(N-(m-1))\,(N-m)+(N-(m-1))\,(2m-1)+(m-1)^2=N^2.$$

We now use the formulation of the process via (3.2) and (3.3) in the first step, normalisation and (4.16) in the second, Lemma 4.2 in the third and finally another normalisation step to calculate

$$\begin{aligned}
\sum_{\mathcal{B}\in\mathbb{P}(S)} \Theta_{\mathcal{A}\mathcal{B}}\, H_{\mathcal{B}}(z) &= \sum_{j\in M}\sum_{\mathcal{J}\in\mathbb{P}_{\leqslant 2}(A_j)} \frac{r_{\mathcal{J}}}{N^{|\mathcal{J}|}} \sum_{\substack{\mathcal{B}\succcurlyeq\mathcal{A}_{M\setminus j}\cup\mathcal{J}\\ \mathcal{B}|_{A_{M\setminus j}}=\mathcal{A}_{M\setminus j}}} \frac{(N-(m-1))!}{(N-|\mathcal{B}|)!}\,(H_{\mathcal{B}}-H_{\mathcal{A}})(z)\\
&= \sum_{j\in M}\sum_{\mathcal{J}\in\mathbb{P}_{\leqslant 2}(A_j)} \frac{r_{\mathcal{J}}}{N^{|\mathcal{J}|}} \left(\left(\sum_{\substack{\mathcal{B}\succcurlyeq\mathcal{A}_{M\setminus j}\cup\mathcal{J}\\ \mathcal{B}|_{A_{M\setminus j}}=\mathcal{A}_{M\setminus j}}} \frac{(N-(m-1))!}{N!}\,\overline{H}_{\mathcal{B}}\right) - N^{|\mathcal{J}|}\,H_{\mathcal{A}}\right)(z)\\
&= \sum_{j\in M}\sum_{\mathcal{J}\in\mathbb{P}_{\leqslant 2}(A_j)} \frac{r_{\mathcal{J}}}{N^{|\mathcal{J}|}} \left(\frac{(N-(m-1))!}{N!}\big(\overline{H}_{\mathcal{A}_{M\setminus j}}\otimes \bar{R}_{\mathcal{J}}\big) - N^{|\mathcal{J}|}\,H_{\mathcal{A}}\right)(z)\\
&= \sum_{j\in M}\sum_{\mathcal{J}\in\mathbb{P}_{\leqslant 2}(A_j)} r_{\mathcal{J}}\left(H_{\mathcal{A}_{M\setminus j}}\otimes R_{\mathcal{J}} - H_{\mathcal{A}}\right)(z).
\end{aligned} \tag{4.17}$$

We now turn to the type distribution process. Here we first evaluate, with Lemma 4.3 (B):

$$\begin{aligned}
&\sum_{y\in\mathbb{X}} z(y)\,[\overline{H}_{\mathcal{A}}(z+\delta_x-\delta_y)-\overline{H}_{\mathcal{A}}(z)]\\
&\qquad = \sum_{y\in\mathbb{X}}(z+\delta_x)(y)\,\overline{H}_{\mathcal{A}}\big((z+\delta_x)-\delta_y\big) - \sum_{y\in\mathbb{X}}(z+\delta_x)(y)\,\overline{H}_{\mathcal{A}}(z)\\
&\qquad = \sum_{y\in\mathbb{X}}(z+\delta_x)(y)\,[\overline{H}_{\mathcal{A}}\big((z+\delta_x)-\delta_y\big) - \overline{H}_{\mathcal{A}}(z+\delta_x) + \overline{H}_{\mathcal{A}}(z+\delta_x) - \overline{H}_{\mathcal{A}}(z)]\\
&\qquad = (N+1-m)\,[\overline{H}_{\mathcal{A}}(z+\delta_x)-\overline{H}_{\mathcal{A}}(z)] - m\,\overline{H}_{\mathcal{A}}(z).
\end{aligned}$$

From this, we obtain via summation against $R_{\mathcal{B}}(z)$ and use of Lemma 4.3 (A) that

$$\begin{aligned}
&\sum_{x\in\mathbb{X}}\sum_{y\in\mathbb{X}} (R_{\mathcal{B}}(z))(x)\,z(y)\,[\overline{H}_{\mathcal{A}}(z+\delta_x-\delta_y)-\overline{H}_{\mathcal{A}}(z)]\\
&\qquad\qquad = (N+1-m)\sum_{j\in M}\Big(\overline{H}_{\mathcal{A}_{M\setminus j}}\otimes R_{\mathcal{B}|_{A_j}}\Big)(z) - m\,\overline{H}_{\mathcal{A}}(z).
\end{aligned} \tag{4.18}$$

We now have to examine $\sum_{z'\in E} \Lambda_{zz'} H_{\mathcal{A}}(z')$ for an arbitrary partition $\mathcal{A}$ of S. To this end, we use (2.11) and normalisation, followed by (4.18) and a change of summation involving (3.1) to calculate

$$\begin{aligned}
\sum_{z'\in E} \Lambda_{zz'}\, H_{\mathcal{A}}(z') &= \sum_{x,y\in\mathbb{X}} \lambda(z;\, y,\, x)\, [H_{\mathcal{A}}(z+\delta_x-\delta_y) - H_{\mathcal{A}}(z)] \\
&= \frac{(N-m)!}{N!} \sum_{\mathcal{B}\in\mathbb{P}_{\leqslant 2}(S)} r_{\mathcal{B}} \sum_{x,y\in\mathbb{X}} (R_{\mathcal{B}}(z))(x)\, z(y)\, [\overline{H}_{\mathcal{A}}(z+\delta_x-\delta_y) - \overline{H}_{\mathcal{A}}(z)] \\
&= \sum_{\mathcal{B}\in\mathbb{P}_{\leqslant 2}(S)} r_{\mathcal{B}} \Big[\Big(\frac{(N-(m-1))!}{N!} \sum_{j\in M} \overline{H}_{\mathcal{A}_{M\setminus j}} \otimes R_{\mathcal{B}|_{A_j}} \Big) - \frac{(N-m)!}{N!}\, m\, \overline{H}_{\mathcal{A}} \Big](z) \\
&= \sum_{\mathcal{B}\in\mathbb{P}_{\leqslant 2}(S)} r_{\mathcal{B}} \sum_{j\in M} \Big(H_{\mathcal{A}_{M\setminus j}} \otimes R_{\mathcal{B}|_{A_j}} - H_{\mathcal{A}} \Big)(z) \\
&= \sum_{j\in M} \sum_{\mathcal{J}\in\mathbb{P}_{\leqslant 2}(A_j)} \sum_{\substack{\mathcal{B}\in\mathbb{P}_{\leqslant 2}(S) \\ \mathcal{B}|_{A_j}=\mathcal{J}}} r_{\mathcal{B}} \Big(H_{\mathcal{A}_{M\setminus j}} \otimes R_{\mathcal{J}} - H_{\mathcal{A}} \Big)(z) \\
&= \sum_{j\in M} \sum_{\mathcal{J}\in\mathbb{P}_{\leqslant 2}(A_j)} r_{\mathcal{J}} \Big(H_{\mathcal{A}_{M\setminus j}} \otimes R_{\mathcal{J}} - H_{\mathcal{A}} \Big)(z),
\end{aligned}$$

which agrees with (4.17) and proves the claim. □

We can now harvest some interesting consequences. First, Equation (4.17) contains a meaningful expression for the derivative:

Corollary 4.1. *For $\mathcal{A}\in\mathbb{P}(S)$, $z\in E$ and the Moran model $(Z_t)_{t\geqslant 0}$, we have*

$$\frac{\mathrm{d}}{\mathrm{d}t}\, \mathbf{E}_{\varphi}\, [H_{\mathcal{A}}(Z_t) \mid Z_0 = z]\,\big|_{t=0} = \sum_{j\in M} \sum_{\mathcal{J}\in\mathbb{P}_{\leqslant 2}(A_j)} r_{\mathcal{J}}^{A_j} \Big(R_{\mathcal{J}}^{A_j} \otimes H_{\mathcal{A}_{M\setminus j}}^{A_{M\setminus j}} - H_{\mathcal{A}}^{S} \Big)(z). \qquad \square$$

The right-hand side has a plausible explanation. Namely, when block A_j splits into $\mathcal{J}$, the other blocks in $\mathcal{A}$ retain their current type distribution (namely, $H_{\mathcal{A}_{M\setminus j}}(z^{A_{M\setminus j}})$). Independently of this, the parts of $\mathcal{J}$ pick their types from *all* individuals (with replacement), including those individuals that already carry other parts of $\mathcal{A}_{M\setminus j}$, which is expressed by the tensor product with $R_{\mathcal{J}}(z^{A_j})$.

Next, since $H_{\mathcal{A}}(z) = \mathbf{E}_{\varphi}\, [H_{\mathcal{A}}(Z_t) \mid Z_t = z]$, Equations (4.3) and (4.14) together with (4.2) give rise to a system of differential equations for the expectations, namely:

Corollary 4.2. *For $\mathcal{A}\in\mathbb{P}(S)$ and the Moran model $(Z_t)_{t\geqslant 0}$, one has*

$$\frac{\mathrm{d}}{\mathrm{d}t}\, \mathbf{E}_{\varphi}\, [H_{\mathcal{A}}(Z_t)] = \sum_{\mathcal{B}\in\mathbb{P}(S)} \Theta_{\mathcal{A}\mathcal{B}}\, \mathbf{E}_{\varphi}\, [H_{\mathcal{B}}(Z_t)]\,. \qquad \square$$

Let us now use our results to obtain information about the continuous-time population process $(Z_t)_{t\geqslant 0}$ by studying the dual process $(\Sigma_t)_{t\geqslant 0}$. With the help of the ODE system in Corollary 4.2, we first investigate the time evolution of expected correlations of sites (linkage disequilibria). Thereafter, we study the asymptotic behaviour of (4.14) to obtain a one-to-one correspondence between the fixation probabilities of $(Z_t)_{t\geqslant 0}$ and the stationary distribution of $(\Sigma_t)_{t\geqslant 0}$.

4.3 Expected linkage disequilibria and type frequencies

In 1960, Lewontin and Kojima [87] introduced the term *linkage disequilibria* for the nonrandom association of two loci. For populations evolving under the concept of random mating (cf. p. 13), linkage disequilibria (LDE) ever since characterise the deviation of allele frequencies at various sites from independence. Such deviations are the result of an intricate interplay between resampling, recombination, selection and other evolutionary forces. Intuitively, one would expect that recombination and resampling are competing forces, where recombination decreases linkage disequilibria since it breaks up the physical connection between sites, and resampling increases disequilibria due to joint inheritance of sites. In fact, we will see that the precise relation is a bit more subtle.

For two sites $i, j \in S$ and a probability measure $p \in \mathcal{P}(\mathbb{X})$, there is an obvious way to define linkage disequilibria. Namely, for a fixed type $x \in \mathbb{X}$, define the LDE between the sites i and j as $p^{\{i,j\}}(x_{\{i,j\}}) - p^{\{i\}}(x_{\{i\}}) \cdot p^{\{j\}}(x_{\{j\}})$. From three sites onwards, many different notions of linkage disequilibria are available. We decided to use as LDEs the general correlation functions, which are widely used in statistical physics, see [39] or [96, Chap. 5.1.1]. Our choice results in an explicit formula for multi-locus LDEs for an arbitrary number of sites in terms of sums of products of marginal frequencies, see also [12, Appendix] or [55].

For any given subset $U \subseteq S$ and $\mathcal{A} \in \mathbb{P}(U)$, we first define *correlation operators* as

$$L^U_{\mathcal{A}} = \sum_{\mathcal{B} \preccurlyeq \mathcal{A}} \mu(\mathcal{B}, \mathcal{A})\, R^U_{\mathcal{B}}, \tag{4.19}$$

where μ is the Möbius function of $(\mathbb{P}(S), \preccurlyeq)$, see (4.9). The restriction to subsystems stems from the fact that one usually considers deviation from independence on small subsets of S. The $L^U_{\mathcal{A}}$ have a product structure, $L^U_{\mathcal{A}} = \prod_{j=1}^{|\mathcal{A}|} L^{A_j}_{\mathbf{1}}$, which is obvious from (4.19) together with the product structure of the recombinators (Fact 4.2 (C)) and that of the Möbius function (Proposition 1.1). The correlation operator from (4.19) has the inverse

$$R^U_{\mathcal{A}} = \sum_{\mathcal{B} \preccurlyeq \mathcal{A}} L^U_{\mathcal{B}} = \sum_{\mathcal{B} \preccurlyeq \mathcal{A}} \prod_{j=1}^{|\mathcal{B}|} L^{B_j}_{\mathbf{1}}$$

due to Möbius inversion from below (see Theorem 1.1). The latter can be reformulated as

$$L^U_{\mathcal{A}} = R^U_{\mathcal{A}} - \sum_{\mathcal{B} \prec \mathcal{A}} \prod_{j=1}^{|\mathcal{B}|} L^{B_j}_{\mathbf{1}}. \tag{4.20}$$

The case $\mathcal{A} = \mathbf{1}|_U$, $U \subseteq S$, now is of special interest. In line with population-genetics understanding, we define the *multi-locus linkage disequilibrium with respect to the sites in* U by letting $L^U_{\mathbf{1}}$ act on the marginal measure ω^U:

$$L^U_{\mathbf{1}}(\omega^U) = \sum_{\mathcal{A} \in \mathbb{P}(U)} \mu(\mathcal{A}, \mathbf{1}|_U)\, R^U_{\mathcal{A}}(\omega^U), \quad \omega \in M_+(\mathbb{X}) \setminus 0, \tag{4.21}$$

cf. (4.19). Obviously, $L^U_{\mathbf{1}}(\omega^U)$ is a measure on $\mathbb{X}_U$ but not positive in general. If the type frequencies at the respective sites in U are independent, we obtain (as it must be) that

$\mathcal{L}^U_{\mathbf{1}}(\omega^U) = 0$ since, due to the independence of the sites, $\omega^J = \otimes_{j\in J}(\omega^{\{j\}})$ for all $J \subseteq U$, so $R^U_{\mathcal{A}}(\omega^U) = \otimes_{j\in U}(\omega^{\{j\}})$ and the statement follows from (1.5). With the help of (4.20), Equation (4.21) can again be reformulated as

$$L^U_{\mathbf{1}}(\omega^U) = R^U_{\mathbf{1}}(\omega^U) - \sum_{\mathcal{B}\prec\mathbf{1}|_U} \prod_{j=1}^{|\mathcal{B}|} L^{B_j}_{\mathbf{1}}(\omega^{B_j}). \tag{4.22}$$

Example 4.1. For $S = \{1,2,3,4\}$ the LDE with respect to the sites in $U = \{1,3,4\}$ reads

$$\begin{aligned}(L^U_{\mathbf{1}}(\omega^{\{1,3,4\}}))(x) = {}& \frac{1}{\|\omega\|}\,\omega(x_1,*,x_3,x_4) - \frac{1}{\|\omega\|^2}\,\omega(x_1,*,*,*)\,\omega(*,*,x_3,x_4)\\ & - \frac{1}{\|\omega\|^2}\,\omega(x_1,*,x_3,*)\,\omega(*,*,*,x_4) - \frac{1}{\|\omega\|^2}\,\omega(x_1,*,*,x_4)\,\omega(*,*,x_3,*)\\ & + 2\,\frac{1}{\|\omega\|^3}\,\omega(x_1,*,*,*)\,\omega(*,*,x_3,*)\,\omega(*,*,*,x_4).\end{aligned}$$ ♢

Comparison to other measures of LDE. Let us compare our definition with other multi-locus linkage disequilibria that appear in the literature in the neutral case (i.e. without selection), see [23, Chap. V.4.2] for a good overview. One of the most familiar choices for LDEs is based on covariances and goes back to Slatkin, see [13, 25, 66, 120]. In our notation, for some $U \subseteq S$, a fixed type $x \in \mathbb{X}$ and a counting measure $\omega \in \mathcal{M}_+(\mathbb{X})$, the definition reads

$$(\Delta^U(\omega^U))(x) = \mathbf{E}\Big[\prod_{i\in U}\Big(\mathbb{1}_{\{\pi_i(x)=x_i\}} - (R^{\{i\}}_{\mathbf{1}|_{\{i\}}}(\omega^{\{i\}}))(x)\Big)\Big],$$

where the expectation is with respect to the entire population. Using the fact that, for all $V \subseteq U$, one has $\mathbf{E}[\mathbb{1}_{\{\pi_V(x)=x_V\}}] = (R^V_{\mathbf{1}|_V}(\omega))(x)$, a straightforward calculation yields

$$\begin{aligned}(\Delta^U(\omega^U))(x) &= \sum_{V\subseteq U}(-1)^{|U|-|V|}\,\Big(R^V_{\mathbf{1}|_V}\otimes R^{U\setminus V}_{\mathbf{0}|_{U\setminus V}}\Big)(\omega(x)) \qquad (4.23)\\ &= \sum_{\substack{V\subseteq U\\ |V|\geqslant 2}}(-1)^{|U|-|V|}\,\Big(R^V_{\mathbf{1}|_V}\otimes R^{U\setminus V}_{\mathbf{0}|_{U\setminus V}}\Big)(\omega(x)) - (-1)^{|U|}(|U|-1)(R^U_{\mathbf{0}|_U}(\omega))(x).\end{aligned}$$

Comparison of the individuals terms shows that the Δ's in (4.23) coincide with our $L_{\mathbf{1}}$'s up to three sites. From four sites onward, they obviously disagree since (4.23) is restricted to partitions in which at most one block contains more than one element. Further relations between Δ and other measures of LDE based on multivariate central moments and multivariate cumulants are given in [23, Eq. (4.28) & (4.29)]. They may be compared with our $L_{\mathbf{1}}$'s using (4.23).

A second class of well-known measures of LDEs goes back to Bennett [14, 27, 53, 64, 89]. For up to three sites, his so-called principal components (cf. p. 14) are precisely those in (4.22). From four sites onwards, the principal components, which are worked out for arbitrary many sites by Dawson and Lyubich [14, 27, 28], depend on recombination parameters and thus disagree not only with our $L_{\mathbf{1}}$'s, but also with the Δ's from (4.23). This disagreement has been overseen by various authors such as by Gorelick and Laubichler [55], whose definition of LDEs is exactly the one given in (4.22) and thus disagrees with Bennetts principal components for more than three sites.

Relation to sampling function. Even though the correlation operator highly resembles the sampling operator, the summation in $L_{\mathcal{A}}$ ranges over all refinements of $\mathcal{A}$. The sampling function, on the contrary, involves all coarsenings of $\mathcal{A}$. Expressing the correlation operator in terms of the sampling function is nevertheless possible. For $z \in E$, Equations (4.13) and (4.19) together with a change of the summation order lead to

$$\begin{aligned} L_{\mathcal{A}}^{U}(z^U) &= \sum_{\mathcal{B} \preccurlyeq \mathcal{A}} \mu(\mathcal{B}, \mathcal{A}) \sum_{\mathcal{C} \succcurlyeq \mathcal{B}} \frac{N!}{(N-|\mathcal{C}|)!\, N^{|\mathcal{B}|}}\, H_{\mathcal{C}}^{U}(z^U) \\ &= \sum_{\mathcal{C} \in \mathbb{P}(U)} \frac{N!}{(N-|\mathcal{C}|)!}\, H_{\mathcal{C}}^{U}(z^U) \sum_{\mathcal{B} \preccurlyeq \mathcal{A} \wedge \mathcal{C}} \frac{1}{N^{|\mathcal{B}|}}\, \mu(\mathcal{B}, \mathcal{A}). \end{aligned} \tag{4.24}$$

Now, let $S(n,k)$ denote the Stirling number of the second kind, i.e. the number of partitions of the set $S = \{1, \dots, n\}$ into exactly k blocks (see Section 1.2.1). Due to the product structure of the lattice of partitions, we know that refining a complete partition is equivalent to refine each block separately (see Section (1.1)). For a fixed partition $\mathcal{A} \in \mathbb{P}(U)$ with $|\mathcal{A}| = m$, the number of partitions into exactly k blocks that are finer than or equal to $\mathcal{A}$ is given by

$$S_{\preccurlyeq \mathcal{A}}(n,k) := \left(\sum_{\substack{k_1 + \cdots + k_m = k \\ 1 \leqslant k_i \leqslant |A_i|}} \prod_{i=1}^{|\mathcal{A}|} S(|A_i|, k_i) \right), \quad k \leqslant n.$$

We can thus read off from (4.24) that, for any $U \subseteq S$ with $|U| = k$, the LDE operator satisfies

$$L_{\mathbf{1}}^{U}(z^U) = \sum_{\mathcal{A} \in \mathbb{P}(U)} \frac{N!}{(N-|\mathcal{A}|)!}\, H_{\mathcal{A}}^{U}(z^U) \sum_{j=|\mathcal{A}|}^{k} \frac{1}{N^k}\; S_{\preccurlyeq \mathcal{A}}(k,j)\; (-1)^j (j-1)!\,,$$

where we used that $\mu(\mathcal{A}, \mathbf{1}|_U) = (-1)^{|\mathcal{A}|-1}(|\mathcal{A}|-1)!$ by (1.6). For two and three sites, the LDE operator and the sampling operator are related as follows

$$L_{\mathbf{1}}^{U}(z^U) = \frac{N!}{N^k (N-k)!} \sum_{\mathcal{A} \in \mathbb{P}(U)} \mu(\mathcal{A}, \mathbf{1}|_U)\, H_{\mathcal{A}}^{U}(z^U), \quad z \in E,\ |U| = k \leqslant 3. \tag{4.25}$$

Let us now consider $L_{\mathcal{A}}^{U}$ for $\mathcal{A} \in \mathbb{P}(U) \setminus \mathbf{1}|_U$. Due to its product structure, the collection of all linkage disequilibria $L_{\mathbf{1}}^{V}(\omega^V)$, $V \subseteq U$, determines all correlation functions $L_{\mathcal{A}}^{U}(\omega^U)$, $\mathcal{A} \in \mathbb{P}(U)$. This is why, for a deterministic ω, the $L_{\mathcal{A}}^{U}(\omega^U)$, $\mathcal{A} \neq \mathbf{1}|_U$, are of no particular interest of their own. This changes, however, when ω is random (like Z_t). For we typically do not know the law of Z_t completely; rather, we have access to the expectation of certain functions of Z_t. More precisely, let φ be the law of Z_t and $\mathbf{E}_\varphi$ denote the expectation with respect to φ. The product structure of the recombined measure does, in general, not carry over to the expectation, i.e.

$$\mathbf{E}_\varphi[R_{\mathcal{A}}^{U}(Z_t^U)] \neq R_{\mathcal{A}}^{U}(\mathbf{E}_\varphi[Z_t^U]), \quad \mathcal{A} \in \mathbb{P}(U),$$

see the discussion in [8]. This is indeed a subtle point that sometimes goes wrong, as in [112], Equation (12), or [22], pp. 471/472. As a consequence, in general, one also has

$\mathbf{E}_\varphi[L_{\mathcal{A}}^U(Z_t^U)] \neq \prod_{i=1}^{|\mathcal{A}|} L_{\mathbf{1}}^{A_i}(\mathbf{E}_\varphi[Z_t^{A_i}])$. In the stochastic case, therefore, it is interesting to consider the $L_{\mathcal{A}}^U$ for $\mathcal{A} \neq \mathbf{1}|_U$ as well. The expectations $\mathbf{E}_\varphi[L_{\mathcal{A}}^U(Z_t^U)]$ contain information on how the mean LDEs in one part of the sequence depend on the mean LDEs in other parts of the sequence.

4.3.1 Time evolution of linkage disequilibria

Our interest in this subsection lies in the time evolution of correlation functions of all orders. The dynamics for linkage disequilibria in the deterministic setting are well studied, see for example [14, 27]. The dynamics for finite populations, however, are challenging due to the interplay of resampling and recombination. It is usually approached forward in time [8, 22, 103, 104, 121]. In the deterministic limit restricted to single-crossover recombination, the system has an explicit solution, both for the type distribution and for correlation functions of all orders [11, 12]. This also provides a decent approximation for large but finite populations [8], but dealing appropriately with the stochasticity of finite populations remains a major challenge. Corollary 4.2 together with (4.24) now yields an efficient way to obtain multi-locus results for expected LDEs by translating the ODE system for the expected sampling functions into a system for expected linkage disequilibria. We will obtain explicit results for the expected linkage disequilibria for two and three sites. We abbreviate $\mathbf{E}_\varphi$ by $\mathbf{E}$ and assume that the initial population Z_0 is deterministic.

Two sites. For $U = S = \{1, 2\}$, there are two partitions $\{\{1, 2\}\}$ and $\{\{1\}, \{2\}\}$. We use the abbreviation $r := r_{\{\{1\},\{2\}\}}$. The ODE system of Corollary 4.2 then reads

$$\begin{aligned} \frac{\mathrm{d}}{\mathrm{d}t}\,\mathbf{E}\big[H_{\{\{1,2\}\}}(Z_t)\big] &= r\,\frac{N-1}{N}\,\mathbf{E}\big[(H_{\{\{1\},\{2\}\}} - H_{\{\{1,2\}\}})(Z_t)\big], \\ \frac{\mathrm{d}}{\mathrm{d}t}\,\mathbf{E}\big[H_{\{\{1\},\{2\}\}}(Z_t)\big] &= \frac{2}{N}\,\mathbf{E}\big[(H_{\{\{1,2\}\}} - H_{\{\{1\},\{2\}\}})(Z_t)\big], \end{aligned} \tag{4.26}$$

where we have dropped the upper index, which is always U. This yields

$$\frac{\mathrm{d}}{\mathrm{d}t}\,\mathbf{E}\big[(H_{\{\{1,2\}\}} - H_{\{\{1\},\{2\}\}})(Z_t)\big] = -\left(\frac{2}{N} + r\,\frac{N-1}{N}\right)\mathbf{E}\big[(H_{\{\{1,2\}\}} - H_{\{\{1\},\{2\}\}})(Z_t)\big]. \tag{4.27}$$

Since $L_{\{\{1,2\}\}} = \frac{N-1}{N}\,(H_{\{\{1,2\}\}} - H_{\{\{1\},\{2\}\}})$ by (4.25), it follows that the expected two-point LDE decays at rate $\frac{2}{N} + \frac{r(N-1)}{N}$. In the case of two alleles per site, an equivalent formula has appeared in [21, Ex. 1]. The corresponding result in the diffusion limit goes back to Ohta and Kimura [103, 104], see also [37, Chap. 8.2]. As noted there, it may seem surprising that the correlations also decay via resampling (even if $r = 0$); but recall that our Moran model with recombination is an absorbing Markov chain where a single type goes to fixation in the long run, that is, Z_t will ultimately end up in a point measure.

Additionally, we can now easily obtain the expected type distribution from (4.26) and (4.27):

$$\begin{aligned} \mathbf{E}\big[H_{\{\{1,2\}\}}(Z_t)\big] &= \mathbf{E}\big[H_{\{\{1,2\}\}}(Z_0)\big] - r\,\frac{N-1}{N}\int_0^t \mathbf{E}\big[(H_{\{\{1,2\}\}} - H_{\{\{1\},\{2\}\}})(Z_\tau)\big]\mathrm{d}\tau \\ &= \frac{Z_0}{N} - \frac{r(N-1)}{r(N-1)+2}\Big(1 - \exp\Big(-\frac{r(N-1)+2}{N}\,t\Big)\Big)\mathbf{E}\big[(H_{\{\{1,2\}\}} - H_{\{\{1\},\{2\}\}})(Z_0)\big]. \end{aligned}$$

Three sites. Consider $U = S = \{1, 2, 3\}$, and let us recall the generator of the continuous-time partitioning process from (3.4), which is

$$\Theta = \begin{pmatrix} -\frac{N-1}{N}(r_1+r_2+r_{12}) & \frac{N-1}{N}r_1 & \frac{N-1}{N}r_2 & \frac{N-1}{N}r_{12} & 0 \\ \frac{2}{N}-\frac{N-1}{N^2}r_2 & -\frac{2}{N}-\frac{(N-1)^2}{N^2}r_2 & \frac{N-1}{N^2}r_2 & \frac{N-1}{N^2}r_2 & \frac{(N-1)(N-2)}{N^2}r_2 \\ \frac{2}{N}-\frac{N-1}{N^2}r_1 & \frac{N-1}{N^2}r_1 & -\frac{2}{N}-\frac{(N-1)^2}{N^2}r_1 & \frac{N-1}{N^2}r_1 & \frac{(N-1)(N-2)}{N^2}r_1 \\ \frac{2}{N}-\frac{N-1}{N^2}(r_1+r_2) & \frac{N-1}{N^2}(r_1+r_2) & \frac{N-1}{N^2}(r_1+r_2) & -\frac{2}{N}-\frac{(N-1)^2}{N^2}(r_1+r_2) & \frac{(N-1)(N-2)}{N^2}(r_1+r_2) \\ 0 & \frac{2}{N} & \frac{2}{N} & \frac{2}{N} & -\frac{6}{N} \end{pmatrix},$$

where $r_1 := r_{\{\{1\},\{2,3\}\}}$, $r_2 := r_{\{\{1,2\},\{3\}\}}$ and $r_{12} := r_{\{\{1,3\},\{2\}\}}$. The partitions of $\mathbb{P}(U)$ are ordered as

$$\{\{1,2,3\}\} \quad \{\{1\},\{2,3\}\} \quad \{\{1,2\},\{3\}\} \quad \{\{1,3\},\{2\}\} \quad \{\{1\},\{2\},\{3\}\}.$$

If we set $H(Z_t) := (H_{\mathcal{A}}^{U}(Z_t))_{\mathcal{A}\in\mathbb{P}(U)}$, we have

$$\frac{\mathrm{d}}{\mathrm{d}t}\,\mathbf{E}[H(Z_t)] = \Theta\,\mathbf{E}[H(Z_t)] \tag{4.28}$$

by Corollary 4.2. Using the translation of the correlation functions in terms of the sampling operator in (4.25) and setting $L(Z_t) := (L_{\mathcal{A}}^{U}(Z_t))_{\mathcal{A}\in\mathbb{P}(U)}$, leads to $L(Z_t) = TH(Z_t)$, where the transformation matrix is of the form

$$T = \tfrac{(N-1)(N-2)}{N^2}\begin{pmatrix} 1 & -1 & -1 & -1 & 2 \\ \frac{1}{N-2} & 1+\frac{1}{N-2} & \frac{-1}{N-2} & \frac{-1}{N-2} & -1 \\ \frac{1}{N-2} & -\frac{1}{N-2} & 1+\frac{1}{N-2} & -\frac{1}{N-2} & -1 \\ \frac{1}{N-2} & -\frac{1}{N-2} & -\frac{1}{N-2} & 1+\frac{1}{N-2} & -1 \\ \frac{1}{(N-1)(N-2)} & \frac{1}{N-2} & \frac{1}{N-2} & \frac{1}{N-2} & 1 \end{pmatrix},$$

as easily verified by combining (4.24) and (4.25). Together with (4.28), this gives us the following ODE system for the expected correlation functions

$$\frac{\mathrm{d}}{\mathrm{d}t}\,\mathbf{E}[L(Z_t)] = T\Theta T^{-1}\,\mathbf{E}[L(Z_t)], \tag{4.29}$$

where $T\Theta T^{-1}$ is given by

$$\begin{pmatrix} -\frac{6}{N}-\frac{(N-1)(N-2)}{N^2}(r_1+r_2+r_{12}) & -\frac{(N-1)(N-2)}{N^2}r_{12} & -\frac{(N-1)(N-2)}{N^2}r_{12} & 0 & 0 \\ \frac{2}{N}-\frac{(N-1)}{N^2}(r_1+r_2+r_{12}) & -\frac{2}{N}-\frac{N-1}{N}r_2-\frac{(N-1)}{N^2}r_{12} & -\frac{(N-1)}{N^2}r_{12} & 0 & 0 \\ \frac{2}{N}-\frac{(N-1)}{N^2}(r_1+r_2+r_{12}) & -\frac{(N-1)}{N^2}r_{12} & -\frac{2}{N}-\frac{N-1}{N}r_1-\frac{(N-1)}{N^2}r_{12} & 0 & 0 \\ \frac{2}{N}-\frac{(N-1)}{N^2}(r_1+r_2+r_{12}) & -\frac{(N-1)}{N^2}r_{12} & -\frac{(N-1)}{N^2}r_{12} & -\frac{2}{N}-\frac{N-1}{N}(r_1+r_2) & 0 \\ -\frac{1}{N^2}(r_1+r_2+r_{12}) & \frac{2}{N}-\frac{1}{N}r_2-\frac{1}{N^2}r_{12} & \frac{2}{N}-\frac{1}{N}r_1-\frac{1}{N^2}r_{12} & \frac{2}{N}-\frac{1}{N}(r_1+r_2) & 0 \end{pmatrix}.$$

Solving (4.29) is inconvenient since $T\Theta T^{-1}$ is far away from being sparse. Nonetheless, we can see that $T\Theta T^{-1}$ has a nice subtriangular structure in the *single-crossover* case, namely, when $r_{12} = 0$. Let us therefore neglect double-crossover events in this example. From the subtriangular structure of $T\Theta T^{-1}$ in the single-crossover case, we can then already read off that the *expected three-point LDE* (cf. (4.25)) decays exponentially according to

$$\frac{\mathrm{d}}{\mathrm{d}t}\,\mathbf{E}[L_{\{\{1,2,3\}\}}(Z_t)] = -\left(\frac{6N + (N-1)(N-2)(r_1+r_2)}{N^2}\right)\mathbf{E}[L_{\{\{1,2,3\}\}}(Z_t)].$$

As in the case of two sites, the decay rate contains contributions from resampling as well as from recombination. To extract more information, we recast $T\Theta T^{-1}$ into the diagonal form $V^{-1}T\Theta T^{-1}V = D$, where the entries of the diagonal matrix D are those on the diagonal of $T\Theta T^{-1}$, i.e., its eigenvalues. Consequently, Equation (4.29) can be rewritten as

$$\frac{\mathrm{d}}{\mathrm{d}t}\, V^{-1}\, \mathbf{E}[L(Z_t)] = DV^{-1}\, \mathbf{E}[L(Z_t)],$$

which, together with $T\Theta T^{-1}$, allows to directly read off the decay rates of linear combinations of $\mathbf{E}[L_{\mathcal{A}}(Z_t)]$'s. With the help of the subtriangular structure of $T\Theta T^{-1}$ in the single-crossover case, the matrix V^{-1} can be calculated explicitly for arbitrary N and arbitrary strength of recombination. It is again subtriangular, but somewhat unwieldy.

To streamline the results, we now turn to the diffusion limit $(\Sigma''_t)_{t\geqslant 0}$ from Definition 3.1 with generator

$$\Theta'' = \begin{pmatrix} -\frac{1}{2}(\rho_1+\rho_2) & \frac{1}{2}\rho_1 & \frac{1}{2}\rho_2 & 0 & 0 \\ 1 & -1-\frac{1}{2}\rho_2 & 0 & 0 & \frac{1}{2}\rho_2 \\ 1 & 0 & -1-\frac{1}{2}\rho_1 & 0 & \frac{1}{2}\rho_1 \\ 1 & 0 & 0 & 1-\frac{1}{2}(\rho_1+\rho_2) & \frac{1}{2}(\rho_1+\rho_2) \\ 0 & 1 & 1 & 1 & -3 \end{pmatrix}, \tag{4.30}$$

where $\rho_i = \lim_{N\to\infty} Nr_i$, $i = 1, 2$. Recall that $\rho_{12} = 0$ due to the single-crossover assumption. In the diffusion limit, T and T^{-1} converge to matrices T'' and $(T'')^{-1}$, respectively, with elements $T''_{\mathcal{AB}} = \mu(\mathcal{B}, \mathcal{A})\, \delta_{\mathcal{B}\preccurlyeq\mathcal{A}}$ and $(T'')^{-1}_{\mathcal{AB}} = \delta_{\mathcal{B}\preccurlyeq\mathcal{A}}$, $\mathcal{A}, \mathcal{B} \in \mathbb{P}(U)$. The former is easily checked, the latter is due to Möbius inversion from below (Theorem 1.1). This yields

$$T''\Theta''(T'')^{-1} = \begin{pmatrix} -(3+\frac{\rho_1}{2}+\frac{\rho_2}{2}) & 0 & 0 & 0 & 0 \\ 1 & -(1+\frac{\rho_2}{2}) & 0 & 0 & 0 \\ 1 & 0 & -(1+\frac{\rho_1}{2}) & 0 & 0 \\ 1 & 0 & 0 & -(1+\frac{\rho_1}{2}+\frac{\rho_2}{2}) & 0 \\ 0 & 1 & 1 & 1 & 0 \end{pmatrix}.$$

The rescaling of time by $\frac{N}{2}$ has already been absorbed in Θ''. In place of V^{-1}, we now get

$$(V'')^{-1} = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ -\frac{4}{(2+\rho_2)(4+\rho_1)} & -\frac{2}{2+\rho_2} & 0 & 0 & 0 \\ -\frac{4}{(2+\rho_1)(4+\rho_2)} & 0 & -\frac{2}{2+\rho_1} & 0 & 0 \\ -\frac{1}{(2+\rho_1+\rho_2)} & 0 & 0 & -\frac{2}{2+\rho_1+\rho_2} & 0 \\ \frac{4(\rho_1\rho_2+(2+\rho_1+\rho_2)(6+\rho_1+\rho_2))}{(2+\rho_1)(2+\rho_2)(2+\rho_1+\rho_2)(6+\rho_1+\rho_2)} & \frac{2}{2+\rho_2} & \frac{2}{2+\rho_1} & \frac{2}{2+\rho_1+\rho_2} & 1 \end{pmatrix},$$

which diagonalises $T\Theta''T^{-1}$. Note that in [40], where a different time scaling (N instead of $\frac{N}{2}$) is used, there are some minus symbols missing for the entries of $(V'')^{-1}$ that are corrected here. In contrast to $|U| = 2$, we see that the linear combinations of $\mathbf{E}[L_{\mathcal{A}}(Z_t)]$'s that decay exponentially have coefficients depending on the recombination rates (with exception of $\mathbf{E}[L_{\{\{1,2,3\}\}}(Z_t)]$). As an example, $(4+\rho_1)\, \mathbf{E}[L_{\{\{1\},\{2,3\}\}}(Z_t)] + 2\, \mathbf{E}[L_{\{\{1,2,3\}\}}(Z_t)]$ is one such combination and decays at rate $\frac{1}{4}(2+\rho_2)$. Solving the complete system is still possible due to the triangular structure. However, it is somewhat tedious since it involves the linear combination given in the last line of $(V'')^{-1}$. Further progress may be possible if alternative scalings are employed, such as the *loose linkage approach* suggested by Jenkins et al. [73] (see Section 2.2.3).

4.4 Fixation probabilities

Our Moran model is an absorbing Markov chain, where a single type will go to fixation in the long run, that is, the entire population will ultimately consist of a single type. In the one-locus case, this type will be one of the types initially present, and it is well known that the fixation probability for a given type equals its initial frequency. If there is recombination, the type that ultimately wins can also be a newly-composed type, but little is known about the fixation probabilities of the many possible types.

The duality relation between the Moran model with recombination and the partitioning process backward in time now facilitates the investigation of the fixation probabilities of $(Z_t)_{t\geqslant 0}$ by studying the limiting behaviour of $(\Sigma_t)_{t\geqslant 0}$ as $t \to \infty$. Since the state spaces E and $\mathbb{P}(S)$ of $(Z_t)_{t\geqslant 0}$ and $(\Sigma_t)_{t\geqslant 0}$ are finite and since $H_{\mathbf{1}}(Z_t) = \frac{Z_t}{N}$, evaluating the duality equation (4.14) for $\Sigma_0 = \mathbf{1}$ yields

$$\mathbf{E}_\varphi\left[\frac{Z_t}{N}\,\middle|\,\frac{Z_0}{N}=\frac{z}{N}\right] = \sum_{z'\in E}\frac{z'}{N}\,\mathbf{P}_\varphi\left[\frac{Z_t}{N}=\frac{z'}{N}\,\middle|\,\frac{Z_0}{N}=\frac{z}{N}\right] \tag{4.31}$$

for the left hand side and

$$\mathbf{E}_\psi\bigl[H_{\Sigma_t}(z)\,|\,\Sigma_0=\mathbf{1}\bigr] = \sum_{\mathcal{A}\in\mathbb{P}(S)}\mathbf{P}_\psi\bigl[\Sigma_t=\mathcal{A}\,|\,\Sigma_0=\mathbf{1}\bigr]\,H_{\mathcal{A}}(Z_0) \tag{4.32}$$

for the right hand side. Passing to the limit $t \to \infty$ in (4.31) and (4.32) for a fixed type $x \in \mathbb{X}$ then leads to

$$\mathbf{P}_\varphi[Z_t(x) \text{ absorbs in } x] = \sum_{\mathcal{A}\in\mathbb{P}(S)}\nu_{\mathcal{A}}\,H_{\mathcal{A}}(Z_0)(x), \tag{4.33}$$

where $\nu := (\nu_{\mathcal{A}})_{\mathcal{A}\in\mathbb{P}(S)}$ is the stationary distribution of the partitioning process $(\Sigma_t)_{t\geqslant 0}$. The convergence of the right hand side of (4.31) to the left hand side of (4.33) is true since $(Z_t)_{t\geqslant 0}$ is an absorbing Markov chain, which means that Z_t will ultimately end up in a point measure ($Z_t(x) = N$ and $Z_t(y) = 0$ for all $y \neq x$, $x, y \in \mathbb{X}$), so $\mathbf{P}_\varphi[Z_t \text{ absorbs in } x] = \lim_{t\to\infty}\mathbf{E}_\varphi\bigl[\frac{Z_t}{N}\bigr](x)$. For the convergence of the right hand side of (4.32) to the right hand side of (4.33), recall that in the partitioning process any partition can be obtained from any other partition by a finite number of splitting and coalescence events. In other words, the generator of $(\Sigma_t)_{t\geqslant 0}$ is irreducible. This guarantees the existence of a unique stationary distribution ν. The corresponding result in the diffusion limit can be found in [59, Sect. 6].

4.4.1 Stationary distribution of the partitioning process

Finding the stationary distribution $\nu = (\nu_{\mathcal{A}})_{\mathcal{A}\in\mathbb{P}(S)}$ of $(\Sigma_t)_{t\geqslant 0}$ with generator Θ, namely the solution to the equation $\nu\,\Theta = 0$ for an arbitrary number of sites, is a difficult task due to the rapid growth of the coupled system of linear equations with n. To see this, recall that Θ is a $|\mathbb{P}(S)| \times |\mathbb{P}(S)|$ matrix, where $|\mathbb{P}(S)|$ is the number of partitions of S and is given by the n-th Bell number B_n, see Section 1.2.1. For three sites, there are $B_3 = 5$ partitions, for six sites already $B_6 = 203$. There is yet no closed solution for an arbitrary number of sites

available, but there are some attempts in the literature. Bobrowski et al. [22], for instance, conclude from simulation studies that the time until the stationary state is reached is of the same order as population size. Griffiths et al. [59, Sec. 6.1], on the other hand, constitute a recursive method to compute the stationary distribution of Θ for n sites, given the solution for $n-1$ sites. In this paragraph, we only look at examples with two or three sites. In the three-site case, this allows to study the influence of trapped material (non-ancestral material enclosed between ancestral material) and double-crossover events. The frequency of trapped material is indeed an interesting object in the partitioning process. Some simplifications of the ancestral recombination process (in the diffusion limit), such as the sequential Markov coalescent (SMC) by McVean and Cardin [95], only serve as good approximations of the original ancestral recombination process (in the diffusion limit) if the amount of trapped material is negligible.

Two sites. For $S = \{1,2\}$, let $r := r_{\{\{1\},\{2\}\}}$ be the probability of a crossover between site 1 and 2, and let $\rho := \rho_{\{\{1\},\{2\}\}}$ be the respective population-scaled rate. The transition matrix Θ of the continuous-time partitioning process $(\Sigma_t)_{t\geqslant 0}$ in the two-site case $(\{\{1,2\}\},\ \{\{1\},\{2\}\})$ is given by

$$\Theta = \begin{pmatrix} -\frac{N-1}{N}\, r & \frac{N-1}{N}\, r \\ \frac{2}{N} & -\frac{2}{N} \end{pmatrix}.$$

The stationary distribution is obtained in a straightforward way and is given by

$$\nu = \left(\frac{2}{2+(N-1)r}\,,\ \frac{(N-1)r}{2+(N-1)r} \right).$$

Using (4.33) and $H_{\mathbf{1}}(Z_0) = \frac{Z_0}{N}$, we can read off the fixation probabilities of the two-locus Moran model $(Z_t)_{t\geqslant 0}$:

$$\mathbf{P}[Z_t \text{ absorbs in } x] = \frac{2}{2+r(N-1)}\, \frac{Z_0(x)}{N} + \frac{r(N-1)}{2+r(N-1)}\, (H_{\{\{1\}\{2\}\}}(Z_0))(x).$$

With probability $\frac{2}{2+r(N-1)}$ (the relative intensity of resampling), the type that wins is drawn from the initial distribution. With probability $\frac{r(N-1)}{2+r(N-1)}$ (the relative intensity of recombination), it is drawn from the distribution that results when the leading and the trailing segments are sampled from the initial population without replacement.

Three sites. Let us abbreviate $r_1 := r_{\{\{1\},\{2,3\}\}}$, $r_2 := r_{\{\{1,2\},\{3\}\}}$ and $r_{12} := r_{\{\{1,3\},\{2\}\}}$. Explicit expressions for $\nu = (\nu_{\mathcal{A}})_{\mathcal{A}\in\mathbb{P}(\{1,2,3\})}$ in the three-site case can be obtained using Mathematica, but they are long and not informative. Even in the diffusion limit, for which the generator Θ'' from Definition 3.1 is of a simple form, formulas are rather involved. Nonetheless, if we set $\rho := \rho_1 = \rho_2$ and $\rho_{12} = 0$ (the ρ's are population scaled recombination rates corresponding to r_1, r_2 and r_{12}) and order the partitions according to

$$\{\{1,2,3\}\}\quad \{\{1\},\{2,3\}\}\quad \{\{1,2\},\{3\}\}\quad \{\{1,3\},\{2\}\}\quad \{\{1\},\{2\},\{3\}\},$$

we obtain a reasonably short expression

$$\nu'' = \frac{1}{(1+\rho)(2+\rho)(3+\rho)} \Big(6+5\rho\,,\ \rho(3+2\rho)\,,\ \rho(3+2\rho)\,,\ \rho^2\,,\ \rho^2(1+\rho)\Big),$$

from which we read off that the full partition $\{\{1,2,3\}\}$ predominates for small recombination rates, whereas the influence of $\{\{1\},\{2\},\{3\}\}$ increases quickly when ρ increases. There is also a parameter space ($\rho \sim 3$), where all partitions (except of $\{\{1,3\},\{2\}\}$) are equally frequent.

To get an impression for the behaviour of the stationary distribution of the partitioning process in the finite multi-crossover case, we studied different scenarios with respect to different population sizes and different recombination strength and represented them in Figure 4.1, Figure 4.2 and Figure 4.3. As it was to be expected, Figure 4.1 and Figure 4.2 show that for high recombination probabilities and large populations, the partitioning process is dominated by the absorbing state $\{\{1\},\{2\},\{3\}\}$ of the deterministic limit; blocks split up quickly, and only a minority of them is involved in coalescence again. Eq. (4.33) then tells us, that the fixation probabilities of $(Z_t)_{t\geqslant 0}$ can be obtained as linear combinations of this curves, weighted with the corresponding value of the sampling function for Z_0.

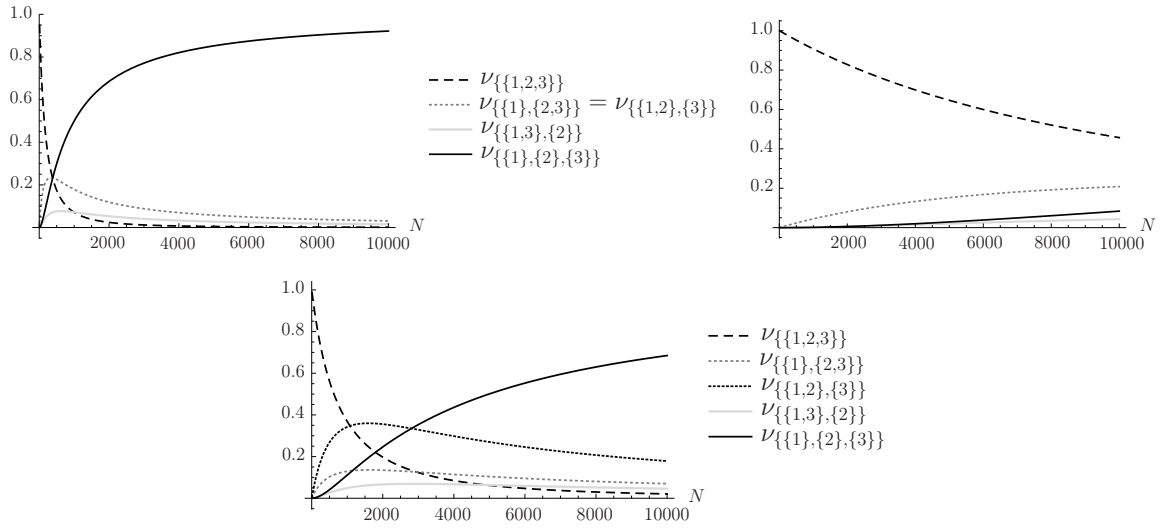


Figure 4.1. Stationary distribution $\nu = (\nu_{\mathcal{A}})_{\mathcal{A}\in\mathbb{P}(\{1,2,3\})}$ of $(\Sigma_t)_{t\geqslant 0}$ with respect to growing population size, comparing different strength of recombination. Left upper panel: high recombination probabilities ($r_1 = r_2 = 0.006$, $r_{12} = 0.00003$). Right upper panel: low recombination probabilities ($r_1 = r_2 = 0.0001$, $r_{12} = 0.000001$). Lower panel: recombination probabilities varying between the sites ($r_1 = 0.0008$, $r_2 = 0.002$, $r_{12} = 0.00003$).

One can also see nicely that the state in which trapped material is present, namely the state $\{\{1,3\},\{2\}\}$, has minor influence even for small populations. This is caused by the fact that transitions to $\{\{1,3\},\{2\}\}$ are either possible via a double crossover (with small probability) from the state $\{\{1,2,3\}\}$ or when blocks separate first and coalesce again (either in one step or in two consecutive steps). Especially for large populations, the reunion of blocks becomes increasingly unlikely.

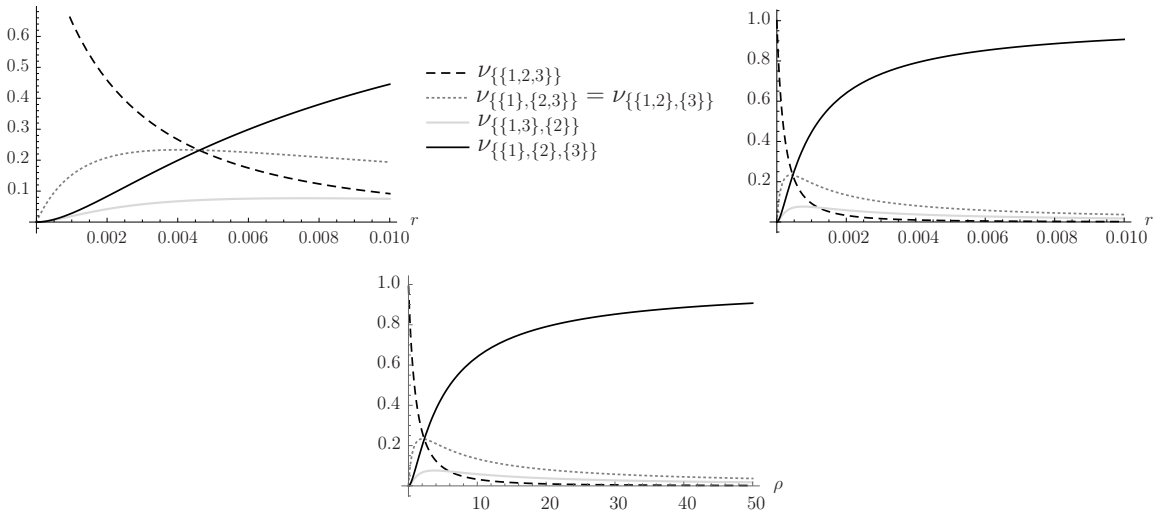


Figure 4.2. Stationary distribution $\nu = (\nu_{\mathcal{A}})_{\mathcal{A}\in\mathbb{P}(\{1,2,3\})}$ of $(\Sigma_t)_{t\geqslant 0}$ with respect to recombination strength ($r := r_1 = r_2$, $r_{12} = 0.8\,r$); comparing different population sizes. Left upper panel: $N = 500$. Right upper panel: $N = 5000$. Lower panel: diffusion limit ($\rho := \rho_1 = \rho_2$, $\rho_{12} = 0.0003\,\rho$).

The example with three sites allows one to study the effect of double crossovers. It can be seen in Figure 4.3 that, even for comparatively large double-crossover probabilities and small populations ($N < 2000$), the effect of double crossovers is small and expresses itself only in the fact that the curve for $\{\{1,3\},\{2\}\}$ increases faster and rises slightly higher if double crossover are allowed. The increased influence reflects the fact that under single-crossover recombination, one-step transitions from $\{\{1,2,3\}\}$ to $\{\{1,3\},\{2\}\}$ do not occur.

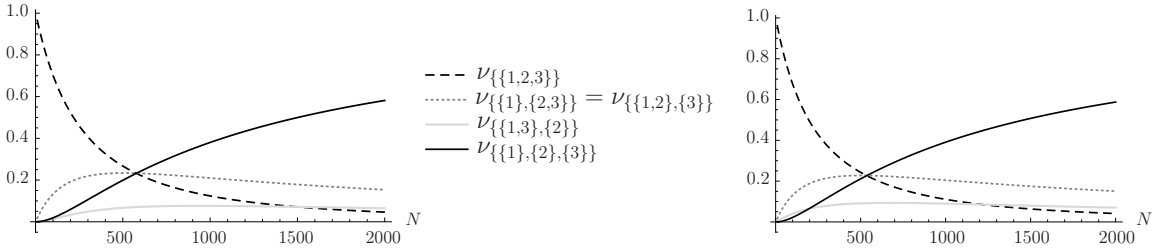


Figure 4.3. Stationary distribution $\nu = (\nu_{\mathcal{A}})_{\mathcal{A}\in\mathbb{P}(\{1,2,3\})}$ with respect to growing population size; comparing the effect of double crossovers. Left: no double crossover ($r_1 = r_2 = 0.004$, $r_{12} = 0$). Right: high probability for a double crossover ($r_1 = r_2 = 0.004$, $r_{12} = 0.001$).

5 Trees in the large population limit

In this chapter, we will continue to investigate the interplay between the forward and backward time perspective of our population models by focussing on the limiting behaviour as population size tends to infinity. More precisely, we will reveal a correspondence between the deterministic forward model and the deterministic limit of the backward model in discrete and continuous time. As in previous chapters, we will stick to the convention that continuous-time variables are indicated by a caron, discrete-time variables by a hat. We omit the indication if statements hold for both of them.

Recall from Section 2.1 that the general (bi-parental) recombination equations in continuous and discrete time are given by

$$\frac{\mathrm{d}}{\mathrm{d}t}\,\check{p}_t = \sum_{\mathcal{A}\in\mathbb{P}_{\leqslant 2}(S)} \varrho_{\mathcal{A}}\,(R_{\mathcal{A}} - \mathbb{1})(\check{p}_t) \quad \text{and} \qquad \hat{p}_{t+1} = \hat{p}_t + \sum_{\mathcal{A}\in\mathbb{P}_2(S)} r_{\mathcal{A}}\,(R_{\mathcal{A}} - \mathbb{1})(\hat{p}_t).$$

For an initial value $p_0 \in \mathcal{P}(\mathbb{X})$, the solution is of the form

$$p_t = \sum_{\mathcal{A}\in\mathbb{P}(S)} a_t(\mathcal{A})\, R_{\mathcal{A}}(p_0), \tag{5.1}$$

where the $a_t(\mathcal{A})$'s need to be determined. Via the usual methods forward in time, it was only possible to find an explicit expression for the continuous-time $\check{a}_t(\mathcal{A})$'s in the single-crossover case. Analogous forward methods in discrete time, even with the additional restriction to single crossovers, did result in a recursive formulation of the $\hat{a}_t(\mathcal{A})$'s only. An explicit expression in the discrete-time, single-crossover setting could nonetheless be found by Baake and von Wangenheim [10, Thm. 4] by studying the backward perspective of the corresponding stochastic model, which is the deterministic limit of the partitioning process restricted to single crossovers. The solution was obtained via a technical calculation and did hint at an underlying principle of inclusion-exclusion that could not be made concrete so far. In this chapter, we present a conceptual proof for this explicit solution based on Möbius inversion on a suitable poset, which reveals the hidden combinatorial and stochastic aspects.

We start with a general connection between the solution of the deterministic recombination equation forward in time and the deterministic limit of the partitioning process backward in time. Recall here, that the deterministic limit of the partitioning process is a process of progressive refinements on the set of partitions of S. Once blocks are separated, they will never come together again, see Section 3.2.2.

Theorem 5.1. *For any $\mathcal{A} \in \mathbb{P}(S)$, the coefficient functions of the solution of the general recombination equation from* (5.1) *satisfy*

$$a_t(\mathcal{A}) = \mathbf{P}\big[\Sigma'_t = \mathcal{A} \,|\, \Sigma'_0 = \mathbf{1}\big], \quad t \geqslant 0,$$

where $(\Sigma'_t)_{t\in\mathbb{T}}$ is the deterministic limit of the partitioning process as in Proposition 3.1 (continuous time) or Proposition 3.2 (discrete time).

A proof for Theorem 5.1 is given in [5] and is based on Haldane linearisation (the properties of the recombinator used in [5] remain true with the slightly different normalisation factor we use in this thesis). A second proof for the discrete-time version of Theorem 5.1 is given by Martinez [93, Thm. 4.2 & Lemma 5.3]. Martinez obtains the connection between $(\widehat{\Sigma}'_t)_{t\in\mathbb{N}_0}$ and the recombination equation by identifying a recursive formulation of the deterministic dynamics in terms of trees whose probability distribution coincides with the one of $(\widehat{\Sigma}'_t)_{t\in\mathbb{N}_0}$. For the continuous-time result, we opt here for an alternative proof based on the duality result in Section 4.2.

Proof. Let $(\check{\Sigma}^{(N)}_t)_{t\geqslant 0}$ be the continuous-time partitioning process from Section 3.2.1 with law ψ and let $(Z^{(N)}_t)_{t\geqslant 0}$ be the Moran model from Definition 2.1 with law φ, where the upper index indicates dependence on population size. Evaluating the duality equation (4.14) for $\check{\Sigma}^{(N)}_0 = \mathbf{1}$, using that the state space of the partitioning process $(\check{\Sigma}^{(N)}_t)_{t\geqslant 0}$ is finite and that $H_{\mathbf{1}}(\check{Z}_t) = \frac{\check{Z}_t}{N}$, yields the following correspondence

$$\mathbf{E}_\varphi\left[\frac{\check{Z}^{(N)}_t}{N} \,\middle|\, \frac{\check{Z}^{(N)}_0}{N} = \frac{z}{N}\right] = \sum_{\mathcal{A}\in\mathbb{P}(S)} \mathbf{P}_\psi\Big[\check{\Sigma}^{(N)}_t = \mathcal{A} \,|\, \check{\Sigma}^{(N)}_0 = \mathbf{1}\Big]\, H_{\mathcal{A}}(\check{Z}^{(N)}_0). \tag{5.2}$$

Now assume that $\lim_{N\to\infty} \frac{\check{Z}^{(N)}_0}{N} = \check{p}_0$ with $\check{p}_0 \in \mathcal{P}(\mathbb{X})$. Starting with the right hand side and using again that the spate space of $(\check{\Sigma}_t)_{t\geqslant 0}$ is finite, gives $\lim_{N\to\infty} H_{\mathcal{A}}(\check{Z}^{(N)}_0) = R_{\mathcal{A}}(\check{p}_0)$ by (4.12). Based on the convergence of $(\check{\Sigma}_t)_{t\geqslant 0}$ to $(\check{\Sigma}'_t)_{t\geqslant 0}$ (Proposition 3.1), we conclude for the right hand side of (5.2) that

$$\lim_{N\to\infty} \sum_{\mathcal{A}\in\mathbb{P}(S)} \mathbf{P}_\psi\Big[\check{\Sigma}^{(N)}_t = \mathcal{A} \,|\, \check{\Sigma}^{(N)}_0 = \mathbf{1}\Big]\, H_{\mathcal{A}}(\check{Z}^{(N)}_0) = \sum_{\mathcal{A}\in\mathbb{P}(S)} \mathbf{P}_\psi\big[\check{\Sigma}'_t = \mathcal{A} \,|\, \check{\Sigma}'_0 = \mathbf{1}\big]\, R_{\mathcal{A}}(\check{p}_0).$$

Since we assumed $\lim_{N\to\infty} \check{Z}^{(N)}_0/N = \check{p}_0$, we obtain by Theorem 2.1 that $\check{Z}^{(N)}_t/N$ converges to $\check{p}_t$ in probability for all $t \geqslant 0$, where $\check{p}_t$ is the solution of the deterministic recombination equation with initial value $\check{p}_0$ from (5.1). This yields the convergence of the left hand side of (5.2) to

$$\lim_{N\to\infty} \mathbf{E}_\varphi\left[\frac{\check{Z}^{(N)}_t}{N} \,\middle|\, \frac{\check{Z}^{(N)}_0}{N} = \frac{z}{N}\right] = \sum_{\mathcal{A}\in\mathbb{P}(S)} \check{a}_t(\mathcal{A})\, R_{\mathcal{A}}(\check{p}_0).$$

Comparison of the coefficients finishes the proof. □

An explicit expression for the $a_t(\mathcal{A})$'s may therefore be obtained by studying the probability distribution of Σ'_t starting in $\Sigma'_0 = \mathbf{1}$. Before we investigate the respective probability distribution in detail, let us first mention some further properties that can be obtained from Theorem 5.1. First, the discrete-time coefficient functions evolve according to the (nonlinear) iteration

$$\hat{a}_{t+1}(\mathcal{A}) = \sum_{\mathcal{B} \succcurlyeq \mathcal{A}} \hat{a}_t(\mathcal{B})\ \hat{\Theta}'_{\mathcal{B}\mathcal{A}},$$

which is the generalisation of the single-crossover iteration in (2.8). Secondly, we may express the type distribution of a sufficiently large population in terms of the ancestral process, that is, if $\lim_{N\to\infty} Z_0^{(N)}/N = p_0 \in \mathcal{P}(\mathbb{X})$, then

$$\lim_{N\to\infty} \frac{Z_t^{(N)}}{N} = \sum_{\mathcal{A}\in\mathbb{P}(S)} \mathbf{P}[\Sigma'_t = \mathcal{A} \,|\, \Sigma'_0 = \mathbf{1}]\ R_{\mathcal{A}}(p_0), \text{ in probability,}$$

for every fixed $t \geqslant 0$. This is the generalisation of Theorem 2 in [10] to multiple crossovers. Since $(\check{\Sigma}_t^{(N)})_{t\in\mathbb{T}}$ has absorbing state $\mathbf{0} = \{\{1\},\{2\},\ldots,\{n\}\}$, the asymptotic behaviour for large populations is given by

$$\lim_{N\to\infty} \frac{Z_\infty}{N} = (\pi_1.\, \check{p}_0) \otimes (\pi_2.\, \check{p}_0) \otimes \ldots \otimes (\pi_n.\, \check{p}_0),$$

which is the counterpart to (2.4).

5.1 Single-crossover recombination: Segmentation process

Apart from some exceptions, most of the properties we studied so far allow multiple crossover events within one reproduction step. For the remaining part of this thesis we now restrict us to single-crossover recombination, for which we are able to use the simplified notation in terms of links rather than partitions of S (cf. Sect. 1.3.2). In Section 5.4, we mention briefly how to apply the method to allow for general, multi-crossover recombination.

Recall here that $L = \{\frac{3}{2},\ldots,\frac{2n-1}{2}\}$ is the set of *links*, where it is understood that link $\alpha \in L$ connects sites $\lfloor\alpha\rfloor$ and $\lceil\alpha\rceil$, see Figure 1.5. Let $G = \{\alpha_1,\ldots,\alpha_{|G|}\}$ be a subset of L with $\alpha_1 < \alpha_2 < \cdots < \alpha_{|G|}$. Every *ordered* partition $\sigma = \{\sigma_1,\ldots,\sigma_{|G|+1}\}$ of $S = \{1,2,\ldots,n\}$ with blocks

$$\sigma_1 = \{1,\ldots,\lfloor\alpha_1\rfloor\}, \quad \sigma_2 = \{\lceil\alpha_1\rceil,\ldots,\lfloor\alpha_2\rfloor\}, \quad \ldots, \quad \sigma_{|G|+1} = \{\lceil\alpha_{|G|}\rceil,\ldots,n\}, \tag{5.3}$$

has a one-to-one correspondence to the set G of removed links. Recall further that r_α (ϱ_α), $\alpha \in L$, is the probability (rate) for a crossover at link α.

Let $(F_t)_{t\in\mathbb{T}}$ be the single-crossover version of $(\Sigma'_t)_{t\in\mathbb{T}}$ with initial value $\Sigma'_0 = \mathbf{1}$. We saw in Section 3.2.2 that if $\Sigma'_0 \in \mathbb{O}(S)$ (the set of ordered partitions of S), then Σ'_t takes values in $\mathbb{O}(S)$ for all times (cf. Fact 3.1). Due to the one-to-one correspondence between $\mathbb{O}(S)$ and $\wp(L)$ (the set of subsets of L), we may thus describe $(F_t)_{t\in\mathbb{T}}$ as a Markov chain on $\wp(L)$, where F_t is the set of links that have been removed until time t, see Figure 5.1. The absorbing state of $(F_t)_{t\in\mathbb{T}}$ is L and obviously $F_{t'} \subseteq F_t$ for all $t' < t$.

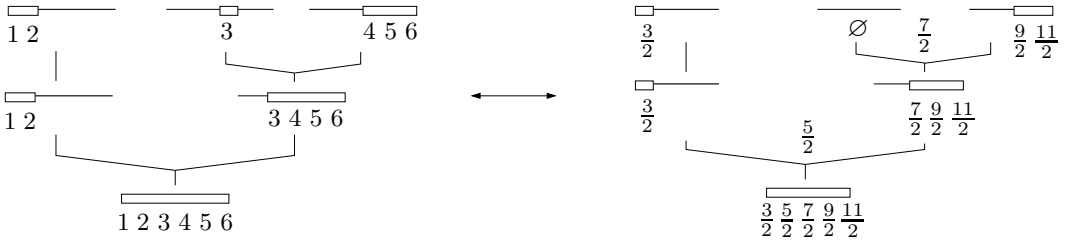


Figure 5.1. Left: Single-crossover version of the deterministic limit of the partitioning process $(\Sigma'_t)_{t\in\mathbb{T}}$ with start in $\Sigma'_0 = \{1, 2 \ldots, 6\}$. Right: Corresponding process $(F_t)_{t\in\mathbb{T}}$ in the link notation; $L = \{\frac{3}{2}, \frac{5}{2}, \ldots, \frac{11}{2}\}$. Here, $\mathcal{L}_{\{\frac{5}{2},\frac{7}{2}\}} = \{\{\frac{3}{2}\}, \varnothing, \{\frac{9}{2}, \frac{11}{2}\}\}$.

If a link of L is removed, the remaining set of links is decomposed into two contiguous subsets of links. If all links in $G = \{\alpha_1, \ldots, \alpha_{|G|}\} \subseteq L$ with $\alpha_1 < \alpha_2 < \cdots < \alpha_{|G|}$ are removed, G induces a decomposition of the set of remaining segments of links into

$$\mathcal{L}_G := \{J_1, \ldots, J_{|G|+1}\}, \tag{5.4}$$

$$J_1 = \{\alpha \in L : \alpha < \alpha_1\}, \quad J_2 = \{\alpha \in L : \alpha_1 < \alpha < \alpha_2\}, \quad \ldots, \quad J_{|G|+1} = \{\alpha \in L : \alpha_{|G|} < \alpha\};$$

in particular, $\mathcal{L}_\varnothing = \{L\}$ and $\mathcal{L}_L = \{\varnothing\}$. Clearly, such a J_i may be empty, and $\mathcal{L}_G \setminus \varnothing$ is a partition of $L \setminus G$. Since $(F_t)_{t\in\mathbb{T}}$ decomposes $L \setminus F_t$ into ordered segments of links, we speak of $(F_t)_{t\in\mathbb{T}}$ as the *segmentation process*.

Translating the partition notation in Proposition 3.1 and Proposition 3.2 to the link notation and assuming single-crossover recombination, leads to the following definitions.

Definition 5.1 (Segmentation process, continuous time)**.** $(\check{F}_t)_{t\geqslant 0}$ is the continuous-time Markov chain with values in $\wp(L)$, initial value $\check{F}_0 = \varnothing$ and transitions $\check{F}_t \longrightarrow \check{F}_t \cup \{\alpha\}$, which occur at rate ϱ_α for every $\alpha \in L \setminus \check{F}_t$ and $t \geqslant 0$. No other transitions are possible.

We may alternatively say that $\check{F}_t \longrightarrow \check{F}_t \cup \{\alpha\}$ occurs at rate ϱ_α for every $\alpha \in L$, which indicates that the transitions are independent of the current state.

Definition 5.2 (Segmentation process, discrete time)**.** $(\widehat{F}_t)_{t\in\mathbb{N}_0}$ is the following discrete-time Markov chain with values in $\wp(L)$: The initial state is $\widehat{F}_0 = \varnothing$ and if $\widehat{F}_{t-1} = G$, then

$$\widehat{F}_t = \widehat{F}_{t-1} \cup \Big(\bigcup_{J\in\mathcal{L}_G} A_t^J \Big).$$

Here $A_t^J = \{\alpha\}$ with probability r_α for all $\alpha \in J$, and $A_t^J = \varnothing$ with probability $1 - \sum_{\alpha\in J} r_\alpha$ independently for all $J \in \mathcal{L}_G$ and all $t \geqslant 1$. $\mathcal{L}_G$ is defined as in (5.4).

The definition deals consistently with empty segments since $A_t^\varnothing = \varnothing$ with probability 1. Links are dependent as long as they belong to the same segment and become independent once they are separated on different segments. We can therefore represent $(\widehat{F}_t)_{t\geqslant t'}$ as

$$\widehat{F}_t^L = \widehat{F}_{t'}^L \cup \Big(\bigcup_{J\in\mathcal{L}_{\widehat{F}_{t'}^L}} \widehat{F}_t^J \Big), \quad t' \geqslant 0,\ t \geqslant t'. \tag{5.5}$$

The $(\widehat{F}_t^J)_{t\geqslant t'}$'s are independent processes with $\widehat{F}_{t'}^J = \varnothing$ and $(\widehat{F}_t^J)_{t\geqslant t'}$ defined in analogy with $(\widehat{F}_t^L)_{t\in\mathbb{N}_0} := (\widehat{F}_t)_{t\in\mathbb{N}_0}$. That is, $(\widehat{F}_t^J)_{t\geqslant t'}$ is the segmentation process defined on the underlying set of links J with removal probabilities r_α, $\alpha \in J$. Throughout, we use the upper index to indicate the underlying set of links and may omit it if the set is L.

The analogue statement of Theorem 5.1 now reads:

Fact 5.1. Let G be a subset of L with $G = \{\alpha_1, \dots, \alpha_{|G|}\}$ and $\alpha_1 < \alpha_2 \cdots < \alpha_{|G|}$. If $\mathcal{A} = \{\sigma_1, \dots, \sigma_{|G|+1}\}$ is an ordered partition of S with blocks as in (5.3), then

$$a_t(\mathcal{A}) = \mathbf{P}[\Sigma'_t = \mathcal{A} \,|\, \Sigma'_0 = \mathbf{1}] = \mathbf{P}[F_t = G] = a_t(G), \quad t \geqslant 0,$$

where a_t is the coefficient function corresponding to (5.1). □

Our interest is therefore in the probability distribution of F_t. We will throughout rely on a formulation via waiting times. Let $\mathcal{T}_\alpha := \min\{t \geqslant 0 : \alpha \in F_t\}$ be the waiting time for the link α to be removed and $\mathcal{T}_K := \min\{\mathcal{T}_\alpha : \alpha \in K\}$ the time at which the first link in $K \subseteq L$ is removed; denote by $\widehat{\mathcal{T}}$ ($\check{\mathcal{T}}$) the corresponding discrete-time (continuous-time) versions. The event $\{F_t = G\}$ then obviously translates into

$$\{F_t = G\} = \{\max\{\mathcal{T}_\alpha : \alpha \in G\} \leqslant t < \mathcal{T}_{L\setminus G}\}, \quad G \subseteq L,\ t \geqslant 0. \tag{5.6}$$

In continuous time, each $\alpha \in L$ is independently removed after an exponential waiting time $\check{\mathcal{T}}_\alpha$ with parameter ϱ_α. The explicit expression for the probability of (5.6) is therefore immediate:

$$\mathbf{P}[\check{F}_t = G] = \prod_{\alpha\in G} \mathbf{P}[\check{\mathcal{T}}_\alpha \leqslant t] \prod_{\beta\in L\setminus G} \mathbf{P}[\check{\mathcal{T}}_\beta > t] = \prod_{\alpha\in G} (1 - \exp(-\varrho_\alpha\, t)) \prod_{\beta\in L\setminus G} \exp(-\varrho_\beta\, t), \tag{5.7}$$

and we rediscover the continuous-time coefficients $\check{a}_t(G)$ from (2.6).

In discrete time, however, the links are dependent: removing a given link forbids to remove any other link in the same segment in the same time step. For each realisation of $(\widehat{F}_t)_{t\in\mathbb{N}_0}$, the order of events therefore matters. One may thus collect all realisations of $(\widehat{F}_t)_{t\in\mathbb{N}_0}$ that agree on the order of events and that end up in the state $\widehat{F}_t = G$ at time t, and represent this set of realisations as a rooted binary tree (cf. Fig. 5.2). Here, addition of an element to $\widehat{F}_t$ is identified with a vertex of that tree in such way that the time series of events of the segmentation process is encoded by the partial order on the vertices of the tree. The root, for instance, represents the link that is removed first (irrespective of the precise time at which it is removed). Obviously, $\mathbf{P}[\widehat{F}_t = G]$ can then be obtained from the sum over all probabilities of trees with vertex set G.

In [10], the probability for each individual tree was obtained from a technical calculation by summing over all possible combinations of branch lengths, i.e. over all possible combinations of times that F_t spends in the various states. This summation led to an alternating sum over terms that reflect a decomposition of the tree into subtrees. The result provided an answer to the problem, but was somewhat unsatisfactory since both the combinatorial and the probabilistic meaning remained in the dark. As to the combinatorial side, the

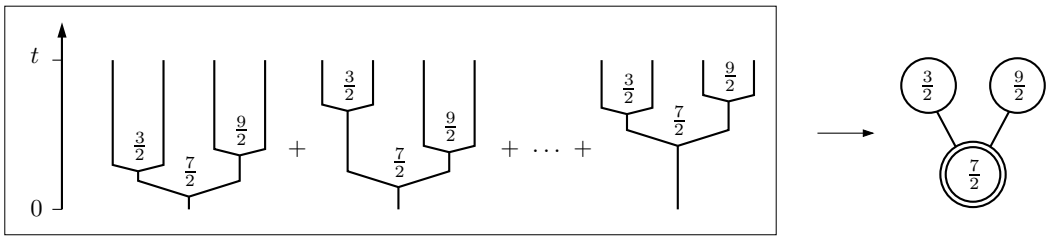


Figure 5.2. All realisations of the segmentation process that end up in the state $F_t = \{\frac{3}{2}, \frac{7}{2}, \frac{9}{2}\}$ at time t and for which $\frac{7}{2}$ is removed before $\frac{3}{2}$ and $\frac{9}{2}$ are represented by the rooted binary tree on the right.

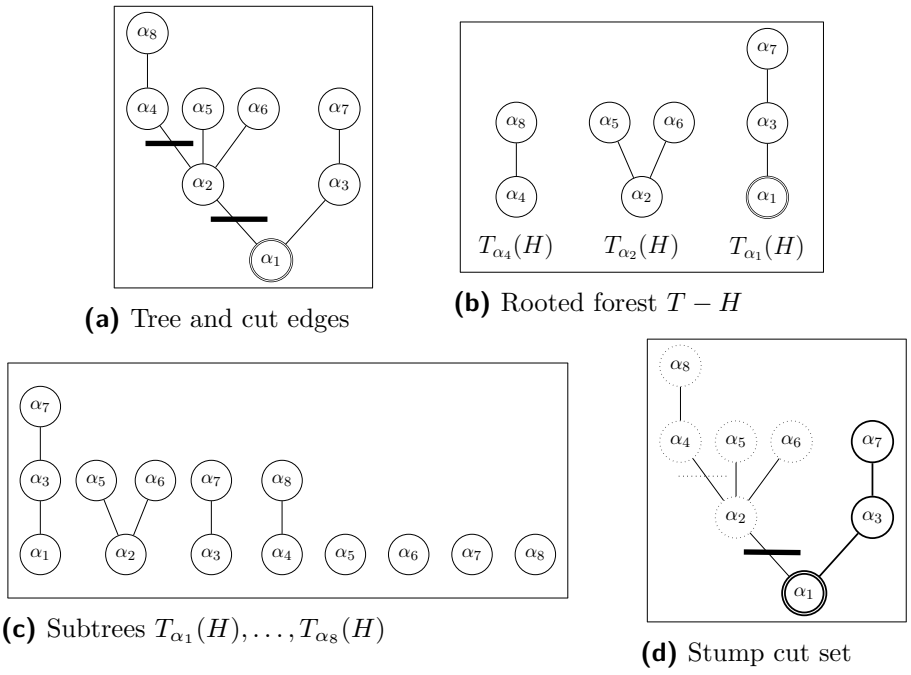
alternating sum hinted at an underlying, yet unidentified, inclusion-exclusion principle. As to the probabilistic side, the terms in the sum hinted at some underlying independence across subtrees, but were hard to interpret in detail. The purpose of the remaining sections is to give a conceptual proof for the distribution of the segmentation process based on the graphical representation via trees.

To this end, we will start with a general investigation of trees and rooted forests independent from the concrete relation to the segmentation process. We then construct a suitable poset (to be called *pruning poset*) on rooted forests, find its Möbius function and give the corresponding Möbius inversion principle. Thereafter (Section 5.3), we relate the rooted trees to sets of realisations of $(F_t)_{t\in\mathbb{T}}$ mentioned above and use Möbius inversion on the pruning poset to obtain an explicit expression for the tree probabilities.

5.2 Möbius inversion on a poset of rooted forests

Let $T = (\gamma, V, E)$ denote a rooted tree with root γ, vertex (or node) set $V = V(T)$ and set of edges $E = E(T) \subseteq V \times V$. The set of vertices together with the standard partial order on rooted trees defines a poset $(V, \preccurlyeq)$ (see Section 1.2 for a brief introduction into poset theory). Namely, for any two nodes $\alpha, \beta \in V$, $\alpha \preccurlyeq \beta$ means that α is on the path from γ to β. Obviously, γ is the minimal element of V with respect to $\preccurlyeq$. If α and β are adjacent and $\alpha \prec \beta$, we write $e = (\alpha, \beta)$ and call α and β the *ends* of e; more precisely, α is the *lower end* and β the *upper end* of e. The partial order on V obviously induces a partial order on E (via the partial order of the upper ends, say), which we will (by slight abuse of notation) also denote by $\preccurlyeq$.

For a fixed tree $T = (\gamma, V, E)$ and a given subset H of E, we denote by $T - H$ the *rooted forest* obtained from T by deleting all edges $e \in H$; we speak of these edges as *cut edges* (see Figure 5.3a). The remaining *connected components* (or *components*) of T are disjoint rooted trees, where the root in each component is the unique vertex that is minimal with respect to $\preccurlyeq$. For all $\alpha \in V$, we denote by $T_\alpha(H)$ the subtree in $T - H$ that is rooted at α, i.e. that consists of α and all its descendants, see Figure 5.3c. By slight abuse of notation, we abbreviate the corresponding vertex and edge sets by $V_\alpha(H) := V(T_\alpha(H))$ and $E_\alpha(H) := E(T_\alpha(H))$, respectively. The rooted forest $T - H$ then is the disjoint collection of all $T_\alpha(H)$ with $\alpha = \gamma$ or α an upper end of some $e \in H$ (cf. Fig. 5.3b).



(a) Tree and cut edges

(b) Rooted forest $T - H$

(c) Subtrees $T_{\alpha_1}(H), \ldots, T_{\alpha_8}(H)$

(d) Stump cut set

Figure 5.3. (a) Tree $T = (\gamma, V, E)$ with root $\gamma = \alpha_1$, vertex set $V = \{\alpha_1, \ldots, \alpha_8\}$ and pruning edges $H = \{(\alpha_1, \alpha_2), (\alpha_2, \alpha_4)\}$. (b) The forest $T - H$ obtained from the tree in (a). The stump tree is $T_{\alpha_1}(H)$ with stump set $R = V_\gamma(H) = \{\alpha_1, \alpha_3, \alpha_7\}$. The root of the stump tree is indicated by a double circle since it coincides with the root of T. (c) Collection of all subtrees $T_\alpha(H)$, $\alpha \in V$, in the forest $T - H$; T and H from (a). (d) The stump cut set for the stump set in (b) is $\partial(R) = \{(\alpha_1, \alpha_2)\}$. The stump set and the stump cut set are indicated in bold.

For a given forest $T - H$, a special role is played by the subtree $T_\gamma(H)$, whose root coincides with the root of T. We call this tree the *stump tree* of the rooted forest and say its vertex set $V_\gamma(H)$ is the *stump set*. Explicitly, for $H = \{e_1, \ldots, e_k\}$ with $e_i = (\alpha_i, \beta_i)$, $1 \leqslant i \leqslant k$,

$$V_\gamma(H) = V \setminus \{\nu \in V : \nu \succ \alpha_i \text{ for some } 1 \leqslant i \leqslant k\}. \tag{5.8}$$

We denote the set of all possible stump sets by

$$\mathcal{R}(T) := \{V_\gamma(H) : H \subseteq E\}. \tag{5.9}$$

Any stump set may be defined via a special set of cut edges. For a given $R \in \mathcal{R}(T)$, we denote by $\partial(R)$ the set of edges that separates R from the remaining set of vertices $V \setminus R$ and call it the *stump cut set* of R, compare Figure 5.3d. Explicitly,

$$\partial(R) := \{(\alpha, \beta) \in E : \alpha \in R,\, \beta \in V \setminus R\};$$

in particular, $\partial(V) = \varnothing$. The set of all stump cut sets is

$$\mathcal{C}(T) := \{\partial(R) : R \in \mathcal{R}(T)\}. \tag{5.10}$$

Obviously, every singleton set $\{e\}$, $e \in E$, is a stump cut set, and every stump cut set C satisfies $C = \partial(V_\gamma(C))$.

Fact 5.1. A subset H of E is a stump cut set if and only if it satisfies $H = M(H)$, where

$$M(H) := \{e \in H : e \text{ is minimal in } H \text{ with respect to } \preccurlyeq\}, \qquad M(\varnothing) := \varnothing. \tag{5.11}$$

Proof. Consider a set $H \subseteq E$. Suppose that $H = M(H)$. Then any two edges $e_i, e_j \in H$ $(i \neq j)$ are incomparable with respect to $\preccurlyeq$ (neither $e_i \preccurlyeq e_j$ nor $e_j \preccurlyeq e_i$, $i \neq j$). As a consequence, the corresponding stump set $R = V_\gamma(H)$ of (5.8) satisfies $\partial(R) = H$, so H is a stump cut set. On the other hand, assume that $H \neq M(H)$. Then there are two edges $e_1 = (\alpha_1, \beta_1)$, $e_2 = (\alpha_2, \beta_2) \in H$ with $e_1 \preccurlyeq e_2$. There is then no $R \in \mathcal{R}(T)$ such that $\alpha_1, \alpha_2 \in R$ and $\beta_1, \beta_2 \in V \setminus R$, so H cannot be a stump cut set. □

Due to Fact 5.1, the set of all stump cut sets may be characterised by $\mathcal{C}(T) = \{H \subseteq E : H = M(H)\}$, and we can rewrite (5.9) as

$$\mathcal{R}(T) = \{V_\gamma(C) : C \in \mathcal{C}(T)\}. \tag{5.12}$$

Fact 5.2. For every $H \subseteq E$, the components of $T - H$ have the following properties:

(A) $\big(T_\gamma(H)\big)_\alpha(K) = T_\alpha(H \cup K)$, $\alpha \in V_\gamma(H)$, $K \subseteq E_\gamma(H)$

(B) $T_\alpha(H) = T_\alpha(H \cup C)$ for $C \in \mathcal{C}(T_\gamma(H))$ and $\alpha \notin V_\gamma(H \cup C)$.

These properties carry over to the corresponding vertex sets of the rooted trees.

Proof. (A) is due to a general property of graph decomposition via recursive edge deletion: the order in which edges are deleted does not affect the final object. So $(T - H)(K) = T - (H \cup K)$ for all $H, K \subseteq E$; in particular, the stump tree is the same in both cases. (B) For every $C \in \mathcal{C}(T_\gamma(H))$ and $\alpha \notin V_\gamma(H \cup C)$, we have $C \cap E_\alpha(\varnothing) = \varnothing$ due to Fact 5.1. But a subtree $T_\alpha(H)$ is not affected by deletion of an edge $e \notin E_\alpha(\varnothing)$. □

For a fixed tree $T = (\gamma, G, E)$, let us now investigate some interesting relations between notions we introduced for rooted forests and notions from general poset theory (see Sect. 1.2).

Fact 5.3. $\mathcal{C}(T)$ is the set of all *antichains* of the poset $(E, \preccurlyeq)$, that is, $\mathcal{C}(T)$ is the set of all subsets of E for which all elements are incomparable with respect to $\preccurlyeq$.

Proof. Consider a subset $H \subseteq E$ of edges for which $e \preccurlyeq f$, $e, f \in H$. Then H cannot be a stump cut set since H is not minimal. Since the empty set and all singletons $\{e\}$, $e \in H$, obviously satisfy the antichain condition, we see that every stump cut set is an antichain. On the other hand, let A be an antichain of $(E, \preccurlyeq)$. Then there exists a stump set $R \in \mathcal{R}(T)$ of the form $R = V_\gamma(A)$, which satisfies $A = \partial(V_\gamma(A))$. We conclude that $A \in \mathcal{C}(T)$. □

Fact 5.4. A subset of vertices of a tree $T = (\gamma, V, E)$ is a stump set if and only if it is a nonempty ideal of $(V, \preccurlyeq)$, that is, if any only if it is a subset I of V for which follows that if $\alpha \in I$ and $\beta \preccurlyeq \alpha$, then $\beta \in I$ (see Section 1.2).

Proof. Consider the poset $(V, \preccurlyeq)$ and let $I \subseteq V$ be a nonempty ideal of $(V, \preccurlyeq)$. Per construction, there is then a set $C = \partial(I)$ that separates I from the remaining set of vertices $V \setminus I$. Moreover, $I = V_\gamma(\partial(I))$ so that $I \in \mathcal{R}(T)$ by (5.12). Since, on the other hand, every $R \in \mathcal{R}(T)$ obviously satisfies the condition for ideals, the claim follows. □

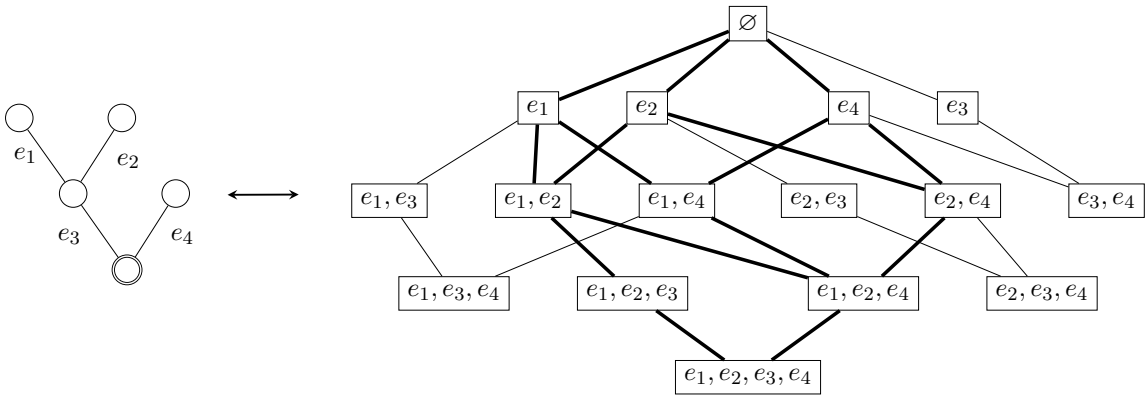


Figure 5.4. Left: Tree T with edges $\{e_1, \dots, e_4\}$. Right: Hasse diagram for the pruning poset $\mathcal{D}(T)$ of T. The subposet $[H, \varnothing]$ is indicated in bold.

5.2.1 Pruning poset

From now on, let $T = (\gamma, V, E)$ be fixed and let us investigate the set $\wp(E)$ of all sets of subsets of E, where each set represents a set of cut edges. We introduce a partial order $\preceq_D$ on $\wp(E)$ and say that $H \preceq_D K$ for any two sets of cut edges $H, K \subseteq E$ when $H = K \cup A$ with $A \subseteq E_\gamma(K)$. In words, $H \preceq_D K$ whenever the additional cuts in $H \setminus K$ occur in the stump tree of the rooted forest $T - K$. The set $\wp(E)$ along with the partial order $\preceq_D$ constitutes a poset $\mathcal{D}(T) := (\wp(E), \preceq_D)$. Since the cut edges prune the tree (in an intuitive way of thinking), we call $\mathcal{D}(T)$ the *pruning poset* of T. A specific example with corresponding Hasse diagram is shown in Figure 5.4. For every $K \subseteq E$, we clearly have the isomorphic relation

$$(\{H : H \preceq_D K\}, \preceq_D) \simeq \mathcal{D}(T_\gamma(K)). \tag{5.13}$$

$\mathcal{D}(T)$ has a maximal element $\varnothing$ but in general no minimal element. As a consequence, $\mathcal{D}(T)$ is, in general, not a lattice. Nonetheless, every embedded subposet or *interval*[1]

$$[H, K]_D := \big(\{I \subseteq E : H \preceq_D I \preceq_D K\}, \preceq_D\big), \quad H \preceq_D K$$

of $\mathcal{D}(T)$ is a lattice. We omit the subscript in what follows. Due to (5.13), we conclude the isomorphic relation $[H, K] \simeq [H \setminus K, \varnothing]$ for any $H \preceq_D K$. It is therefore sufficient to investigate the properties of $[H, \varnothing]$ for every $H \subseteq E$. The interval $[H, \varnothing]$ obviously has maximal element $\varnothing$ and minimal element H. Every path (top to bottom) in $[H, \varnothing]$ represents the possibility to add elements from H in nonincreasing order with respect to $\preccurlyeq$.

Remark 5.1. The idea of successively cutting edges within the stump tree of a rooted forest is reminiscent of the cutting-down procedure introduced by Meir and Moon [97]. In [97], the root of a random tree is isolated by uniformly cutting edges of the tree until the tree is reduced to the root. In the resulting line of research (see, for example, [35, 72, 97, 109]), one is interested in the distribution and limiting behaviour of the number of cuts required to isolate the root for various classes of random trees. In contrast, for our pruning, we keep track of the entire rooted forest, rather than the stump tree alone. ♢

[1] In contrast to Section 1.2, we define intervals here as *subposets* rather than subsets.

There is again an interesting relation to notions from general poset theory.

Fact 5.5. Let $T = (\gamma, V, E)$ be given and consider a subset of edges $H \subseteq E$. If $(H, \preccurlyeq)^*$ is the dual poset of $(H, \preccurlyeq)$, then

$$[H, \varnothing] = \big(J\big((H, \preccurlyeq)^*\big), \subseteq\big)^*,$$

where $J\big((H, \preccurlyeq)^*\big)$ denotes the set of all (possibly empty) ideals of $(H, \preccurlyeq)^*$.

Proof. Let $(H, \preccurlyeq)$ be a subposet of $(E, \preccurlyeq)$ and I be an ideal of the corresponding dual poset $(H, \preccurlyeq)^*$. Then, due to the definition of ideals (cf. Section 1.2) and the reversed partial order, the following condition holds:

$$x \in I \text{ and } y \succcurlyeq x \;\Rightarrow\; y \in I \subseteq H. \tag{5.14}$$

In other words, it follows for every $x \in I$ and $y \in H \setminus I$ that $y \not\succcurlyeq x$. Since per definition $E_\gamma(I) = \{y : y \not\succcurlyeq x \text{ for all } x \in I\}$, Equation (5.14) is equivalent to

$$H = I \cup X \text{ with } X \subseteq E_\gamma(I)$$

and thus to $I \in [H, \varnothing]$. This gives the equality of the sets

$$J\big((H, \preccurlyeq)^*\big) = \{K \subseteq E : H \preceq_D K\}.$$

In the end, since the poset $(J((H, \preccurlyeq)^*), \subseteq)$ is ordered by inclusion, for any two elements $I_1, I_2 \in J((H, \preccurlyeq)^*)$ we have $I_1 \preccurlyeq_J I_2$ if $I_2 \subseteq I_1$, where $\preccurlyeq_J$ denotes the partial order on $\big(J\big((H, \preccurlyeq)^*\big), \subseteq\big)^*$. Analogously, it follows for two elements $K_1, K_2 \in [H, \varnothing]$ that $K_1 \preceq_D K_2$ if they satisfy $K_2 \subseteq K_1$. Altogether, the two posets agree on their underlying sets as well on their partial order. □

Since the lattice of (order) ideals of a poset equipped with the setinclusion order is distributive [123, p. 106], we conclude that every interval $[H, \varnothing]$, $H \subseteq E$, of $\mathcal{D}(T)$ is distributive.

Every atom of $[H, \varnothing]$ is of the form $H \setminus \{\nu\}$, where ν is minimal in H with respect to $\preccurlyeq$. As illustrated in Figure 5.5, $M(H)$, the set of all minimal edges of H from (5.11), induces a product structure of $[H, \varnothing]$:

$$[H, \varnothing] = \bigtimes_{e \in M(H)} [H_e, \varnothing], \quad \text{with} \quad H_e := \{f \in H : f \succcurlyeq e\}. \tag{5.15}$$

For $H = \varnothing$, this remains true with the convention that the empty product is the empty set. With the help of this product structure, we can now calculate the Möbius function for the pruning poset.

Theorem 5.2. *For a given tree $T = (\gamma, V, E)$, the Möbius function for the pruning poset $\mathcal{D}(T)$ is, for every $H, K \subseteq E$ with $H \preceq_D K$, given by*

$$\mu(H, K) = \begin{cases} (-1)^{|H|-|K|}, & \text{if } H \setminus K \in \mathcal{C}(T_\gamma(K)), \\ 0, & \text{otherwise.} \end{cases} \tag{5.16}$$

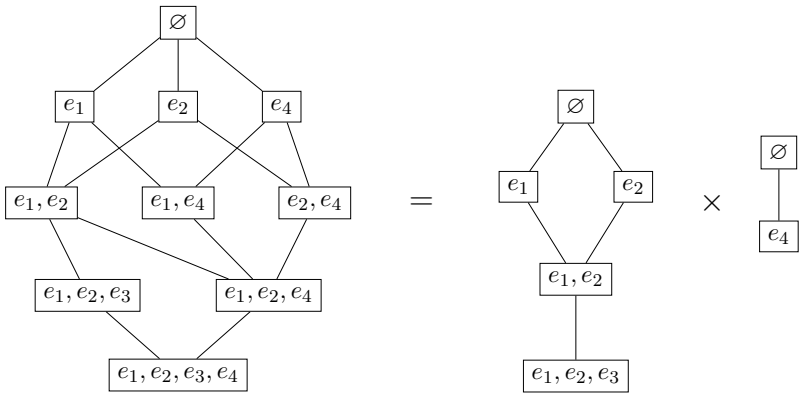


Figure 5.5. Left: The embedded subposet $[H, \varnothing]$, $H = \{e_1, e_2, e_3, e_4\}$, of $\mathcal{D}(T)$ from Figure 5.4. Here, $M(H) = \{\alpha_3, \alpha_4\}$ with $H_{\alpha_3} = \{\alpha_1, \alpha_2, \alpha_3\}$ and $H_{\alpha_4} = \{\alpha_4\}$. As illustrated on the right hand side, the interval $[H, \varnothing]$ can be represented as the direct product $[H_{\alpha_3}, \varnothing] \times [H_{\alpha_4}, \varnothing]$.

Proof. Consider the interval $[H, \varnothing]$ for $H \subseteq E$. Equation (5.15) together with the elementary product theorem for Möbius functions (Proposition 1.1) allows one to decompose the Möbius function into

$$\mu(H, \varnothing) = \prod_{e \in M(H)} \mu(H_e, \varnothing), \tag{5.17}$$

where $M(H)$ and H_e are defined as in (5.11) and (5.15). If $H = \varnothing$, (5.17) remains true under the usual convention that the empty product is 1. Now assume $H \neq \varnothing$, fix an $e \in M(H)$ and consider the interval $[H_e, \varnothing]$. In contrast to $[H, \varnothing]$, the subinterval $[H_e, \varnothing]$ has a *unique* atom for every choice of e and H; this is $H_e \setminus \{e\}$ (cf. Fig. 5.5). It follows immediately from (1.4) that $\mu(H_e, H_e) = 1$ and $\mu(H_e, H_e \setminus \{e\}) = -1$. If $H_e \neq \{e\}$, there is at least one element K with $K \succ_D H_e$ that covers the atom $H_e \setminus \{e\}$. Each interval $[H_e, K]$ is therefore a chain of length two and $\mu(H_e, K) = 0$ by (1.4). Again by (1.4), the property $\mu(H_e, I) = 0$ carries over to every other element I with $I \succ_D H_e \setminus \{e\}$. Together with the isomorphic relation $[H, K] \simeq [H \setminus K, \varnothing]$ for $H \preceq_D K$, this yields for every $e \in M(H)$:

$$\mu(H_e, \varnothing) = \begin{cases} -1, & \text{if } H_e = \{e\}, \\ 0, & \text{otherwise.} \end{cases} \tag{5.18}$$

For $\varnothing \neq H \subseteq E$, the statement $H_e = \{e\}$ for all $e \in H$ is equivalent to $H = M(H)$. We therefore conclude from (5.17) and (5.18) that for $H \subseteq E$:

$$\mu(H, \varnothing) = \begin{cases} (-1)^{|H|}, & \text{if } H = M(H), \\ 0, & \text{otherwise,} \end{cases} \tag{5.19}$$

which also includes the case $H = \varnothing$ mentioned initially. Let now $H, K \subseteq E$. Again, due to the isomorphism in (5.13), we obtain from (5.19) that $\mu(H, K) = (-1)^{|H|-|K|}$ if $H \preceq_D K$ and $H \setminus K = M(H \setminus K)$; $\mu(H, K) = 0$ otherwise. Finally, $H \preceq_D K$ entails $H \setminus K \subseteq E(T_\gamma(K))$. The claim follows from Fact 5.1. □

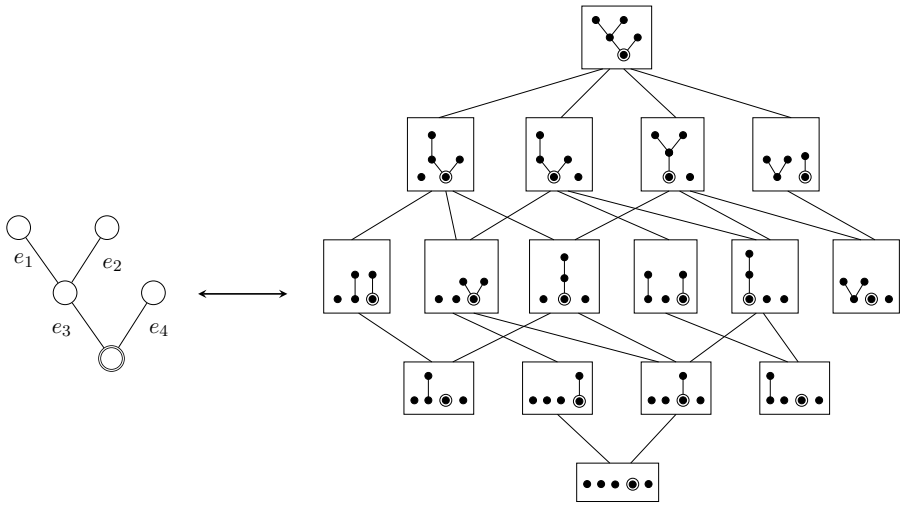


Figure 5.6. Poset $(\mathbf{F}(T), \preceq_F)$ of rooted forests obtained from T by deleting edges in the stump tree only. The connected components in each rooted forest are ordered according to the left-to-right order on the vertices of the tree (so every forest is a planar forest).

Remark 5.2. Using the more abstract representation of $[H, \varnothing]$ from Fact 5.5 together with the antichain property of $\mathcal{C}(T)$ from Fact 5.3, one can also deduce (5.16) from a general result for Möbius functions of lattices of the form $(\{I : I \text{ ideal of } P\}, \subseteq)$, P any poset; see Example 3.9.6 in [123]. We opt for the direct approach here since it is simple and yields additional insight into the structure of the pruning poset. ♢

Now that we have an explicit expression for the Möbius function, we can use Möbius inversion (see Thm. 1.1) on $\mathcal{D}(T)$, which for any two functions $f, g\colon \mathcal{D}(T) \to \mathbb{R}$ and any two subsets $H, K \subseteq E$ reads

$$g(K) = \sum_{H \preceq_D K} f(H) \;\Leftrightarrow\; f(K) = \sum_{H \preceq_D K} \mu(H, K)\; g(H). \tag{5.20}$$

So far, we focussed on the set $\wp(E)$ of all possible subsets of edges of a given tree $T = (\gamma, V, E)$. Let us now shift the perspective to the set $\mathbf{F}(T)$ of all rooted forests that can be obtained from T by edge deletion. Obviously, there is a one-to-one correspondence between the elements of $\mathbf{F}(T)$ and those of $\wp(E)$. We may thus equip $\mathbf{F}(T)$ with a partial order $\preceq_F$ by specifying that $T - H \preceq_F T - K$ precisely if $H \preceq_D K$ for $H, K \subseteq E$, see Figure 5.6. It is clear that $(\mathbf{F}(T), \preceq_F)$ is isomorphic to $\mathcal{D}(T) := (\wp(E), \preceq_D)$ by construction. All properties of $\mathcal{D}(T)$, such as isomorphism, the Möbius function in (5.16), as well as the Möbius inversion formula in (5.20) therefore carry over to $(\mathbf{F}(T), \preceq_F)$.

The poset $(\mathbf{F}(T), \preceq_F)$ is a special case of the *poset of planar forests* introduced by Foissy [50], restricted to what he calls transformations of the second kind and applied to the stump tree only. Foissy also uses Möbius inversion on his more general poset of planar forests. He calculates the Möbius function for small examples, but does not give a general formula. Fortunately, our special case has enough structure to allow for a simple, general and explicit result. This will be the key to an explicit expression for the tree probabilities in the context of the segmentation process $(F_t)_{t \geqslant 0}$.

5.3 Segmentation trees

Let us now return to the segmentation process and relate sets of realisations of $(F_t)_{t\in\mathbb{T}}$ that agree on the time series of events until time t to the trees from Section 5.2. To this end, note that the trees discussed in Section 5.2 did not assume any left-to-right order on the vertices of the tree. From now, let us consider the vertex set of a tree as a subset of removed links $L = \{\frac{1}{2}, \frac{3}{2}, \dots, \frac{2n-1}{2}\}$, denote it by G and equip it with a left-to-right order according to $\leqslant$. Any tree $T = (\gamma, G, E)$ is then a plane oriented tree. Let us, moreover, add information to the plane oriented trees about the link set L and the segments induced by subsets of G as follows.

Let $\mathcal{S} := \bigcup_{R\in\mathcal{R}(T)\cup\varnothing} \mathcal{L}_R$, where $\mathcal{R}(T)$ is the set of stump sets as in (5.9) or (5.12), and where $\mathcal{L}_R$ is defined as in (5.4). Cleary, $\mathcal{S}$ depends on T, but we suppress the dependence on T in the notation. A *segmentation tree* $T^L := (\gamma, G, E, L)$ corresponding to the tree $T = (\gamma, G, E)$ is then the augmented planted plane tree constructed as follows (see Fig. 5.7 for an example):

- Add additional lines to T such that every vertex $\alpha \in G$ has exactly two lines emanating from it. We call these additional lines *branches* and distinguish them from edges. More precisely, a *branch* has a lower end and no upper end in the vertex set of T and an edge always connects two vertices.
- Add a *phantom* node r to the tree. That is, r is the parent of γ, but does not count as a vertex (this makes T^L a planted plane tree [34]). Connect r and γ by a branch.
- Associate every line (edge or branch) with a segment $J \in \mathcal{S}$ according to the following rules: Start with the line between r and γ and identify it with $I_\gamma = L$. Next, associate the two lines emanating from γ with the segments $I'_\gamma := \{\beta \in I_\gamma : \beta < \gamma\}$ and $I''_\gamma := \{\beta \in I_\gamma : \beta > \gamma\}$: so I'_γ is the left and I''_γ the right branch or edge, and $I_\gamma = I'_\gamma \cup \{\gamma\} \cup I''_\gamma$ as well as $\mathcal{L}_{\{\gamma\}} = \{I'_\gamma, I''_\gamma\}$. If γ has a child $\alpha \in G$ ($\beta \in G$) with $\alpha < \gamma$ ($\gamma < \beta$), we set $I'_\gamma =: I_\alpha$ ($I''_\gamma =: I_\beta$) and proceed up the tree in a recursive way by identifying all remaining lines with the (possibly empty) segments $J \in \mathcal{S}$ in an analogous way, starting with the lines emanating from the child(ren) of γ.

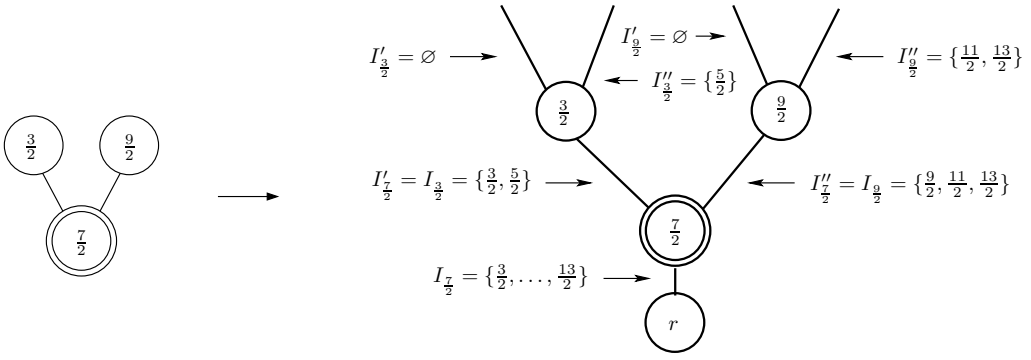


Figure 5.7. Left: Plane oriented tree $T = (\gamma, G, E)$ with vertex set $G = \left\{\frac{3}{2}, \frac{7}{2}, \frac{9}{2}\right\}$ and root $\gamma = \frac{7}{2}$. Right: The corresponding segmentation tree $T^L = (\gamma, G, E, L)$ for $L = \left\{\frac{3}{2}, \dots, \frac{13}{2}\right\}$.

Clearly, $\mathcal{S}$ is the set of all (possibly empty) segments that emerge when links are removed from L in the order prescribed by T. For every $\alpha \in G$, the segment I_α is the smallest segment in $\mathcal{S}$ that contains α, i.e. the particular segment that is cut next at link $\alpha \in G$. I_α will be understood as *internal segment*. The segments in $\mathcal{L}_G$, namely those that are associated with branches rather than edges, will be termed *external segments*. External segments $J \in \mathcal{L}_G$ can be either *full* (if $J \neq \varnothing$) or *empty* (if $J = \varnothing$). For $G = \varnothing$, the only segmentation tree is the empty planted tree (with no node except the phantom node r and the single line I_γ).

Due to the above description, we can rewrite $\mathcal{S}$ in various ways, namely,

$$\mathcal{S} = \bigcup_{R \in \mathcal{R}(T)} \mathcal{L}_R = \{I_\gamma\} \cup \{I'_\alpha, I''_\alpha : \alpha \in G\} = \{I_\alpha : \alpha \in G\} \cup \mathcal{L}_G \,.$$

In a similar manner, we can write $\mathcal{L}_R$, $R \in \mathcal{R}(T)$, as a collection of external segments and internal segments, namely

$$\mathcal{L}_R = \Big(\mathcal{L}_G \setminus \Big(\bigcup_{\alpha \in M(G \setminus R)} \mathcal{L}^{I_\alpha}_{G_\alpha(\varnothing)} \Big)\Big) \cup \{I_\alpha : \alpha \in M(G \setminus R)\}, \tag{5.21}$$

where $M(G \setminus R)$ is the set of vertices $G \setminus R$ that are minimal with respect to $\preccurlyeq$ (set $M(\varnothing) := \varnothing$); that is, M is the vertex counterpart of (5.11) and denoted by the same symbol by slight abuse of notation. Note that $G_\alpha(\varnothing) = G \cap I_\alpha$.

Remark 5.3. Our segmentation trees correspond to the *tree topologies* that occurred in [10]. In the genealogical context, a tree topology means an unweighted tree. We slightly adjusted the notation here for compatibility with the general usage in graph theory. ♢

Since the notions stump set, stump cut set, etc. from Section 5.2 depend on edges and vertices alone, they are not affected by additional lines that are attached to the trees. All notions from Section 5.2 therefore carry over to segmentation trees. Edge cutting, as the word suggests, still refers to cutting edges, not branches. For every $H \subseteq E$ and $\alpha \in G$, the rooted segmentation tree $T^L_\alpha(H)$ with vertex set $G_\alpha(H)$ contains information about the segments in $\mathcal{S}^{I_\alpha} := \bigcup_{R \in \mathcal{R}(T_\alpha(H))} \mathcal{L}^{I_\alpha}_R$, where $\mathcal{L}^{I_\alpha}_R$ is defined as in (5.4) with L replaced by I_α, $\alpha \in G$. The segmentation tree $T^L_\alpha(H)$ has phantom node r_α (where we set $r_\gamma := r$ for consistency). A *segmentation forest* $T^L - H$ of $T^L = (\gamma, G, E, L)$ is then the disjoint collection of segmentation trees $T^L_\alpha(H)$, where either $\alpha = \gamma$, or α is an upper end of an edge in H. Since edge deletion for segmentation trees is performed in the same way as for the trees in Section 5.2, the poset of segmentation forests $(\{T^L - H : H \subseteq E\}, \preceq_F)$ equipped with the partial order defined in Section 5.2.1, is once more isomorphic to the pruning poset $\mathcal{D}(T)$.

We now match realisations of the segmentation process with segmentation trees. Recall that $\mathcal{T}_\alpha$ is the waiting time until link α is removed and that $\mathcal{T}_K$, $K \subseteq L$, is the waiting time until the first link in K is removed.

Definition 5.3. For a given $t \geqslant 0$, we say that $(F_{t'})_{0 \leqslant t' \leqslant t}$ *matches* the segmentation tree $T^L = (\gamma, G, E, L)$ if $F_t = G$ and $\mathcal{T}_\alpha \leqslant \mathcal{T}_\beta$ precisely for those $\alpha, \beta \in G$ with $\alpha \preccurlyeq \beta$. In words, if the partial order of the waiting times in $(F_t)_{t \in \mathbb{T}}$ agrees with the partial order of the vertices of T^L up to time t.

Let now $\tau(G, L)$ be the set of all segmentation trees with vertex set G and underlying link set L (the cardinality of this set is the Catalan number $C_{|G|} = \frac{1}{|G|+1}\binom{2|G|}{|G|}$). We can then expand (5.6) into

$$\{F_t = G\} = \bigcup_{T^L \in \tau(G,L)} \mathcal{F}_t(T^L), \quad G \subseteq L$$

where

$$\mathcal{F}_t(T^L) = \{\{\max\{\mathcal{T}_\alpha : \alpha \in G\} \leqslant t < \mathcal{T}_{L\setminus G}\}, \{\mathcal{T}_\alpha = \mathcal{T}_{G_\alpha(\varnothing)} \text{ for all } \alpha \in G\}\} \tag{5.22}$$

is the event that $(F_t)_{0\leqslant t'\leqslant t}$ matches T^L. Indeed, the *inequalities* in (5.22) ensure that precisely the vertices in G have been removed before t. The *equalities* then enforce the partial order within the tree by requiring that α be the first link to be removed in the subtree with root α (which has vertex set $G_\alpha(\varnothing)$); it is sufficient to look at the links in $G_\alpha(\varnothing)$ since we know from the inequalities that those in $I_\alpha \setminus G_\alpha(\varnothing)$ are not cut until t anyway.

The task for the remainder of this section is to find an explicit expression for $\mathbf{P}[\mathcal{F}_t(T^L)]$. To this end, with the help of Möbius inversion on the pruning poset, we will write the maximum in (5.22) in terms of minima over certain subsets of G. This is motivated by the fact that the minimum of a collection of independent exponential (or geometric) random variables is independent of the order in which the events take place, whereas the respective maximum is not. But the details are quite different in continuous and discrete time. We therefore first set up a more general framework that covers both situations.

5.3.1 Möbius inversion for segmentation trees

Consider a segmentation tree $T^L = (\gamma, G, E, L)$. Let $\Gamma := G \cup \mathcal{S}$ and assign to every element $s \in \Gamma$ some event (in the sense of a finite set) $\mathcal{E}(s)$. We will throughout abbreviate $\mathcal{E}(\{\alpha\}) =: \mathcal{E}(\alpha)$. At this point, we neither give a meaning nor a law to the events, but will assume that the events are *nested* according to the set structure, i.e., that

$$\mathcal{E}(s_1) \subseteq \mathcal{E}(s_2) \text{ if and only if } s_1 \subseteq s_2 \subseteq \Gamma. \tag{5.23}$$

Note that in general $\mathcal{E}(s_1) \cup \mathcal{E}(s_2) \neq \mathcal{E}(s_1 \cup s_2)$, in particular $\mathcal{E}(I'_\alpha) \cup \mathcal{E}(\alpha) \cup \mathcal{E}(I''_\alpha) \subseteq \mathcal{E}(I_\alpha)$, but equality need not hold. Let Ξ be the set generated from $\{\mathcal{E}(s) : s \in \Gamma\}$ by arbitrary unions and set exclusions. We abbreviate the composite event $\bigcup_{J\in\mathcal{L}^{I_\alpha}_{G_\alpha(H)}} \mathcal{E}(J) =: \mathcal{E}(\mathcal{L}^{I_\alpha}_{G_\alpha(H)})$ for $\alpha \in G$, $H \subseteq E$. Furthermore, the event $\mathcal{E}(I_\alpha) \setminus \mathcal{E}(\mathcal{L}^{I_\alpha}_{G_\alpha(H)})$ will often be required. Let us state the following fact.

Fact 5.6. For events nested according to (5.23) we have $\mathcal{E}(\mathcal{L}_A) \subseteq \mathcal{E}(\mathcal{L}_G)$ for $G \subseteq A \subseteq L$. Moreover, for every $\alpha \in G$, $H \subseteq E$, the following properties hold:

(A) $\mathcal{E}(\mathcal{L}^{I_\alpha}_{G_\alpha(H)}) \subseteq \mathcal{E}(\mathcal{L}^{I_\alpha}_{G_\alpha(H\cup K)}) \subseteq \mathcal{E}(\mathcal{L}^{I_\alpha}_{\varnothing}) = \mathcal{E}(I_\alpha)$ for $K \subseteq E$.

(B) $\mathcal{E}(\beta) \subseteq \mathcal{E}(I_\alpha) \setminus \mathcal{E}(\mathcal{L}^{I_\alpha}_{G_\alpha(H)})$ for all $\beta \in G_\alpha(H)$.

(C) $\mathcal{E}(\mathcal{L}^{I_\beta}_{G_\beta(H)}) \subseteq \mathcal{E}(\mathcal{L}^{I_\alpha}_{G_\alpha(H)})$ for all $\beta \in G_\alpha(H)$.

Proof. Let $G \subseteq A \subseteq L$. By definition of $\mathcal{L}_A$ and $\mathcal{L}_G$, for any $J \in \mathcal{L}_A$ there is an $I \in \mathcal{L}_G$ such that $J \subseteq I$ and thus $\mathcal{E}(J) \subseteq \mathcal{E}(I)$ by (5.23). (A) follows from the latter statement since $\varnothing \subseteq G_\alpha(H \cup K) \subseteq G_\alpha(H)$ for any $\alpha \in G$, $H, K \subseteq E$, and because $\mathcal{L}_\varnothing^{I_\alpha} = \{I_\alpha\}$. (B): Let $\beta \in G_\alpha(H)$ for some $\alpha \in G$. Since $\beta \in I_\beta \subseteq I_\alpha$, we know $\mathcal{E}(\beta) \subseteq \mathcal{E}(I_\beta) \subseteq \mathcal{E}(I_\alpha)$ by (5.23). On the other hand, $\mathcal{L}_{G_\alpha(H)}^{I_\alpha} \setminus \varnothing$ is a partition of $I_\alpha \setminus G_\alpha(H)$, so $\beta \notin J$ for any $J \in \mathcal{L}_{G_\alpha(H)}^{I_\alpha}$ and thus $\mathcal{E}(\beta) \nsubseteq \mathcal{E}(\mathcal{L}_{G_\alpha(H)}^{I_\alpha})$ by (5.23). (C) follows from (5.23) and the fact that $\mathcal{L}_{G_\beta(H)}^{I_\beta} \subseteq \mathcal{L}_{G_\alpha(H)}^{I_\alpha}$ for all $\beta \in G_\alpha(H)$. □

Let now $\mathcal{T} : \Xi \to \mathbb{R}_{\geqslant 0}$ be a function that assigns a scalar to each event in Ξ. Later, $\mathcal{T}$ will turn into the waiting time for the event, but here we are not tied to an underlying process. Let us write $\mathcal{T}_\mathcal{G} := \mathcal{T}(\mathcal{G})$ and assume that

$$\mathcal{T}_\mathcal{G} \leqslant \mathcal{T}_\mathcal{H} \text{ if and only if } \mathcal{G} \supseteq \mathcal{H}, \quad \mathcal{H}, \mathcal{G} \in \Xi. \tag{5.24}$$

Our object of interest in this section is the event $\{\mathrm{Max}_{t,\mathcal{E}}(G), m_\mathcal{E}(\varnothing)\}$, where

$$\mathrm{Max}_{t,\mathcal{E}}(G) := \Big\{ \max\{\mathcal{T}_{\mathcal{E}(\alpha)} : \alpha \in G\} \leqslant t < \mathcal{T}_{\mathcal{E}(\mathcal{L}_G)} \Big\}, \quad G \subseteq L, \quad t \geqslant 0, \tag{5.25}$$

and

$$m_\mathcal{E}(H) := \bigcap_{\alpha \in G} \Big\{ \mathcal{T}_{\mathcal{E}(\alpha)} = \mathcal{T}_{\mathcal{E}(I_\alpha) \setminus \mathcal{E}(\mathcal{L}_{G_\alpha(H)}^{I_\alpha})} \Big\}, \quad H \subseteq E. \tag{5.26}$$

We will see later that $\{\mathrm{Max}_{t,\mathcal{E}}(G), m_\mathcal{E}(\varnothing)\}$ generalises the tree event in (5.22). Let us only mention here that (5.26) may be understood as an order relation within each of the connected components of $T^L - H$. Our aim is to express $\{\mathrm{Max}_{t,\mathcal{E}}(G), m_\mathcal{E}(\varnothing)\}$ in terms of a collection of certain minima combined with order relations, via an inclusion-exclusion principle. The order relations are those just defined, and the minima are analogous to the maxima, namely

$$\mathrm{Min}_{t,\mathcal{E}}(G) := \{ \min\{\mathcal{T}_{\mathcal{E}(\alpha)} : \alpha \in G\} \leqslant t < \mathcal{T}_{\mathcal{E}(\mathcal{L}_G)} \}, \quad G \subseteq L, \quad t \geqslant 0. \tag{5.27}$$

We will proceed in the opposite direction and start with a decomposition of the joint event of the form $\{\mathrm{Min}_{t,\mathcal{E}}(G),\ m_\mathcal{E}(H)\}$ into a collection of maxima and then apply Möbius inversion on $\mathcal{D}(T)$ from (5.20). Anticipating that the stump set will play a special role in our final tree probabilities, we formulate the following lemma.

Lemma 5.1. *Let $T^L = (\gamma, G, E, L)$ be a segmentation tree and $K \subseteq E$. If* (5.23) *and* (5.24) *are satisfied, then*

$$\mathbf{P}\big[\mathrm{Min}_{t,\mathcal{E}}(G_\gamma(K)),\ m_\mathcal{E}(K)\big] = \sum_{C \in \mathcal{C}(T_\gamma(K))} \mathbf{P}\big[\mathrm{Max}_{t,\mathcal{E}}(G_\gamma(K \cup C)),\ m_\mathcal{E}(K \cup C)\big], \tag{5.28}$$

where $\mathbf{P}$ denotes a probability measure on Ξ and $\mathcal{C}(T_\gamma(K))$ is the set of all stump cut sets of $T_\gamma(K)$.

Proof. We will decompose the probability for the joint event $\{\mathrm{Min}_{t,\mathcal{E}}(G_\gamma(K)),\, m_{\mathcal{E}}(K)\}$ part by part. We first express the minimum in $\mathrm{Min}_{t,\mathcal{E}}(G_\gamma(K))$ in terms maxima using the well-known disjoint decomposition, which here reads

$$\begin{aligned}&\mathbf{P}\big[\min\{\mathcal{T}_{\mathcal{E}(\alpha)} : \alpha \in G_\gamma(K)\} \leqslant t\big]\\ &\quad= \sum_{\varnothing \neq A \subseteq G_\gamma(K)} \mathbf{P}\big[\max\{\mathcal{T}_{\mathcal{E}(\alpha)} : \alpha \in A\} \leqslant t < \min\{\mathcal{T}_{\mathcal{E}(\beta)} : \beta \in G_\gamma(K) \setminus A\}\big].\end{aligned} \tag{5.29}$$

We now add the ordering relation $m_{\mathcal{E}}(K)$ on both sides of (5.29). Since

$$m_{\mathcal{E}}(K) \text{ implies } \mathcal{T}_{\mathcal{E}(\alpha)} \leqslant \mathcal{T}_{\mathcal{E}(\beta)} \text{ for all } \alpha \in G \text{ and } \beta \in G_\alpha(K) \tag{5.30}$$

by Fact 5.6 (B), we have

$$\mathbf{P}\big[\max\{\mathcal{T}_{\mathcal{E}(\alpha)} : \alpha \in A\} \leqslant t < \min\{\mathcal{T}_{\mathcal{E}(\beta)} : \beta \in G_\gamma(K) \setminus A\},\, m_{\mathcal{E}}(K)\big] = 0$$

for every subset $A \subseteq G$ that does not contain the root, or is not contiguous with respect to the partial order on $T^L_\gamma(K)$, that is, if A is not a stump set of $T^L_\gamma(K)$. Using (5.30) once more, we conclude that

$$\min\{\mathcal{T}_{\mathcal{E}(\beta)} : \beta \in G_\gamma(K) \setminus R\} \cap m_{\mathcal{E}}(K) = \min\{\mathcal{T}_{\mathcal{E}(\beta)} : \beta \in M(G_\gamma(K) \setminus R)\} \cap m_{\mathcal{E}}(K),$$

where $M(G_\gamma(K) \setminus R)$ is the set of vertices in $G_\gamma(K) \setminus R$ that are minimal with respect to $\preccurlyeq$. We may thus write

$$\begin{aligned}\mathbf{P}\big[\mathrm{Min}_{t,\mathcal{E}}(G_\gamma(K)),\, m_{\mathcal{E}}(K)\big] &= \sum_{R \in \mathcal{R}(T^L_\gamma(K))} \mathbf{P}\Big[\max\{\mathcal{T}_{\mathcal{E}(\alpha)} : \alpha \in R\} \leqslant t < \mathcal{T}_{\mathcal{E}(\mathcal{L}_{G_\gamma(K)})},\, m_{\mathcal{E}}(K),\\ &\qquad\qquad t < \min\{\mathcal{T}_{\mathcal{E}(\beta)} : \beta \in M(G_\gamma(K) \setminus R)\}\}\Big]\\ &= \sum_{R \in \mathcal{R}(T^L_\gamma(K))} \mathbf{P}\Big[\max\{\mathcal{T}_{\mathcal{E}(\alpha)} : \alpha \in R\} \leqslant t < \mathcal{T}_{\mathcal{E}(\mathcal{L}_{G_\gamma(K)})},\, m_{\mathcal{E}}(K),\\ &\qquad\qquad t < \min\Big\{\mathcal{T}_{\mathcal{E}(I_\beta) \setminus \mathcal{E}(\mathcal{L}^{I_\beta}_{G_\beta(K)})} : \beta \in M(G_\gamma(K) \setminus R)\Big\}\Big]\\ &= \sum_{R \in \mathcal{R}(T^L_\gamma(K))} \mathbf{P}\big[\mathrm{Max}_{t,\mathcal{E}}(G_\gamma(K)),\, m_{\mathcal{E}}(K)\big].\end{aligned}$$

The second equality is due to $m_{\mathcal{E}}(K)$, see (5.26). In the third step, we used that

$$\mathcal{E}(\mathcal{L}_{G_\gamma(K)}) \bigcup_{\beta \in M(G_\gamma(K) \setminus R)} \mathcal{E}(I_\beta) \setminus \mathcal{E}(\mathcal{L}^{I_\beta}_{G_\beta(K)}) = \mathcal{E}(\mathcal{L}_R)\,,$$

which follows by (5.21) applied to the stump tree $T^L_\gamma(K)$ with the help of Fact 5.2 (A). Altogether this gives

$$\mathbf{P}\big[\mathrm{Min}_{t,\mathcal{E}}(G_\gamma(K)),\, m_{\mathcal{E}}(K)\big] = \sum_{C \in \mathcal{C}(T_\gamma(K))} \mathbf{P}\big[\mathrm{Max}_{t,\mathcal{E}}(G_\gamma(K \cup C)),\, m_{\mathcal{E}}(K)\big] \tag{5.31}$$

due (5.12) and Fact 5.2 (A). Let us finally consider the ordering relation $m_{\mathcal{E}}(K)$ in the joint event on the right-hand side of (5.31). Consider first an $\alpha \notin G_\gamma(K \cup C)$, in which

case we obtain $\mathcal{E}(I_\alpha) \setminus \mathcal{E}(\mathcal{L}^{I_\alpha}_{G_\alpha(K)}) = \mathcal{E}(I_\alpha) \setminus \mathcal{E}(\mathcal{L}^{I_\alpha}_{G_\alpha(K\cup C)})$ by Fact 5.2 (B). Let now $\alpha \in G_\gamma(K\cup C)$. Given $\mathrm{Max}_{t,\mathcal{E}}(G_\gamma(K\cup C))$, we then have $\mathcal{T}_{\mathcal{E}(\alpha)} < \mathcal{T}_{\mathcal{E}(\mathcal{L}_{G_\gamma(K\cup C)})}$. Since furthermore $\mathcal{T}_{\mathcal{E}(\mathcal{L}_{G_\gamma(K\cup C)})} \leqslant \mathcal{T}_{\mathcal{E}(\mathcal{L}^{I_\alpha}_{G_\alpha(K\cup C)})}$ by Fact 5.6 (C), we can conclude

$$\Big\{\mathcal{T}_{\mathcal{E}(\alpha)} = \mathcal{T}_{\mathcal{E}(I_\alpha)\setminus\mathcal{E}(\mathcal{L}^{I_\alpha}_{G_\alpha(K)})}\Big\} = \Big\{\mathcal{T}_{\mathcal{E}(\alpha)} = \mathcal{T}_{\left(\mathcal{E}(I_\alpha)\setminus\mathcal{E}(\mathcal{L}^{I_\alpha}_{G_\alpha(K)})\right)\setminus\mathcal{E}(\mathcal{L}^{I_\alpha}_{G_\alpha(K\cup C)})}\Big\}.$$

Since furthermore

$$(\mathcal{E}(I_\alpha) \setminus \mathcal{E}(\mathcal{L}^{I_\alpha}_{G_\alpha(K)})) \setminus \mathcal{E}(\mathcal{L}^{I_\alpha}_{G_\alpha(K\cup C)}) = \mathcal{E}(I_\alpha) \setminus \mathcal{E}(\mathcal{L}^{I_\alpha}_{G_\alpha(K\cup C)})$$

by Fact 5.6 (A), we can rewrite the joint event as

$$\{\mathrm{Max}_{t,\mathcal{E}}(G_\gamma(K\cup C)),\, m_{\mathcal{E}}(K)\} = \{\mathrm{Max}_{t,\mathcal{E}}(G_\gamma(K\cup C)),\, m_{\mathcal{E}}(K\cup C)\}.$$

Together with (5.31) this completes the proof. □

Remark 5.4. For $K = \varnothing$, Equation (5.28) leads to the recursion

$$\mathbf{P}\big[\mathrm{Max}_{t,\mathcal{E}}(G),\, m_{\mathcal{E}}(\varnothing)\big] = \mathbf{P}\big[\mathrm{Min}_{t,\mathcal{E}}(G),\, m_{\mathcal{E}}(\varnothing)\big] - \sum_{R\in\mathcal{R}(T)\setminus\{G\}} \mathbf{P}\big[\mathrm{Max}_{t,\mathcal{E}}(R),\, m_{\mathcal{E}}(\partial(R))\big].$$

The ordering relation within a given segmentation tree may therefore be separated into an ordering within the stump tree and an ordering within the remaining part. Iteratively, this leads to decompositions of the ordering relation that correspond to rooted forests obtained from the original segmentation tree via edge deletion in the stump tree. The procedure justifies the construction of the pruning poset in Section 5.2.1. ♢

Theorem 5.3. *Under the conditions of Lemma 5.1, the following holds for every $K \subseteq E$:*

$$\mathbf{P}\big[\mathrm{Max}_{t,\mathcal{E}}(G_\gamma(K)),\, m_{\mathcal{E}}(K)\big] = \sum_{H\subseteq E_\gamma(K)} (-1)^{|H|}\, \mathbf{P}\big[\mathrm{Min}_{t,\mathcal{E}}(G_\gamma(H\cup K)),\, m_{\mathcal{E}}(H\cup K)\big].$$

Proof. Recall the Möbius function μ for the pruning poset $\mathcal{D}(T)$ in (5.16) and rewrite it as $\mu(H,K)\,(-1)^{|H|-|K|} = \mathbb{1}_{\{H\setminus K\,\in\,\mathcal{C}(T_\gamma(K))\}}$ for $H, K \subseteq E, H \preceq_D K$. This allows to reformulate (5.28) from Lemma 5.1 as

$$\begin{aligned}(-1)^{|K|}\, &\mathbf{P}\big[\mathrm{Min}_{t,\mathcal{E}}(G_\gamma(K)),\, m_{\mathcal{E}}(K)\big] \\ &= (-1)^{|K|} \sum_{H\subseteq E} \mathbb{1}_{\{H\setminus K\,\in\,\mathcal{C}(T_\gamma(K))\}}\, \mathbf{P}\big[\mathrm{Max}_{t,\mathcal{E}}(G_\gamma(H)),\, m_{\mathcal{E}}(H)\big] \\ &= \sum_{H\preceq_D K} \mu(H,K)\, (-1)^{|H|}\, \mathbf{P}\big[\mathrm{Max}_{t,\mathcal{E}}(G_\gamma(H)),\, m_{\mathcal{E}}(H)\big],\end{aligned} \tag{5.32}$$

where the last equality is due to isomorphism on $\mathcal{D}(T_\gamma(H))$ in (5.13). Möbius inversion on $\mathcal{D}(T)$ (cf. (5.20)) then yields the inverse of (5.32):

$$\begin{aligned}(-1)^{|K|}\, \mathbf{P}\big[\mathrm{Max}_{t,\mathcal{E}}(G_\gamma(K)),\, m_{\mathcal{E}}(K)\big] &= \sum_{H\preceq_D K} (-1)^{|H|}\, \mathbf{P}\big[\mathrm{Min}_{t,\mathcal{E}}(G_\gamma(H)),\, m_{\mathcal{E}}(H)\big] \\ &= \sum_{H\subseteq E_\gamma(K)} (-1)^{|H\cup K|}\, \mathbf{P}\big[\mathrm{Min}_{t,\mathcal{E}}(G_\gamma(H\cup K)),\, m_{\mathcal{E}}(H\cup K)\big],\end{aligned}$$

where the last equality is once more isomorphism on $\mathcal{D}(T_\gamma(H))$. □

We can now use Theorem 5.3 to evaluate the tree probabilities in (5.22). To this end, we define events for the segmentation process as $\mathcal{E}_F(s) := \{s\}$ for all $s \in \Gamma$, so that $\mathcal{T}_{\mathcal{E}_F(s)} = \mathcal{T}_s$ is the waiting time at which the first link in s is removed under $(F_t)_{t\in\mathbb{T}}$. Since $L\setminus G = \cup_{J\in\mathcal{L}_G} J$, we then have

$$\mathcal{T}_{L\setminus G} = \min\{\mathcal{T}_J : J \in \mathcal{L}_G\} = \min\{\mathcal{T}_{\mathcal{E}_F(J)} : J \in \mathcal{L}_G\} = \mathcal{T}_{\mathcal{E}_F(\mathcal{L}_G)}. \tag{5.33}$$

Likewise, since

$$G_\alpha(H) = I_\alpha \setminus \{J : J \in \mathcal{L}^{I_\alpha}_{G_\alpha(H)}\} = \mathcal{E}_F(I_\alpha) \setminus \mathcal{E}_F(\mathcal{L}^{I_\alpha}_{G_\alpha(H)}), \tag{5.34}$$

one has

$$\mathcal{T}_{G_\alpha(H)} = \mathcal{T}_{\mathcal{E}_F(I_\alpha)\setminus\mathcal{E}_F(\mathcal{L}^{I_\alpha}_{G_\alpha(H)})}. \tag{5.35}$$

These seemingly more complicated expressions allow us to rewrite $\mathcal{F}_t(T^L)$ from (5.22) as the generalised tree event

$$\mathcal{F}_t(T^L) = \{\mathrm{Max}_{t,\mathcal{E}_F}(G), m_{\mathcal{E}_F}(\varnothing)\} \tag{5.36}$$

with $\mathrm{Max}_{t,\mathcal{E}_F}(G)$ and $m_{\mathcal{E}_F}(\varnothing)$ as defined in (5.25) and (5.26), and $\mathcal{E}$ replaced by $\mathcal{E}_F$.

Corollary 5.1. *Let $T^L = (\gamma, G, E, L)$ and $t \geqslant 0$ be given. The probability that $(F_t)_{0\leqslant t'\leqslant t}$ matches T^L is then given by*

$$\begin{aligned}\mathbf{P}[\mathcal{F}_t(T^L)] &= \sum_{H\subseteq E} (-1)^{|H|}\, \mathbf{P}[\mathrm{Min}_{t,\mathcal{E}_F}(G_\gamma(H)),\, m_{\mathcal{E}_F}(H)] \\ &= \sum_{H\subseteq E} (-1)^{|H|}\, \mathbf{P}[\mathcal{T}_{G_\gamma(H)} \leqslant t < \mathcal{T}_{L\setminus G_\gamma(H)},\ \mathcal{T}_\alpha = \mathcal{T}_{G_\alpha(H)}\ \forall\ \alpha \in G].\end{aligned} \tag{5.37}$$

The probability of a segmentation tree T^L can thus be expressed as an alternating sum over all probabilities corresponding to segmentation forests that can be obtained from T^L by edge deletion. For every given segmentation forest $T^L - H$, the ordering relation may be rewritten as

$$m_{\mathcal{E}_F}(H) = \bigcap_{T^L_\alpha(H)\in T^L - H} \quad \bigcap_{\nu\in G_\alpha(H)} \{\mathcal{T}_\nu = \mathcal{T}_{G_\nu(H)}\},$$

which shows that the ordering is now prescribed within each component of $T^L - H$, in contrast to (5.22), which prescribes the ordering within the entire tree. The joint event $\{\mathrm{Min}_{t,\mathcal{E}_F}(G_\gamma(H)),\, m_{\mathcal{E}_F}(H)\}$ thus means that at least one link in the stump tree has been removed until time t, all the links in $L \setminus G_\gamma(H)$ are still intact and that the events corresponding to the vertices in G happen in the prescribed order within each component. This interpretation holds for discrete and continuous time alike.

Proof. Choosing $\mathcal{E}_F(s) = \{s\}$ for all $s \in \Gamma$ clearly satisfies the nesting condition (5.23). Furthermore, choosing $\mathcal{T}$ as the waiting-time for the events $\mathcal{E}_F$ guarantees (5.24). We may thus use Theorem 5.3 and apply it to (5.36), that is, for $K = \varnothing$. This yields

$$\mathbf{P}[\mathcal{F}_t(T^L)] = \mathbf{P}[\mathrm{Max}_{t,\mathcal{E}_F}(G),\, m_{\mathcal{E}_F}(\varnothing)] = \sum_{H\subseteq E} (-1)^{|H|}\, \mathbf{P}[\mathrm{Min}_{t,\mathcal{E}_F}(G_\gamma(H)),\, m_{\mathcal{E}_F}(H)]$$

with $\mathrm{Min}_{t,\mathcal{E}_F}$ from (5.27). Employing (5.33) and (5.35) once more, this time in the reverse direction, completes the proof. □

5.3.2 Tree probabilities in continuous time

An explicit expression for the tree probabilities in continuous time now transpires without much effort due to the independence of the waiting times. Let us denote by $\breve{\mathcal{F}}_t(T^L)$ the continuous-time version of (5.22).

Proposition 5.1. *For a given segmentation tree $T^L = (\gamma, G, E, L)$ and a fixed $t \geqslant 0$, one has $\mathbf{P}[\breve{\mathcal{F}}_t(T^L)] = \exp(-\sum_{\alpha\in L} \varrho_\alpha\, t)$ for $G = \varnothing$ and, for every $\varnothing \neq G \subseteq L$,*

$$\mathbf{P}[\breve{\mathcal{F}}_t(T^L)] = \sum_{H\subseteq E} (-1)^{|H|}\Big(1 - \exp\Big(-\sum_{\alpha\in G_\gamma(H)} \varrho_\alpha\, t\Big)\Big) \exp\Big(-\sum_{\beta\in L\setminus G_\gamma(H)} \varrho_\beta\, t\Big) \prod_{\alpha\in G} \frac{\varrho_\alpha}{\sum_{\nu\in G_\alpha(H)} \varrho_\nu}.$$

Proof. For a fixed $H \subseteq E$, consider the event $\{\mathrm{Min}_{t,\mathcal{E}_F}(G_\gamma(H)),\, m_{\mathcal{E}_F}(H)\}$ on the right-hand side of (5.37), which is $\Big\{\widetilde{\mathrm{Min}}_{t,\mathcal{E}_F}(G_\gamma(H)),\, \widetilde{m}_{\mathcal{E}_F}(H)\Big\}$ in continuous time. Recall that the waiting times for the links are independent and that the minimum of a collection of independent exponential waiting times is independent of the order in which the events appear. We obtain

$$\mathbf{P}\Big[\widetilde{\mathrm{Min}}_{t,\mathcal{E}_F}(G_\gamma(H)),\, \widetilde{m}_{\mathcal{E}_F}(H)\Big] = \mathbf{P}\Big[\min\{\breve{\mathcal{T}}_\alpha : \alpha \in G_\gamma(H)\} \leqslant t\Big]\ \mathbf{P}\Big[\breve{\mathcal{T}}_\beta > t,\ \beta \in L\setminus G_\gamma(H)\Big] \times \prod_{\alpha\in G} \mathbf{P}\Big[\breve{\mathcal{T}}_\alpha \leqslant \breve{\mathcal{T}}_\beta,\ \beta \in G_\alpha(H)\Big],$$

which can be evaluated in an elementary manner, with the help of the independent exponential laws of the $\breve{\mathcal{T}}$'s. Together with Corollary 5.1, this completes the proof. □

As it must be, we rediscover the explicit expression for the probability distribution of $\breve{F}_t$ in (5.7) by summing over all probabilities of matching events of segmentation trees whose vertex set is G.

Fact 5.7. For $G \subseteq L$ and fixed $t \geqslant 0$:

$$\sum_{T^L\in\tau(G,L)} \mathbf{P}[\breve{\mathcal{F}}_t(T^L)] = \prod_{\alpha\in G} (1 - \exp(-\varrho_\alpha\, t)) \prod_{\beta\in L\setminus G} \exp(-\varrho_\beta\, t) = \mathbf{P}[\breve{F}_t = G].$$

Proof. For every $\varnothing \neq G \subseteq L$, we first show that

$$\sum_{T^L\in\tau(G,L)} \mathbf{P}[\breve{\mathcal{F}}_t(T^L)] = \sum_{\varnothing\neq H\subseteq G} (-1)^{|G|-|H|}\Big(1 - \exp\Big(-\sum_{\nu\in H} \varrho_\nu\, t\Big)\Big) \exp\Big(-\sum_{\nu\in L\setminus H} \varrho_\nu\, t\Big) \quad (5.38)$$

We use induction and employ the recursion formula for binary trees, which in this case gives

$$\mathbf{P}[\breve{\mathcal{F}}_t(T^L)] = \int_{y=0}^{t} \varrho_\gamma \exp\Big(-\sum_{\alpha\in L} \varrho_\alpha\, y\Big)\, \mathbf{P}[\breve{\mathcal{F}}_{t-y}(T^{I'_\gamma})]\, \mathbf{P}[\breve{\mathcal{F}}_{t-y}(T^{I''_\gamma})]\, \mathrm{d}y. \quad (5.39)$$

Equation (5.39) is the continuous-time analogue of Equation (38) in [10], where $T^{I'_\gamma}$ is the left subtree of the segmentation tree T^L with vertex set $G_{<\gamma} := \{\alpha \in G : \alpha < \gamma\}$ and $T^{I''_\gamma}$

the right subtree with vertex set $G_{>\gamma} := \{\alpha \in G : \alpha > \gamma\}$. Assume now that the equality in (5.38) holds for any subtree of T^L. Due to the independence of the subtrees, we find

$$\begin{aligned}\sum_{T^L \in \tau(G,L)} \mathbf{P}[\breve{\mathcal{F}}_t(T^L)] &= \sum_{\alpha \in G} \sum_{\substack{T^L \in \tau(G,L),\\ T^L \text{ has root } \alpha}} \mathbf{P}[\breve{\mathcal{F}}_t(T^L)] \\ &= \sum_{\alpha \in G} \varrho_\alpha \int_{y=0}^{t} \exp\Big(-\sum_{\beta \in L} \varrho_\beta\, y\Big) \Bigg(\sum_{T^{I'_\gamma} \in \tau(G_{<\gamma}, I'_\gamma)} \mathbf{P}[\breve{\mathcal{F}}_t(T^{I'_\gamma})]\Bigg) \\ &\qquad \times \Bigg(\sum_{T^{I''_\gamma} \in \tau(G_{>\gamma}, I''_\gamma)} \mathbf{P}[\breve{\mathcal{F}}_t(T^{I''_\gamma})]\Bigg) \mathrm{d}y.\end{aligned}$$

Inserting the induction hypothesis and calculating out the integral shows (5.38). The claim then follows from a straightforward calculation using

$$\sum_{H \subseteq G} (-1)^{|H|} \exp\Big(-\sum_{\nu \in H} \varrho_\nu\, t\Big) = \prod_{\beta \in G} (1 - \exp(-\varrho_\beta\, t)) \quad G \subseteq L. \qquad \square$$

In continuous time, simultaneous events are automatically excluded. The partial order that encodes the time series of removed links until time t is therefore in fact a total order. One may thus obtain the probability for a set of realisations of $(\breve{F}_t)_{t \in \mathbb{T}}$ that agree on the order of events more easily by calculating out the convolution of the respective exponential distributions, i.e. one may avoid all the extensive framework about segmentation trees, the pruning poset and Möbius inversion. The tree probabilities in Proposition 5.1 will, however, facilitate the comparison to the discrete-time case.

Discrete time In discrete time, the links are dependent as long as they belong to the same segment. The probability that nothing happens in a given time step is

$$\mathbf{P}\Big[\widehat{F}_{t+1} = G \mid \widehat{F}_t = G\Big] = \prod_{J \in \mathcal{L}_G} (1 - r_J) =: \lambda_G, \quad r_J := \sum_{\alpha \in J} r_\alpha. \tag{5.40}$$

Due to the triangular structure, the λ_G's are the eigenvalues of the Markov transition matrix of $\widehat{F}$. The λ_G's have previously been identified by Bennett [14], Ljubič [89, Sect. 6.4] and Dawson [27] in the context of the deterministic recombination equation.

The law of links to be added to $\widehat{F}_t$ changes over time and (5.37) can not be evaluated in a straightforward manner. Suppose that, up to a particular time, the link $\gamma \notin \{\frac{3}{2}, \frac{2n-1}{2}\}$ is removed. Then L splits into the two nonempty segments $I'_\gamma = \{\beta \in L : \beta < \gamma\}$ and $I''_\gamma = \{\beta \in L : \beta > \gamma\}$. After removal of γ, the joint probability that a link in I'_γ or I''_γ is removed is $1 - \lambda^L_\gamma = r_{I'_\gamma} + r_{I''_\gamma} - r_{I'_\gamma} \cdot r_{I''_\gamma}$, whereas before removal of γ it is $r_{I'_\gamma} + r_{I''_\gamma} = 1 - \lambda^L_\gamma + r_{I'_\gamma} \cdot r_{I''_\gamma}$. We may thus think of $1 - \lambda^L_\gamma$ as the probability for a removal in $L \setminus \{\gamma\}$ when I'_γ and I''_γ are independent and of $r_{I'_\gamma} \cdot r_{I''_\gamma}$ as the additional probability for the case that the segments are still dependent. We generalise the idea of a decomposition into dependent and independent parts in the next section.

5.3.3 The auxiliary process

We now construct an auxiliary process which is state independent and which jointly represents all transitions of interest of the discrete-time segmentation process for a given segmentation tree. The method is reminiscent of that used by Clifford and Sudbury [26]. We then use the auxiliary process to construct realisations of $(\widehat{F}_t)_{t\in\mathbb{N}_0}$ that are compatible with a given segmentation tree up to time t and express matching events of the segmentation process in terms of matching events of the auxiliary process.

Construction of the auxiliary process

Fix a segmentation tree $T^L = (\gamma, G, E, L)$. We aim at a construction of a sequence of i.i.d. random variables $(X_t)_{t\in\mathbb{N}_0}$ where, for all $t \in \mathbb{N}_0$, X_t will be a family $X_t = (X_t^J)_{J\in\mathcal{S}}$, and the X_t^J's will have a specific dependence for the J's. We construct this collection for every $t \in \mathbb{N}_0$ inductively, starting with the (full or empty) external segments of the tree and proceeding in a top-down manner.

For the start, let $t > 0$ be fixed and define X_t^J for each external segment $J \in \mathcal{L}_G$ independently for each J on $\Omega^J := \{\omega_\varnothing^J, \omega_J^J\}$ with

$$X_t^J = \begin{cases} \omega_\varnothing^J, & \text{with probability } 1 - r_J, \\ \omega_J^J, & \text{with probability } r_J. \end{cases} \tag{5.41}$$

If $J = \varnothing$, then obviously $X_t^J = \omega_\varnothing^\varnothing$ with probability 1 (for consistency, set $r_\varnothing := 0$). Now consider the internal segments I_α, $\alpha \in G$. As already mentioned, every segment may be pieced together from its two descendant segments $I'_\alpha = \{\beta \in I_\alpha : \beta < \alpha\}$ and $I''_\alpha = \{\beta \in I_\alpha : \beta > \alpha\}$. Namely, $I_\alpha = I'_\alpha \cup \{\alpha\} \cup I''_\alpha$; I'_α and I''_α may be internal segments or (empty or full) external segments. We now proceed down the tree inductively by taking, in every step, one α, for which $X_t^{I'_\alpha}$ and $X_t^{I''_\alpha}$ have already been defined (as independent processes on $\Omega^{I'_\alpha}$ and $\Omega^{I''_\alpha}$). $X_t^{I_\alpha}$ will live on the state space $\Omega^{I_\alpha} := \left\{\omega_\varnothing^{I_\alpha}, \omega_\alpha^{I_\alpha}, \omega_{\mathrm{ind}}^{I_\alpha}, \omega_{\mathrm{dep}}^{I_\alpha}\right\}$, and we define the composite event $\omega_{I_\alpha}^{I_\alpha} := \Omega^{I_\alpha} \setminus \omega_\varnothing^{I_\alpha}$, which blends in with the notation in (5.41). By slight abuse of notation, we will sometimes write $X_t^J = \omega_J^J$ for a $J \in \mathcal{S}$ even though the correct statement would be $X_t^J \in \omega_J^J$ if J is an internal segment, and $X_t^J = \omega_J^J$ if J is an external segment. We now construct $X_t^{I_\alpha}$ by specifying

$$X_t^{I_\alpha} = \omega_{\mathrm{ind}}^{I_\alpha} \quad \text{if } X_t^{I'_\alpha} = \omega_{I'_\alpha}^{I'_\alpha} \quad \text{or} \quad X_t^{I''_\alpha} = \omega_{I''_\alpha}^{I''_\alpha}$$

(this is the case with probability $1 - \lambda_\alpha^{I_\alpha}$) and, if $X_t^{I'_\alpha} = \omega_\varnothing^{I'_\alpha}$ and $X_t^{I''_\alpha} = \omega_\varnothing^{I''_\alpha}$ (which is the case with probability $\lambda_\alpha^{I_\alpha}$), we set

$$X_t^{I_\alpha} = \begin{cases} \omega_\varnothing^{I_\alpha}, & \text{with probability } {}^{(1-r_{I_\alpha})}/_{\lambda_\alpha^{I_\alpha}}, \\ \omega_\alpha^{I_\alpha}, & \text{with probability } {}^{r_\alpha}/_{\lambda_\alpha^{I_\alpha}}, \\ \omega_{\mathrm{dep}}^{I_\alpha}, & \text{with probability } {}^{r_{I'_\alpha} \cdot r_{I''_\alpha}}/_{\lambda_\alpha^{I_\alpha}} \end{cases}$$

independently of what has been decided for the previous segments. Here, $\lambda_\alpha^{I_\alpha}$ is defined as in Equation (5.40), but with respect to the link set I_α of the tree with root α. This means that $\Omega^{I'_\alpha} \times \Omega^{I''_\alpha} \setminus (\omega_\varnothing^{I'_\alpha}, \omega_\varnothing^{I''_\alpha})$ is identified with the event $\omega_{\mathrm{ind}}^{I_\alpha}$, whereas the remaining element $(\omega_\varnothing^{I'_\alpha}, \omega_\varnothing^{I''_\alpha})$ is 'split up' into the elements of $\Omega^{I_\alpha} \setminus \omega_{\mathrm{ind}}^{I_\alpha}$.

We see that under this construction, $X_t^{I_\alpha}$ has the law

$$X_t^{I_\alpha} = \begin{cases} \omega_\varnothing^{I_\alpha}, & \text{with probability } 1 - r_{I_\alpha}, \\ \omega_\alpha^{I_\alpha}, & \text{with probability } r_\alpha, \\ \omega_{\mathrm{dep}}^{I_\alpha}, & \text{with probability } r_{I'_\alpha} \cdot r_{I''_\alpha}, \\ \omega_{\mathrm{ind}}^{I_\alpha} & \text{with probability } 1 - \lambda_\alpha^{I_\alpha}. \end{cases} \tag{5.42}$$

The construction is completed when $X_t^L = X_t^{I_\gamma}$ has been reached. Altogether, we then have the family $X_t = (X_t^J)_{J \in \mathcal{S}}$ with state space $\Omega := \bigtimes_{J \in \mathcal{S}} \Omega^J$. The sequence of random variables $X = (X_t)_{t \in \mathbb{N}_0}$ is defined to be i.i.d. in t. The X_t^J are independent for all disjoint segments; in particular, for every stump set $R \in \mathcal{R}(T^L)$, the family $(X_t^J)_{J \in \mathcal{L}_R}$ with state space $\bigtimes_{J \in \mathcal{L}_R} \Omega^J$ is independent. In contrast, for nondisjoint segments there are dependencies, such as

$$X_t^{I_\alpha} \in \Omega^{I_\alpha} \setminus \omega_{\mathrm{ind}}^{I_\alpha} \text{ for } \alpha \in G \text{ implies } X_t^J = \omega_\varnothing^J \text{ for all } J \in \mathcal{S}^{I_\alpha} \setminus I_\alpha. \tag{5.43}$$

The other way round, this means

$$X_t^J \neq \omega_\varnothing^J \text{ for some } J \in \mathcal{S}^{I_\alpha} \setminus I_\alpha \text{ implies } X_t^{I_\alpha} \in \omega_{I_\alpha}^{I_\alpha}. \tag{5.44}$$

Events and Waiting times. We now define events $\mathcal{E}_X(s)$ for all $s \in \Gamma$ based on the process X_t. To this end, recall that $\pi_I : \Omega \to \Omega^I$, $I \in \mathcal{S}$, is the canonical projection. We set for all $\alpha \in G$ and $J \in \mathcal{S}$:

$$\mathcal{E}_X(\alpha) := \Big\{\omega \in \Omega : \pi_{I_\alpha}(\omega) = \omega_\alpha^{I_\alpha}\Big\}, \qquad \mathcal{E}_X(J) := \Big\{\omega \in \Omega : \pi_J(\omega) = \omega_J^J\Big\}.$$

Due to (5.43) and (5.44), these events satisfy the nesting condition (5.23).

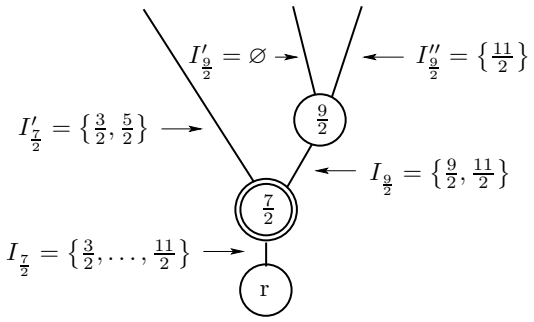


Figure 5.8. Segmentation tree with vertex set $G = \{\frac{7}{2}, \frac{9}{2}\}$, link set $L = \{\frac{3}{2}, \ldots, \frac{11}{2}\}$, internal segments $I_{\frac{7}{2}}$ and $I_{\frac{9}{2}}$, and external segments $I'_{\frac{7}{2}}$, $I'_{\frac{9}{2}}$ and $I''_{\frac{9}{2}}$. Here, $\mathcal{L}_{G_{\frac{9}{2}}(\varnothing)}^{I_{\frac{9}{2}}} = \{I'_{\frac{9}{2}}, I''_{\frac{9}{2}}\}$.

Example 5.1. Consider the segmentation tree in Figure 5.8. For every $t \in \mathbb{N}_0$, X_t is given by the family $X_t = \Big(X_t^{I'_{\frac{7}{2}}}, X_t^{I'_{\frac{9}{2}}}, X_t^{I''_{\frac{9}{2}}}, X_t^{I_{\frac{9}{2}}}, X_t^{I_{\frac{7}{2}}}\Big)$. Events $\omega \in \Omega$ that satisfy $\mathbf{P}[X_t = \omega] > 0$

are

$$\Big(\omega_{\varnothing}^{I'_{\frac{7}{2}}},\omega_{\varnothing}^{I'_{\frac{9}{2}}},\omega_{\varnothing}^{I''_{\frac{9}{2}}},\omega_{\varnothing}^{I_{\frac{9}{2}}},\omega_{\varnothing}^{I_{\frac{7}{2}}}\Big),\quad \Big(\omega_{\varnothing}^{I'_{\frac{7}{2}}},\omega_{\varnothing}^{I'_{\frac{9}{2}}},\omega_{\varnothing}^{I''_{\frac{9}{2}}},\omega_{\varnothing}^{I_{\frac{9}{2}}},\omega_{I_{\frac{7}{2}}}^{I_{\frac{7}{2}}}\Big),\quad \Big(\omega_{\varnothing}^{I'_{\frac{7}{2}}},\omega_{\varnothing}^{I'_{\frac{9}{2}}},\omega_{\varnothing}^{I''_{\frac{9}{2}}},\omega_{\varnothing}^{I_{\frac{9}{2}}},\omega_{\mathrm{dep}}^{I_{\frac{7}{2}}}\Big),$$
$$\Big(\omega_{\varnothing}^{I'_{\frac{7}{2}}},\omega_{\varnothing}^{I'_{\frac{9}{2}}},\omega_{\varnothing}^{I''_{\frac{9}{2}}},\omega_{I_{\frac{9}{2}}}^{I_{\frac{9}{2}}},\omega_{\mathrm{ind}}^{I_{\frac{7}{2}}}\Big),\quad \Big(\omega_{\varnothing}^{I'_{\frac{7}{2}}},\omega_{\varnothing}^{I'_{\frac{9}{2}}},\omega_{\varnothing}^{I''_{\frac{9}{2}}},\omega_{\mathrm{dep}}^{I_{\frac{9}{2}}},\omega_{\mathrm{ind}}^{I_{\frac{7}{2}}}\Big),\quad \Big(\omega_{I_{\frac{7}{2}}}^{I'_{\frac{7}{2}}},\omega_{\varnothing}^{I'_{\frac{9}{2}}},\omega_{\varnothing}^{I''_{\frac{9}{2}}},\omega_{\varnothing}^{I_{\frac{9}{2}}},\omega_{\mathrm{ind}}^{I_{\frac{7}{2}}}\Big),$$
$$\Big(\omega_{I'_{\frac{7}{2}}}^{I'_{\frac{7}{2}}},\omega_{\varnothing}^{I'_{\frac{9}{2}}},\omega_{\varnothing}^{I''_{\frac{9}{2}}},\omega_{I_{\frac{9}{2}}}^{I_{\frac{9}{2}}},\omega_{\mathrm{ind}}^{I_{\frac{7}{2}}}\Big),\quad \Big(\omega_{I'_{\frac{7}{2}}}^{I'_{\frac{7}{2}}},\omega_{\varnothing}^{I'_{\frac{9}{2}}},\omega_{\varnothing}^{I''_{\frac{9}{2}}},\omega_{\mathrm{dep}}^{I_{\frac{9}{2}}},\omega_{\mathrm{ind}}^{I_{\frac{7}{2}}}\Big),\quad \Big(\omega_{\varnothing}^{I'_{\frac{7}{2}}},\omega_{\varnothing}^{I'_{\frac{9}{2}}},\omega_{I''_{\frac{9}{2}}}^{I''_{\frac{9}{2}}},\omega_{\mathrm{ind}}^{I_{\frac{9}{2}}},\omega_{\mathrm{ind}}^{I_{\frac{7}{2}}}\Big),$$
$$\Big(\omega_{I'_{\frac{7}{2}}}^{I'_{\frac{7}{2}}},\omega_{\varnothing}^{I'_{\frac{9}{2}}},\omega_{I''_{\frac{9}{2}}}^{I''_{\frac{9}{2}}},\omega_{\mathrm{ind}}^{I_{\frac{9}{2}}},\omega_{\mathrm{ind}}^{I_{\frac{7}{2}}}\Big).$$

Events of interest are for example

$$\mathcal{E}_X(\tfrac{9}{2}) = \Big\{\Big(\omega_{\varnothing}^{I'_{\frac{7}{2}}},\omega_{\varnothing}^{I'_{\frac{9}{2}}},\omega_{\varnothing}^{I''_{\frac{9}{2}}},\omega_{I_{\frac{9}{2}}}^{I_{\frac{9}{2}}},\omega_{\mathrm{ind}}^{I_{\frac{7}{2}}}\Big),\ \Big(\omega_{I'_{\frac{7}{2}}}^{I'_{\frac{7}{2}}},\omega_{\varnothing}^{I'_{\frac{9}{2}}},\omega_{\varnothing}^{I''_{\frac{9}{2}}},\omega_{I_{\frac{9}{2}}}^{I_{\frac{9}{2}}},\omega_{\mathrm{ind}}^{I_{\frac{7}{2}}}\Big)\Big\}$$

and

$$\mathcal{E}_X(I_{\frac{9}{2}}) \setminus \mathcal{E}_X\Big(\mathcal{L}_{G_{\frac{9}{2}}(\varnothing)}^{I_{\frac{9}{2}}}\Big) = \Big\{\Big(\omega_{\varnothing}^{I'_{\frac{7}{2}}},\omega_{\varnothing}^{I'_{\frac{9}{2}}},\omega_{\varnothing}^{I''_{\frac{9}{2}}},\omega_{I_{\frac{9}{2}}}^{I_{\frac{9}{2}}},\omega_{\mathrm{ind}}^{I_{\frac{7}{2}}}\Big),\Big(\omega_{\varnothing}^{I'_{\frac{7}{2}}},\omega_{\varnothing}^{I'_{\frac{9}{2}}},\omega_{\varnothing}^{I''_{\frac{9}{2}}},\omega_{\mathrm{dep}}^{I_{\frac{9}{2}}},\omega_{\mathrm{ind}}^{I_{\frac{7}{2}}}\Big),$$
$$\Big(\omega_{I'_{\frac{7}{2}}}^{I'_{\frac{7}{2}}},\omega_{\varnothing}^{I'_{\frac{9}{2}}},\omega_{\varnothing}^{I''_{\frac{9}{2}}},\omega_{I_{\frac{9}{2}}}^{I_{\frac{9}{2}}},\omega_{\mathrm{ind}}^{I_{\frac{7}{2}}}\Big),\Big(\omega_{I'_{\frac{7}{2}}}^{I'_{\frac{7}{2}}},\omega_{\varnothing}^{I'_{\frac{9}{2}}},\omega_{\varnothing}^{I''_{\frac{9}{2}}},\omega_{\mathrm{dep}}^{I_{\frac{9}{2}}},\omega_{\mathrm{ind}}^{I_{\frac{7}{2}}}\Big)\Big\}. \qquad \diamondsuit$$

Let $\mathcal{T}_{\mathcal{E}_X(s)}$ denote the waiting time for the event $\mathcal{E}_X(s)$, $s \in \Gamma$ (condition (5.24) is then obviously satisfied). By construction, $\mathcal{T}_{\mathcal{E}_X(\alpha)}$ and $\mathcal{T}_{\mathcal{E}_X(J)}$ are geometrically distributed with parameters r_α and r_J, $\alpha \in G$, $J \in \mathcal{S}$. Since for every $\alpha \in G, H \subseteq E$, the family $\big(X_t^J\big)_{J \in \mathcal{L}_{G_\alpha(H)}^{I_\alpha}}$ is independent, the family of waiting times $\big(\mathcal{T}_{\mathcal{E}_X(J)}\big)_{J \in \mathcal{L}_{G_\alpha(H)}^{I_\alpha}}$ is independent as well, and, as a minimum of independent geometric variables, $\mathcal{T}_{\mathcal{E}_X(\mathcal{L}_{G_\alpha(H)}^{I_\alpha})}$ is geometrically distributed with parameter $1 - \lambda_{G_\alpha(H)}^{I_\alpha}$. For any $H \subseteq E$, the waiting time $\mathcal{T}_{\mathcal{E}_X(I_\alpha)\setminus\mathcal{E}_X(\mathcal{L}_{G_\alpha(H)}^{I_\alpha})}$ is geometric with parameter $r_{I_\alpha} - (1 - \lambda_{G_\alpha(H)}^{I_\alpha}) = \lambda_{G_\alpha(H)}^{I_\alpha} - \lambda_{\varnothing}^{I_\alpha}$ (recall that $\mathcal{E}_X(\mathcal{L}_{G_\alpha(H)}^{I_\alpha}) \subseteq \mathcal{E}_X(I_\alpha)$ by Fact (A)). Since the conditions of Theorem 5.3 are satisfied, we can directly conclude:

Corollary 1. *Let $T^L = (\gamma, G, E, L)$ be a segmentation tree. Then*

$$\mathbf{P}\big[\mathrm{Max}_{t,\mathcal{E}_X}(G),\, m_{\mathcal{E}_X}(\varnothing)\big] = \sum_{H \subseteq E} (-1)^{|H|}\, \mathbf{P}\big[\mathrm{Min}_{t,\mathcal{E}_X}(H),\, m_{\mathcal{E}_X}(H)\big],$$

with $m_{\mathcal{E}_X}(H)$, $\mathrm{Min}_{t,\mathcal{E}_X}(G)$ and $\mathrm{Max}_{t,\mathcal{E}_X}(G)$ as in (5.25)–(5.27), and $\mathcal{E}$ replaced by $\mathcal{E}_X$.

Constructing the segmentation process from the auxiliary process.

We now present a pathwise construction for realisations of $(\widehat{F}_t)_{t\in\mathbb{N}_0}$ that have the correct law as long as they are compatible with a given segmentation tree $T^L = (\gamma, G, E, L)$. We

say that $\widehat{F}_t$ is *compatible* with T^L if $\widehat{F}_t \in \mathcal{R}(T^L)$. In this case, $(\widehat{F}_{t'})_{0\leqslant t'\leqslant t}$ matches a stump tree of T^L. We use the auxiliary process $(X_t)_{t\in\mathbb{N}_0}$ for the construction.

Recall that the transition from $\widehat{F}_{t-1}$ to $\widehat{F}_t$ is determined by the family of independent random variables $(A_t^J)_{J\in\mathcal{L}_{\widehat{F}_{t-1}}}$ (see Definition 5.2). Now fix a tree $T^L = (\gamma, G, E, L)$ and construct the enlarged family $\big(A_t^J\big)_{J\in\mathcal{S}}$ from X_t by prescribing that, for all $t > 0$,

$$\begin{aligned} A_t^J = \varnothing, \text{ if and only if } X_t^J = \omega_\varnothing^J, \text{ for all } J \in \mathcal{S}, \\ A_t^{I_\alpha} = \{\alpha\}, \text{ if and only if } X_t^{I_\alpha} = \omega_\alpha^{I_\alpha}, \text{ for all } \alpha \in G. \end{aligned} \tag{5.45}$$

This entails that $A_t^J = \varnothing$ with probability $1 - r_J$ for all $J \in \mathcal{S}$, and $A_t^{I_\alpha} = \{\alpha\}$ with probability r_α, $\alpha \in G$. On the other hand, it implies that

$$A_t^J \in J, \text{ if and only if } X_t^J = \omega_J^J, \text{ for } J \in \mathcal{L}_G, \tag{5.46}$$

which happens with probability r_J, and

$$A_t^{I_\alpha} \in I_\alpha \setminus \alpha, \text{ if and only if } X_t^{I_\alpha} \in \omega_{I_\alpha}^{I_\alpha} \setminus \omega_\alpha^{I_\alpha}, \text{ for } \alpha \in G, \tag{5.47}$$

which is the case with probability $r_{I_\alpha\setminus\alpha}$. If we want to know the precise event in these cases, we can use additional chance to decide for $A_t^J = \{\beta\}$ with probability r_β for all $\beta \in J \in \mathcal{L}_G$, and $A_t^{I_\alpha} = \{\beta\}$, $\beta \in I_\alpha \setminus \alpha$, for all $\alpha \in G$, but this is never required in our construction; what matters is that, under the construction in (5.45), each A_t^J has the right probabilities for the *compatible events* (those in (5.45)) and their complements, for every given t and every given $J \in \mathcal{S}$. Also, the A_t^J inherit from the X_t^J the i.i.d. property over t and the independence across disjoint segments.

We now proceed as follows. Start with $\widehat{F}_0 = \varnothing$, which is certainly compatible with the given T^L. If $\widehat{F}_{t-1}$ is compatible, then construct $\widehat{F}_t$ from $\widehat{F}_{t-1}$ according to Definition 5.2, but use the $\big(A_t^J\big)_{J\in\mathcal{S}}$ from (5.45)–(5.47). If only *compatible events* occur for all $J \in \mathcal{L}_{\widehat{F}_{t-1}}$, then $\widehat{F}_t$ is compatible as well. If at least one *incompatible* event occurs (at least one event of those in (5.46) or (5.47)), then $\widehat{F}_t$ is incompatible. We say the construction *fails* at time t and discontinue it. Since the subfamily $(A_t^J)_{J\in\mathcal{L}_{\widehat{F}_{t-1}}}$ has the right law for the compatible events, we know that $(\widehat{F}_t)_{t\in\mathbb{N}_0}$ has the right law for all $t < t_f$, where t_f is the failure time.

Proposition 5.2. *For every given segmentation tree $T^L = (\gamma, G, E, L)$ and the pathwise construction of $\widehat{F}$ described above, we have*

$$\Big\{\widehat{\mathrm{Max}}_{t,\mathcal{E}_F}(G),\, \widehat{m}_{\mathcal{E}_F}(\varnothing)\Big\} \quad = \quad \Big\{\mathrm{Max}_{t,\mathcal{E}_X}(G),\, m_{\mathcal{E}_X}(\varnothing)\Big\}, \quad t \in \mathbb{N}_0 \tag{5.48}$$

and

$$\mathbf{P}\Big[\widehat{\mathrm{Max}}_{t,\mathcal{E}_F}(G),\, \widehat{m}_{\mathcal{E}_F}(\varnothing)\Big] = \mathbf{P}\Big[\mathrm{Max}_{t,\mathcal{E}_X}(G),\, m_{\mathcal{E}_X}(\varnothing)\Big], \quad t \in \mathbb{N}_0, \tag{5.49}$$

where $\widehat{\mathrm{Max}}_{t,\mathcal{E}_F}(G)$ and $\widehat{m}_{\mathcal{E}_F}(\varnothing)$ are the discrete-time versions of (5.25) *and* (5.26) *with $\mathcal{E}$ replaced by $\mathcal{E}_F$.*

The description in terms of the waiting times of the auxiliary process offers a great advantage since this law is known and does not change over time.

Proof. We start by considering the events $\mathrm{Max}_{t,\mathcal{E}}(G)$ and $m_{\mathcal{E}}(\varnothing)$ for general $\mathcal{E}$. Given $\mathrm{Max}_{t,\mathcal{E}}(G)$, we know $\mathcal{T}_{\mathcal{E}(\alpha)} < \mathcal{T}_{\mathcal{E}(\mathcal{L}_G)}$ for all $\alpha \in G$. Since by Fact 5.6 (C) furthermore $\mathcal{T}_{\mathcal{E}(\mathcal{L}_G)} \leqslant \mathcal{T}_{\mathcal{E}(\mathcal{L}^{I_\alpha}_{G_\alpha(\varnothing)})}$ for all $\alpha \in G$, we obtain, given $m_{\mathcal{E}}(\varnothing)$:

$$\Big\{\mathcal{T}_{\mathcal{E}(\alpha)} = \mathcal{T}_{\mathcal{E}(I_\alpha)\setminus\mathcal{E}(\mathcal{L}^{I_\alpha}_{G_\alpha(\varnothing)})}\Big\} = \Big\{\mathcal{T}_{\mathcal{E}(\alpha)} = \min\Big\{\mathcal{T}_{\mathcal{E}(I_\alpha)\setminus\mathcal{E}(\mathcal{L}^{I_\alpha}_{G_\alpha(\varnothing)})}, \mathcal{T}_{\mathcal{E}(\mathcal{L}^{I_\alpha}_{G_\alpha(\varnothing)})}\Big\}\Big\} = \Big\{\mathcal{T}_{\mathcal{E}(\alpha)} = \mathcal{T}_{\mathcal{E}(I_\alpha)}\Big\}$$

for every $\alpha \in G$. We can therefore rewrite

$$\Big\{\mathrm{Max}_{t,\mathcal{E}}(G),\, m_{\mathcal{E}}(\varnothing)\Big\} = \Big\{\mathrm{Max}_{t,\mathcal{E}}(G),\, \bigcap_{\alpha\in G}\{\mathcal{T}_{\mathcal{E}(\alpha)} = \mathcal{T}_{\mathcal{E}(I_\alpha)}\}\Big\}. \tag{5.50}$$

The choice $\mathcal{E} = \mathcal{E}_F$ or $\mathcal{E} = \mathcal{E}_X$ in (5.50) turns the claim (5.48) into

$$\Big\{\widehat{\mathrm{Max}}_{t,\mathcal{E}_F}(G),\, \bigcap_{\alpha\in G}\{\widehat{\mathcal{T}}_{\mathcal{E}_F(\alpha)} = \widehat{\mathcal{T}}_{\mathcal{E}_F(I_\alpha)}\}\Big\} = \Big\{\mathrm{Max}_{t,\mathcal{E}_X}(G),\, \bigcap_{\alpha\in G}\{\mathcal{T}_{\mathcal{E}_X(\alpha)} = \mathcal{T}_{\mathcal{E}_X(I_\alpha)}\}\Big\}. \tag{5.51}$$

Recall that $\mathcal{E}_F(s) = \{s\}$ for all $s \in \Gamma$, such that $\widehat{\mathcal{T}}_{\mathcal{E}_F(s)} = \widehat{\mathcal{T}}_s = \min\{\widehat{\mathcal{T}}_\alpha : \alpha \in s\}$ is the time at which the first link in s is removed. Now, assume that we have shown the identification

$$\widehat{\mathcal{T}}_J = \mathcal{T}_{\mathcal{E}_X(J)} \text{ for all } J \in \mathcal{S} \text{ given } \bigcap_{\alpha\in G}\{\widehat{\mathcal{T}}_\alpha = \widehat{\mathcal{T}}_{I_\alpha}\}. \tag{5.52}$$

Due to (5.45)–(5.47), it then follows under the pathwise construction of $\widehat{F}$ from the auxiliary process that $\{\widehat{\mathcal{T}}_\alpha = \widehat{\mathcal{T}}_{I_\alpha}\} = \{\mathcal{T}_{\mathcal{E}_X(\alpha)} = \mathcal{T}_{\mathcal{E}_X(I_\alpha)}\}$. Together with (5.52), this implies $\widehat{\mathcal{T}}_\alpha = \mathcal{T}_{\mathcal{E}_X(\alpha)}$ for all $\alpha \in G$. Equation (5.52) therefore entails (5.51), so it suffices to show (5.52).

We first show the relation (5.52) for all internal segments (i.e. for all I_α, $\alpha \in G$). Start with the set of links $I_\gamma = L$ and initial value $\widehat{F}^L_0 = \{\varnothing\}$. For $t \geqslant 1$, the first event $\{\widehat{F}^L_t \neq \varnothing\}$ happens when $\{A^L_t \in L\}$ for the first time; this happens at $t = \widehat{\mathcal{T}}_L$. Under (5.45) $\widehat{\mathcal{T}}_L$ corresponds to the first time at which $\{X^L_t \in \omega^L_L\}$ (at time $\mathcal{T}_{\mathcal{E}_X(L)}$); this gives $\widehat{\mathcal{T}}_L = \mathcal{T}_{\mathcal{E}_X(L)}$.

Now consider a link $\beta \in G \setminus \{\gamma\}$, and assume that we have already identified $\widehat{\mathcal{T}}_{I_\nu} = \mathcal{T}_{\mathcal{E}_X(I_\nu)}$ for the parent node ν of β. Given $\widehat{\mathcal{T}}_\nu = \widehat{\mathcal{T}}_{I_\nu}$, we conclude $\widehat{\mathcal{T}}_{I_\nu} < \widehat{\mathcal{T}}_{I_\beta}$ since $\nu \notin I_\beta \subset I_\nu$. This yields $\widehat{F}^{I_\beta}_{\widehat{\mathcal{T}}_{I_\nu}} = \varnothing$ by (5.5). Now consider the first time $t' > \widehat{\mathcal{T}}_{I_\nu}$ the event $\{\widehat{F}^{I_\beta}_{t'} \neq \varnothing\}$ occurs. Again by (5.5), this time is $\widehat{\mathcal{T}}_{I_\beta}$. Due to (5.45), the event $\{\widehat{F}^{I_\beta}_{t'} \neq \varnothing\}$ with $t' > \widehat{\mathcal{T}}_{I_\nu}$ happens when $\{X^{I_\beta}_{t'} \in \omega^{I_\beta}_{I_\beta}\}$ for the first time, which is at time $\mathcal{T}_{\mathcal{E}_X(I_\beta)}$. Since we assumed that $\widehat{\mathcal{T}}_{I_\nu} = \mathcal{T}_{\mathcal{E}_X(I_\nu)}$ and since $\mathcal{T}_{\mathcal{E}_X(I_\nu)} \leqslant \mathcal{T}_{\mathcal{E}_X(I_\beta)}$ due to (5.23), we conclude $\mathcal{T}_{\mathcal{E}_X(I_\beta)} > \widehat{\mathcal{T}}_{I_\nu}$, which gives $\mathcal{T}_{\mathcal{E}_X(I_\beta)} = \widehat{\mathcal{T}}_{I_\beta}$.

It remains to show the equality of the waiting times for the full external segments $J \in \mathcal{L}_G$. For each such segment J, denote by $\delta := \delta_J \in G$ the unique link for which $J \in \{I'_\delta, I''_\delta\}$.

Assume that $\widehat{\mathcal{T}}_\delta = \widehat{\mathcal{T}}_{I_\delta}$ and one has already identified $\widehat{\mathcal{T}}_{I_\delta} = \mathcal{T}_{\mathcal{E}_X(I_\delta)}$. With the same arguments as above, we conclude that under the given assumption $\widehat{\mathcal{T}}_J = \mathcal{T}_{\mathcal{E}_X(J)}$.

To finally show (5.49) recall that, for a given segmentation tree T^L, $(\widehat{F}_t)_{t\in\mathbb{N}_0}$ has the right law for all $t < t_f$, where t_f is the first time at which $\widehat{F}_{t_f}$ fails to be compatible with the tree. Since $\{\widehat{\mathrm{Max}}_{t,\mathcal{E}_F}(G), \widehat{m}_{\mathcal{E}_F}(\varnothing)\}$ describes a sequence of events that are all compatible with T^L, (5.49) follows. □

Before we give an explicit expression for the tree probabilities in discrete time, let us comment on the meaning of the joint event $\{\mathrm{Max}_{t,\mathcal{E}_X}(G),\, m_{\mathcal{E}_X}(\varnothing)\}$ for the auxiliary process and compare it with the corresponding joint event in the segmentation process. We start with the ancestral relation and compare the event $\mathcal{E}(I_\alpha) \setminus \mathcal{E}(\mathcal{L}^{I_\alpha}_{G_\alpha(H)})$ for the choices $\mathcal{E} = \mathcal{E}_F$ and $\mathcal{E} = \mathcal{E}_X$ for general $H \subseteq E$ and $\alpha \in G$. Recall that, for the segmentation process, we set $\mathcal{E}_F(s) = \{s\}$ and obtained in (5.34) that $\mathcal{E}_F(I_\alpha) \setminus \mathcal{E}_F(\mathcal{L}_{G_\alpha(H)}) = G_\alpha(H)$. For the auxiliary process the events are more subtle. We find

$$\begin{aligned}
\mathcal{E}_X(I_\alpha) \setminus \mathcal{E}_X(\mathcal{L}^{I_\alpha}_{G_\alpha(H)}) &= \{\omega \in \Omega : \pi_{I_\alpha}(\omega) \in \omega^{I_\alpha}_{I_\alpha}, \\
&\qquad \text{and } \pi_J(\omega) = \omega^J_\varnothing \text{ for all } J \in \mathcal{L}^{I_\alpha}_{G_\alpha(H)}\} \qquad (5.53)\\
&= \bigcup_{\beta\in G_\alpha(H)} \Big\{\omega \in \Omega : \pi_{I_\beta}(\omega) \in \{\omega^{I_\beta}_\beta, \omega^{I_\beta}_{\mathrm{dep}}\}, \\
&\qquad \text{and } \pi_J(\omega) = \omega^J_\varnothing \text{ for all } J \in \mathcal{L}^{I_\alpha}_{G_\alpha(H)}\Big\} \\
&= \bigcup_{\beta\in G_\alpha(H)} \Big\{\omega \in \Omega : \pi_{I_\beta}(\omega) \in \{\omega^{I_\beta}_\beta, \omega^{I_\beta}_{\mathrm{dep}}\}, \\
&\qquad \text{and } \pi_J(\omega) = \omega^J_\varnothing \text{ for all } J \in \mathcal{L}^{I_\alpha}_{G_\alpha(H)} \setminus \mathcal{L}^{I_\beta}_{G_\beta(H)}\Big\}, \qquad (5.54)
\end{aligned}$$

see also Example 5.1. The first equality is a reformulation of $\mathcal{E}_X(I_\alpha) \setminus \mathcal{E}_X(\mathcal{L}^{I_\alpha}_{G_\alpha(H)})$ in terms of the explicit events $\mathcal{E}_X(I_\alpha) \setminus \mathcal{E}_X(\mathcal{L}^{I_\alpha}_{G_\alpha(H)})$ is composed of. The second equality follows inductively: $\pi_{I_\alpha}(\omega) \in \omega^{I_\alpha}_{I_\alpha}$ is equivalent to either $\pi_{I_\alpha}(\omega) \in \{\omega^{I_\alpha}_\alpha, \omega^{I_\alpha}_{\mathrm{dep}}\}$, or $\pi_{I_\alpha}(\omega) = \omega^{I_\alpha}_{\mathrm{ind}}$. In the first case, this implies $\pi_I(\omega) = \omega^I_\varnothing$ for all $I \in \mathcal{S}^{I_\alpha} \setminus \{I_\alpha\}$ by (5.43). In the second case, there exists (at least) one segment $K \in \{I'_\alpha, I''_\alpha\}$ such that $\pi_K(\omega) = \omega^K_K$. Climbing up the tree in a bottom-up manner yields the expression. The third equality follows again from the fact that $\pi_{I_\beta}(\omega) \in \{\omega^{I_\beta}_\beta, \omega^{I_\beta}_{\mathrm{dep}}\}$ implies $\pi_J(\omega) = \omega^J_\varnothing$ for all $J \in \mathcal{L}^{I_\beta}_{G_\beta(H)}$ by (5.43). It is interesting to note here that the formulation in (5.54) requires information about the complete partial order within the component $T_\alpha(H)$, whereas in (5.53), it is sufficient to know the vertex set $G_\alpha(H)$ of the subtree (analogous to the situation in the segmentation process).

Regarding now the ancestor relation $m_{\mathcal{E}_X}(H)$, we see from (5.54) that, for a given $\alpha \in G$, $\mathcal{E}_X(\alpha)$ not only competes with all $\mathcal{E}_X(\beta)$, $\beta \in G_\alpha(H)$, but also with the corresponding dependent events — but the result only counts when 'nothing happens' in the disjoint segments in the *same* subtree ($\pi_J = \omega^J_\varnothing$ for all $J \in \mathcal{L}^{I_\alpha}_{G_\alpha(H)} \setminus \mathcal{L}^{I_\beta}_{G_\beta(H)}$). These conditions are far more intricate compared to $m_{\mathcal{E}_F}(H)$, where α simply needs to be the first link that is removed in the vertex set $G_\alpha(H)$ of $T_\alpha(H)$.

In analogy with $\mathrm{Max}_{t,\mathcal{E}_F}(G)$, $\mathrm{Max}_{t,\mathcal{E}_X}(G)$ is the event that all $\mathcal{E}_X(\alpha)$, $\alpha \in G$, appear before t and none of the $\mathcal{E}_X(J)$, where the $J \in \mathcal{L}_G$ are the external segments. This means in particular that 'dependent' events in the internal segments are allowed to show up before t, provided nothing happens in the external segments. Altogether, the joint event $\{\mathrm{Max}_{t,\mathcal{E}_X}(G),\ m_{\mathcal{E}_X}(\varnothing)\}$ therefore says that every $\mathcal{E}_X(\alpha)$ needs to appear before the corresponding dependent event (provided nothing happens to the disjoint segments in the same subtree) and before t, but once $\mathcal{E}_X(\alpha)$ appeared, neither this nor the corresponding dependent part has an effect.

5.3.4 Tree probabilities in discrete time

We can now harvest the consequences and state an explicit expression for tree probabilities in discrete time. Denote by $\widehat{\mathcal{F}}_t(T^L)$ the discrete-time version of (5.22).

Proposition 5.3. *For a given segmentation tree $T^L = (\gamma, G, E, L)$ and $t \in \mathbb{N}_0$, one has $\mathbf{P}[\widehat{\mathcal{F}}_t(T^L)] = (1 - r_L)^t = (\lambda^L_\varnothing)^t$ for $G = \varnothing$, and, for $G \neq \varnothing$,*

$$\mathbf{P}[\widehat{\mathcal{F}}_t(T^L)] = \sum_{H \subseteq E} (-1)^{|H|} \left[(\lambda^L_{G_\gamma(H)})^t - (\lambda^L_\varnothing)^t \right] \prod_{\alpha \in G} \frac{r_\alpha}{\lambda^{I_\alpha}_{G_\alpha(H)} - \lambda^{I_\alpha}_\varnothing},$$

where the λ's are defined as in (5.40).

Proof. We first employ Proposition 5.2 together with Corollary 5.1 to rewrite the matching probability corresponding to the segmentation process in terms of the auxiliary process:

$$\mathbf{P}[\widehat{\mathcal{F}}_t(T^L)] = \mathbf{P}[\mathrm{Max}_{t,\mathcal{E}_X}(G),\, m_{\mathcal{E}_X}(\varnothing)] = \sum_{H \subseteq E} (-1)^{|H|}\, \mathbf{P}[\mathrm{Min}_{t,\mathcal{E}_X}(G_\gamma(H)),\, m_{\mathcal{E}_X}(H)]. \tag{5.55}$$

Now fix a set of edges $H \subseteq E$ and consider the event $\mathrm{Min}_{t,\mathcal{E}_X}(G_\gamma(H))$ on the right-hand side of (5.55). In contrast to the continuous-time case, the family $\big(\mathcal{T}_{\mathcal{E}_X(\alpha)}\big)_{\alpha \in G_\gamma(H)}$ is not independent, so $\min\{\mathcal{T}_{\mathcal{E}_X(\alpha)} : \alpha \in G_\gamma(H)\}$ is not a simple geometric waiting time. But, given $m_{\mathcal{E}_X}(H)$, we can use that $\min\{\mathcal{T}_{\mathcal{E}_X(\alpha)} : \alpha \in G_\gamma(H)\} = \mathcal{T}_{\mathcal{E}_X(\gamma)}$ by Fact 5.6 (B) and, again due to $m_{\mathcal{E}_X}(H)$, that $\mathcal{T}_{\mathcal{E}_X(\gamma)} = \mathcal{T}_{\mathcal{E}_X(\mathcal{L}_\varnothing)\setminus\mathcal{E}_X(\mathcal{L}_{G_\gamma(H)})}$ since $I_\gamma = L = \mathcal{L}_\varnothing$. This gives

$$\begin{aligned}
\Big\{\mathrm{Min}_{t,\mathcal{E}_X}(G_\gamma(H)),\, m_{\mathcal{E}_X}(H)\Big\} &= \Big\{\mathcal{T}_{\mathcal{E}_X(\mathcal{L}_\varnothing)\setminus\mathcal{E}_X(\mathcal{L}_{G_\gamma(H)})} \leqslant t < \mathcal{T}_{\mathcal{E}_X(\mathcal{L}_{G_\gamma(H)})},\, m_{\mathcal{E}_X}(H)\Big\} \\
&= \Big\{ \min\big\{\mathcal{T}_{\mathcal{E}_X(\mathcal{L}_\varnothing)\setminus\mathcal{E}_X(\mathcal{L}_{G_\gamma(H)})}, \mathcal{T}_{\mathcal{E}_X(\mathcal{L}_{G_\gamma(H)})}\big\} \leqslant t < \mathcal{T}_{\mathcal{E}_X(\mathcal{L}_{G_\gamma(H)})}, \\
&\qquad m_{\mathcal{E}_X}(H)\Big\} \\
&= \Big\{\mathcal{T}_{\mathcal{E}_X(\mathcal{L}_\varnothing)} \leqslant t < \mathcal{T}_{\mathcal{E}_X(\mathcal{L}_{G_\gamma(H)})},\, m_{\mathcal{E}_X}(H)\Big\}.
\end{aligned}$$

Let us now investigate the connection between $\{\mathcal{T}_{\mathcal{E}_X(\mathcal{L}_\varnothing)} \leqslant t < \mathcal{T}_{\mathcal{E}_X(\mathcal{L}_{G_\gamma(H)})}\}$ and $m_{\mathcal{E}_X}(H)$. To this end, consider first an $\alpha \notin G_\gamma(H)$. For this we know that there is a $J \in \mathcal{L}_{G_\gamma(H)}$ such that $I_\alpha \subseteq J$; thus $\mathcal{E}_X(\alpha) \subseteq \mathcal{E}_X(I_\alpha) \setminus \mathcal{E}_X(\mathcal{L}^{I_\alpha}_{G_\alpha(H)}) \subseteq \mathcal{E}_X(\mathcal{L}_{G_\gamma(H)}) \subseteq \mathcal{E}_X(\mathcal{L}_\varnothing)$ by (5.23).

Since the minimum of a collection of events is independent of the order in which (some of) the events occur, we obtain the independence of $\{\mathcal{T}_{\mathcal{E}_X(\mathcal{L}_\varnothing)} \leqslant t < \mathcal{T}_{\mathcal{E}_X(\mathcal{L}_{G_\gamma(H)})}\}$ and $\{\mathcal{T}_{\mathcal{E}_X(\alpha)} = \mathcal{T}_{\mathcal{E}_X(I_\alpha)\setminus\mathcal{E}_X(\mathcal{L}^{I_\alpha}_{G_\alpha(H)})}\}$ for every $\alpha \notin G_\gamma(H)$. Consider now $\alpha \in G_\gamma(H)$. We can then obviously decompose the event $\{\mathcal{T}_{\mathcal{E}_X(\mathcal{L}_\varnothing)} \leqslant t < \mathcal{T}_{\mathcal{E}_X(\mathcal{L}_{G_\gamma(H)})}\}$ into

$$\Big\{\mathcal{T}_{\mathcal{E}_X(\mathcal{L}_\varnothing)} \leqslant t\Big\} \cap \Big\{\mathcal{T}_{\mathcal{E}_X(\mathcal{L}^L_{G_\gamma(H)}\setminus\mathcal{L}^{I_\alpha}_{G_\alpha(H)})} > t\Big\} \cap \Big\{\mathcal{T}_{\mathcal{E}_X(\mathcal{L}^{I_\alpha}_{G_\alpha(H)})} > t\Big\}.$$

Due to the independence of the X_t^J's for disjoint sets J, we conclude that the event $\{\mathcal{T}_{\mathcal{E}_X(\alpha)} = \mathcal{T}_{\mathcal{E}_X(I_\alpha)\setminus\mathcal{E}_X(\mathcal{L}^{I_\alpha}_{G_\alpha(H)})}\}$ is independent of the event $\{\mathcal{T}_{\mathcal{E}_X(\mathcal{L}_{G_\gamma(H)}\setminus\mathcal{L}^{I_\alpha}_{G_\alpha(H)})} > t\}$. The independence of the event $\{\mathcal{T}_{\mathcal{E}_X(\alpha)} = \mathcal{T}_{\mathcal{E}_X(I_\alpha)\setminus\mathcal{E}_X(\mathcal{L}^{I_\alpha}_{G_\alpha(H)})}\}$ of $\{\mathcal{T}_{\mathcal{E}_X(\mathcal{L}^{I_\alpha}_{G_\alpha(H)})} > t\}$ is obvious since the respective events $\mathcal{E}_X(I_\alpha)\setminus\mathcal{E}_X(\mathcal{L}^{I_\alpha}_{G_\alpha(H)})$ and $\mathcal{E}_X(\mathcal{L}^{I_\alpha}_{G_\alpha(H)})$ are disjoint; the independence of $\{\mathcal{T}_{\mathcal{E}_X(\mathcal{L}_\varnothing)} \leqslant t\}$ follows again by the argument that the minimum of a collection of events is independent of the order in which (some of) the events occur. Altogether, we obtain

$$\mathbf{P}\big[\mathrm{Min}_{t,\mathcal{E}_X}(G_\gamma(H)),\, m_{\mathcal{E}_X}(H)\big] = \Big[\mathbf{P}\big[\mathcal{T}_{\mathcal{E}_X(\mathcal{L}_{G_\gamma(H)})} > t\big] - \mathbf{P}\big[\mathcal{T}_{\mathcal{E}_X(\mathcal{L}_\varnothing)} > t\big]\Big] \times \mathbf{P}\big[m_{\mathcal{E}_X}(H)\big],$$

where we used that $\mathcal{E}_X(\mathcal{L}_{G_\gamma(H)}) \subseteq \mathcal{E}_X(\mathcal{L}_\varnothing)$ by Fact 5.6 (A). Since for $\alpha, \beta \in G$ with $\alpha \prec \beta$, $\mathcal{E}_X(\alpha) \notin \mathcal{E}_X(I_\beta)$ and hence $\mathcal{E}_X(\alpha) \notin \mathcal{E}_X(I_\beta)\setminus\mathcal{E}_X(\mathcal{L}_{G_\alpha(H)})$, we can furthermore decompose the probability for $m_{\mathcal{E}_X}(H)$ into independent factors:

$$\mathbf{P}\big[m_{\mathcal{E}_X}(H)\big] = \prod_{\alpha\in G} \mathbf{P}\Big[\mathcal{T}_{\mathcal{E}_X(\alpha)} = \mathcal{T}_{\mathcal{E}_X(I_\alpha)\setminus\mathcal{E}_X(\mathcal{L}^{I_\alpha}_{G_\alpha(H)})}\Big].$$

Now recall, that each $\mathcal{T}_{\mathcal{E}_X(\mathcal{L}^{I_\alpha}_{G_\alpha(H)})}$ is geometric with parameter $1-\lambda^{I_\alpha}_{G_\alpha(H)}$ and that each $\mathcal{T}_{\mathcal{E}_X(I_\alpha)\setminus\mathcal{E}_X(\mathcal{L}^{I_\alpha}_{G_\alpha(H)})}$ is geometric with parameter $\lambda^{I_\alpha}_{G_\alpha(H)} - \lambda^{I_\alpha}_{\varnothing}$. All in all, we obtain

$$\begin{aligned}\mathbf{P}\big[\mathrm{Min}_{t,\mathcal{E}_X}(G_\gamma(H)),\, m_{\mathcal{E}_X}(H)\big] &= \Big[\big(1-(1-\lambda^L_{G_\gamma(H)})\big)^t - \big(1-(1-\lambda^L_{\varnothing})\big)^t\Big] \prod_{\alpha\in G} \frac{r_\alpha}{\lambda^{I_\alpha}_{G_\alpha(H)} - \lambda^{I_\alpha}_{\varnothing}} \\ &= \Big[\big(\lambda^L_{G_\gamma(H)}\big)^t - \big(\lambda^L_{\varnothing}\big)^t\Big] \prod_{\alpha\in G} \frac{r_\alpha}{\lambda^{I_\alpha}_{G_\alpha(H)} - \lambda^{I_\alpha}_{\varnothing}}\,.\end{aligned}$$

Equation (5.55) then completes the proof. □

Remark 5.5. For better comparison of the probability of a segmentation tree in continuous time (Proposition 5.1) and discrete time (Proposition 5.3), we can use the reformulation

$$\lambda^{I_\alpha}_{G_\alpha(H)} - \lambda^{I_\alpha}_{\varnothing} = \prod_{J\in\mathcal{L}^{I_\alpha}_{G_\alpha(H)}} (1-r_J) - \Big(1-\sum_{\nu\in I_\alpha} r_\nu\Big) = \sum_{\nu\in G_\alpha(H)} r_\nu + \sum_{\substack{J\subseteq\mathcal{L}^{I_\alpha}_{G_\alpha(H)},\\ |J|>1}} (-1)^{|J|} \prod_{I\in J} r_I. \;\diamondsuit$$

Corollary 2. *Let $\hat{a}_t$ be the coefficient function corresponding to the solution of the discrete-time deterministic recombination equation from* (5.1)*. For every $G \subseteq L$,*

$$\hat{a}_t(G) = \mathbf{P}\big[\widehat{F}_t = G\big] = \sum_{T^L \in \tau(G,L)} \mathbf{P}\big[\widehat{\mathcal{F}}_t(T^L)\big], \tag{5.56}$$

with $\mathbf{P}\big[\widehat{\mathcal{F}}_t(T^L)\big]$ as in Proposition 5.3 and where $\tau(G,L)$ is the set of all segmentation trees with vertex set G and underlying link set L.

Again, in contrast to the continuous-time case, there is (in general) no simple explicit expression for the sum in (5.56). But Remark 5.5 shows that there is one exception, namely the case $|\mathcal{L}^{I_\alpha}_{G_\alpha(H)}| \leqslant 1$ for every $\alpha \in G$ and every $H \subseteq E$. If $L = \{\frac{3}{2}, \dots, \frac{2n-1}{2}\}$, this is true for $G \subseteq \{\frac{3}{2}, \frac{2n-1}{2}\}$, in which case

$$\hat{a}_t(G) = \mathbf{P}\big[\widehat{F}_t = G\big] = \sum_{\varnothing \neq H \subseteq G} (-1)^{|H|} \, \big[(\lambda^L_H)^t - (\lambda^L_\varnothing)^t\big].$$

As it must be, we rediscover the $\hat{a}_t(G)$'s from (2.9) obtained via forward methods. The difference between two or three sites compared to a larger number of sites, and thus the possibility of finding an explicit solution via forward methods in the three-site case, now becomes clear in the light of our event structure: $\alpha \in \{\frac{3}{2}, \frac{2n-1}{2}\}$, implies that either $I'_\alpha = \varnothing$ or $I''_\alpha = \varnothing$, so that the probability for $\omega^{I_\alpha}_{\text{dep}}$ vanishes (cf. (5.42)). Any subset of links that only contains the 'ends' of L therefore induces significant simplifications.

Remark 5.6. A closed form expression for the $\hat{a}_t(G)$'s in terms of sums of probabilities of trees is also given in [93, Thm. 4.2] (even for general multi-crossover recombination). The trees considered in [93] represent single realisations of the segmentation process and encode the state of this realisation at each time step. The probability for any given tree is decomposed into the probability for each path of the tree (from the root to a leaf). For each path, the probability is a product of one-step transitions from node to node and not simplified any further. The number of trees that need to be considered increases incredibly fast with t. In contrast, our trees represent *sets* of realisations of the segmentation process (cf. Fig. 5.2). The number of trees that need to be considered thus depends on the state of the segmentation process at time t only, irrespective of how long any realisation may stay in the various intermediate states. ♢

5.4 Outlook: Multi-crossover recombination

So far, we restricted ourselves to single-crossover recombination, for which the partitioning process (starting with an appropriate initial state) takes values in the set of ordered partitions and has a one-to-one correspondence to the segmentation process on the powerset of removed links. Let us now describe how the quite general framework that we established for single crossovers can be transferred to obtain results in the general recombination case, for which the description via links is no longer sufficient.

If we allow multiple crossovers, the deterministic limit of the partitioning process $(\Sigma'_t)_{t\in\mathbb{T}}$ as defined in Proposition 3.1 and Proposition 3.2 takes values in $\mathbb{P}(S)$, the set of partitions of

$S = \{1, \dots, n\}$. Since $(\Sigma'_t)_{t\in\mathbb{T}}$ is a process of progressive refinements, we can again represent each set of realisations of $(\Sigma'_t)_{t\in\mathbb{T}}$ that start in $\Sigma'_t = \mathbf{1}$ and that agree on the partial order of events until time t by a particular rooted tree. In contrast to the single-crossover case, each node in the rooted tree no longer represents a single link but rather a partition of a subset of sites into (exactly) two blocks. We will denote such a tree by $T = (\mathcal{C}, \mathbb{G}, E)$, where $\mathbb{G}$ is the set of vertices, $\mathcal{C}$ is the root and E is the edge set. For simplicity, let T be equipped with a left-to-right order defined as follows: for any vertex $\mathcal{A} = \{A_1, A_2\}$, define the child corresponding to the partition in $\mathbb{P}_2(A_1)$ as the left child and the child corresponding to $\mathbb{P}_2(A_2)$ as the right child of $\mathcal{A}$. The notation for subtrees and rooted forests from Section 5.2 carries over.

Analogous to the single-crossover case, we can augment every tree $T = (\mathcal{C}, \mathbb{G}, E)$ with information about the relevant segments, which are here simply the blocks of the partitions in the tree. Let us collect all such segments into $\mathcal{S} := \{A : A \in \mathcal{A},\ \mathcal{A} \in \mathbb{G} \cup \mathbf{1}\}$. Similar to the construction in Section 5.3, we construct a segmentation tree $T^S := (\mathcal{C}, \mathbb{G}, E, \mathcal{S})$ corresponding to $T = (\mathcal{C}, \mathbb{G}, E)$ as follows:

- Add additional branches to T such that every vertex $\mathcal{A} \in \mathbb{G}$ has exactly two lines emanating from it. Add a phantom node r to the tree which corresponds to the coarsest partition in $\mathbb{P}(S)$, namely $\mathbf{1} = \{S\}$. Connect $\mathbf{1}$ and $\mathcal{C}$ by a single branch.
- Associate every line (edge or branch) with a segment $J \in \mathcal{S}$ according to the following rules. Start with the line between $\mathbf{1}$ and $\mathcal{C}$ and identify it with $I_{\mathcal{C}} = \{S\}$. If $\mathcal{C} = \{C_1, C_2\}$, associate the two lines emanating from $\mathcal{C}$ with the segments $I'_{\mathcal{C}} := C_1$ and $I''_{\mathcal{C}} := C_2$; so $I'_{\mathcal{C}}$ is the left and $I''_{\mathcal{C}}$ is the right branch or edge; obviously $I_{\mathcal{C}} = \{I'_{\mathcal{C}}, I''_{\mathcal{C}}\}$. Proceed up the tree as described in Section 5.3.

Via this construction, every internal segment $I_{\mathcal{A}}$ satisfies $I_{\mathcal{A}} = \bigcup_{A\in\mathcal{A}} A$, $\mathcal{A} \in \mathbb{G}$. All external segments are captured in the set $\mathcal{L}_{\mathbb{G}} := \bigwedge_{\mathcal{A}\in\mathbb{G}} (\mathcal{A} \cup (S \setminus I_{\mathcal{A}}))$, where $\wedge$ denotes the greatest lower bound of a set of partitions (see Section 1.2.1). $\mathcal{L}_{\mathbb{G}}$ is by construction a partition of $\mathbb{P}(S)$.

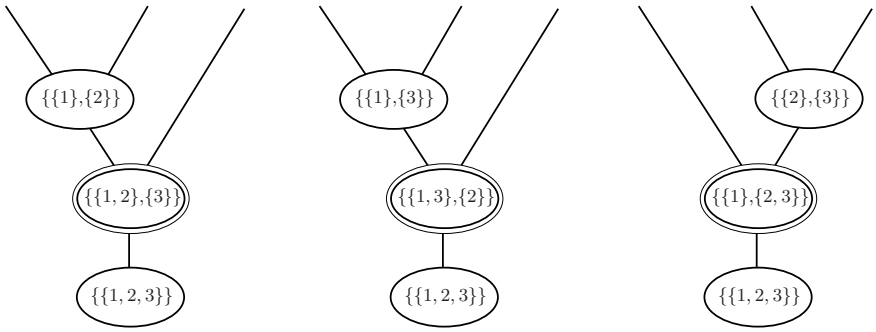


Figure 5.9. Three different segmentation trees that lead to the state $\Sigma'_t = \{\{1\}, \{2\}, \{3\}\}$ at some time $t \geqslant 0$. The first two trees share the tree topology.

Aiming at an explicit probability distribution of Σ'_t, we can again rely on a formulation via waiting times. For any $U \subseteq S$ and any partition $\mathcal{A} \in \mathbb{P}_2(U)$, let us therefore define $\mathcal{T}_{\mathcal{A}} := \min\{t \geqslant 0 : \Sigma'_t|_U = \mathcal{A}\}$ as the first time a splitting-event according to $\mathcal{A}$ occurs. For every $U \subseteq S$, denote by $\mathcal{T}_U := \min\{\mathcal{T}_{\mathcal{B}} : \mathcal{B} \in \mathbb{P}_2(U)\}$ the first time that the segment U is

split up, and by $\mathcal{T}_{\mathcal{L}_\mathbb{G}}$ the first time any segment in $\mathcal{L}_\mathbb{G}$ is split up. The event that $(\Sigma'_t)_{0\leqslant t'\leqslant t}$ matches T^S in the general recombination case is then

$$\mathcal{F}_t(T^S) = \Big\{ \max\{\mathcal{T}_\mathcal{A} : \mathcal{A} \in \mathbb{G}\} \leq t < \mathcal{T}_{\mathcal{L}_\mathbb{G}},\ \mathcal{T}_\mathcal{A} = \min\{\mathcal{T}_\mathcal{B} : \mathcal{B} \in \mathbb{G}_\mathcal{A}(\varnothing)\}\ \forall\ \mathcal{A} \in \mathbb{G}\Big\}, \quad (5.57)$$

where $\mathbb{G}_\mathcal{A}(\varnothing)$ is the vertex set of the subtree with root $\mathcal{A}$. Let $\tau(\mathcal{A}, S)$ be the set of all segmentation trees that lead to a partition $\mathcal{A}$, namely the set of all trees with internal node set $\mathbb{G}$ that satisfies $\mathcal{L}_\mathbb{G} = \mathcal{A}$ (cf. Figure 5.9). The event that Σ'_t is in the state $\mathcal{A} \in \mathbb{P}(S)$ at time $t \geqslant 0$ then obviously translates into

$$\{\Sigma'_t = \mathcal{A}\} = \bigcup_{T^S \in \tau(\mathcal{A},S)} \mathcal{F}_t(T^S).$$

In contrast to the single-crossover case, trees in $\tau(\mathcal{A}, S)$ are no longer uniquely defined via their topology (see Figure 5.9). It is easy to see that the cardinality $\tau_{|\mathcal{A}|}$ of $\tau(\mathcal{A}, S)$ follows the recursion

$$\tau_{|\mathcal{A}|} = \sum_{k=1}^{\lfloor |\mathcal{A}|/2 \rfloor} c_k \binom{|\mathcal{A}|}{k} \tau_k \cdot \tau_{|\mathcal{A}|-k}, \qquad c_k = \begin{cases} 1, & \text{if } k \neq |\mathcal{A}|/2, \\ 1/2, & \text{if } k = |\mathcal{A}|/2, \end{cases}$$

with boundary conditions $\tau_1 = \tau_2 = 1$. The sequence $1, 1, 3, 18, 120, 1080, \ldots$ grows a lot faster than the sequence of Catalan-numbers $1, 1, 2, 5, 14, 42, \ldots$, which describes the number of segmentation trees in the single-crossover case.

The next step would be to define events corresponding to the waiting-times in (5.57). Via Theorem 5.3, one may then obtain an analogue of Corollary 5.1 for arbitrary partitions. Once this is done, it might be helpful to study the continuous-time case first, for which explicit expressions for the tree probabilities in the single-crossover case were obtained without much effort.

Let us finally note that our method does not hinge heavily on the assumption of bi-parental inheritance. The pruning poset is constructed for arbitrary rooted trees. It will be a matter of notation only to drop the restriction to binary trees and thus to also allow for multiple parents as done in [5, 6, 27, 28, 94].

6 Summary and discussion

We presented the forward and backward picture of the dynamics of a haploid population of finite size N evolving under general multi-crossover recombination and verified the formal relationship between the forward and the backward approach via duality.

The forward evolution is described in Chapter 2. For finite populations and continuous time, a Moran model with multi-crossover recombination is introduced, which is strongly related to the Moran model in [8, 9]. The discrete-time counterpart is the Wright-Fisher model with recombination, which is the generalisation of the respective single-crossover version in [10]. As a counterpart to the statements in [8, 10], we observe that the type frequencies of the stochastic models converge to the solution of the deterministic recombination equation as $N \to \infty$ (Thm. 2.1 & Thm. 2.2). In the diffusion limit, where $N \to \infty$ and time is rescaled by N, we obtain convergence to the Wright-Fisher diffusion with general recombination (Thm. 2.3). This convergence result generalises previous results with respect to the number of considered loci.

The ancestral process corresponding to the Moran model or the Wright-Fisher model is described in Chapter 3. Unlike common approaches, we do not start with the diffusion limit (called ancestral recombination graph), but start with the finite model, which allows not only to draw conclusions for the entire parameter space, but also to take different limits efficiently in the end. Instead of considering the full (and complicated) multi-locus ancestral recombination process, we describe a marginalised version in which every locus is sampled in a single individual only. The marginalised ancestral recombination process is defined as a process on the set of partition of sites. The transitions of this partitioning process are described in terms of splitting, coalescence and mixed splitting-coalescence events. An explicit representation of the generator for an arbitrary number of loci is given in continuous time. An analogous expression in discrete time can be obtained in a straightforward way. As $N \to \infty$, only the pure splitting events survive and the partitioning process turns into a process of pure refinements (Prop. 3.1 & Prop. 3.2). In the diffusion limit, splitting and coalescence events act in isolation and the partitioning process converges to a marginalised version of the reduced ancestral recombination graph (Prop. 3.3).

The formal duality between the Moran model with multi-crossover recombination forward in time and the partitioning process backward in time is proved in Chapter 4 (Thm. 4.1). The associated duality function (called sampling function) describes a specific sampling procedure related to sampling with replacement. The sampling function is obtained as the

Möbius inverse of the recombination operator, which represents the respective sampling procedure without replacement. To the best of our knowledge, there is no comparable duality result for finite populations available.
Three main conclusions are derived from the duality result: Together with the marginal ancestral recombination process, it firstly reveals the genealogical structure hidden in the work of Bobrowski et al. [22], who approached the matter by functional-analytic means and forward in time (Rem. 3.1). It allows secondly to write down an explicit and *closed* system of ordinary differential equations for the expected sampling functions (Corol. 4.2). Studying the expected time evolution of a population with the help of this ODE system provides a promising alternative for parameter estimation of recombination rates compared to current methods, which are usually tied to the situation in which a stationary state is reached, see [124] for a general overview, or [38, 48, 61, 74, 95, 124, 129]. The ODE system for the expected sampling functions can, moreover, be translated into an ODE system for expected linkage disequilibria of all orders (Sect. 4.3). Explicit results in the two-site and three-site case show that the expected linkage disequilibria decay exponentially even in the absence of recombination.
The duality equation is thirdly employed to investigate the fixation probabilities of the Moran model by studying the stationary distribution of the partitioning process (Sect. 4.4). A tiny example with three sites leads to the suggestion that the effect of double-crossovers on the long-term behaviour of the genetic composition of a population is rather small. All results in Chapter 4 are expected to hold in discrete time as well.

On the grounds of the duality relation from Chapter 4, we rediscovered in Chapter 5 an explicit solution for the recombination equation in the single-crossover case (first stated in [10]) by studying the probability distribution of the partitioning process as $N \to \infty$ (called segmentation process). It turned out to be useful to represent sets of realisations of the segmentation process that agree on the partial order of events via rooted segmentation trees, whose partial order on the set of vertices encodes the time series of events of the segmentation process. Summing over all tree probabilities yields an explicit expression for the probability distribution of the segmentation process and thereby an explicit expression for the solution of the single-crossover recombination equation. Möbius inversion on a specifically defined poset (called pruning poset) of all rooted forests of segmentation trees helped to decompose the probability of each individual segmentation tree into related probabilities whose explicit expression is known. In discrete time, this required the construction of an auxiliary process, from which the probabilities for the trees are read off. The conceptual proof of the probability distribution of the segmentation process revealed the hidden probabilistic and combinatorial aspects of the solution in [10]. The approach is promising to be easily generalised to allow also for multi-crossover recombination (Sect. 5.4). This way, one may finally be able to state a closed solution for the deterministic recombination equation in the general multi-crossover case in a compact way.

Throughout the thesis, we compared discrete and continuous-time approaches. In most cases, additional dependencies in discrete time impede the analysis and, as for the probability distribution of the segmentation process, long for further tools in order to obtain explicit results.

6.1 Concluding remarks on the model

Let us finally comment on the assumptions of the model with respect to the question whether they may or may not capture the actual biological mechanisms. First of all, it is known that there are dependencies between the positions at which recombination events occur (a phenomenon called interference). Secondly, we know that recombination events are not uniformly distributed along the genome (as assumed in some models [68, 95, 131, 132, 133]) [80, 102, 111]. Starting with the fairly general recombination distribution $\{r_{\mathcal{A}}\}_{\mathcal{A}\in\mathbb{P}_{\leqslant 2}(S)}$, where $r_{\mathcal{A}}$ is the probability of a recombination event with respect to the partition $\mathcal{A}$, $\mathcal{A} \in \mathbb{P}_{\leqslant 2}(S)$, enables to take care of both of these phenomena. Additionally, it enables one to include noncrossover outcomes to the recombination analysis without any further effort. As mentioned in the Introduction, there are two possible outcomes after a recombination event: crossovers and noncrossovers. A crossover refers to a reciprocal exchange of genetic material between maternal and paternal chromosomes. A noncrossover means that genetic information is only transferred from one parental chromosome to the other (Fig. 6.1). Noncrossovers usually affect only a small region of a chromosome and are in general harder to detect than crossovers. Noncrossover hotspots create holes of reduced linkage to their surroundings, which is why Mancera et al. [90] postulate incorporating noncrossovers into linkage analysis. Since noncrossover outcomes, however, agree with double-crossover outcomes that occur nearby, they can easily be included in our model.

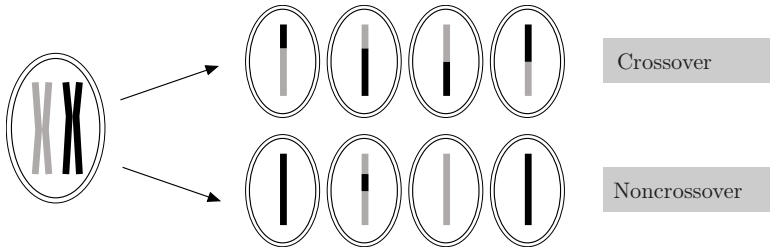


Figure 6.1. Two different end-products of recombination. During meiosis, each cell produces four genetically distinct (haploid) gamete cells. A crossover outcome results from a reciprocal exchange between homologs. A noncrossover outcome is the result of a nonreciprocal exchange between homologs.

The model may further be improved by taking into account the huge sex-differences in recombination rates [29] or by including further evolutionary forces, such as mutation or selection. Including mutation, or at least specific kinds of mutation, is expected to be fairly easy since mutation events can be independently superimposed on the model. Allowing selection, on the other hand, is assumed to be anything but simple.

Bibliography

[1] Aigner, M. *Combinatorial Theory.* Grundlehren der mathematischen Wissenschaften, 234. Springer, Berlin (1979).

[2] Alberts, B., Bray, D., Hopkin, K., Johnson, A., Lewis, J., Raff, M., Roberts, K., AND Walter, P. *Essential Cell Biology.* 4th ed. Garland Science, New York (2014).

[3] Andrews, G.E. *The Theory of Partitions.* Encyclopedia of mathematics and its applications, 2 (Section: Number theory). Addison-Wesley, Reading (1976).

[4] Baake, E. Deterministic and stochastic aspects of single-crossover recombination. *Proceedings of the international Congress of Mathematicians,* Hyderabad, India, Vol. IV (2010), pp. 3037–3053.

[5] Baake, E., AND Baake, M. Haldane linearisation done right: Solving the nonlinear recombination equation the easy way. *Discrete Contin. Dyn. Syst.* 36, 12 (2016), pp. 6645–6656.

[6] Baake, E., Baake, M., AND Salamat, M. The general recombination equation in continuous time and its solution. *Discrete Contin. Dyn. Syst.* 36, 1 (2016), pp. 63–95.

[7] Baake, E., AND Esser, M. Fragmentation process, pruning poset for rooted forests, and Möbius inversion. *arXiv:1702.03173.* (2017).

[8] Baake, E., AND Herms, I. Single-crossover dynamics: finite versus infinite populations. *Bull. Math. Biol.* 70, 2 (2008), pp. 603–624.

[9] Baake, E., AND Hustedt, T. Moment closure in a Moran model with recombination. *Markov Process. Relat. Fields* 17 (2011), pp. 429–446.

[10] Baake, E., AND VON Wangenheim, U. Single-crossover recombination and ancestral recombination trees. *J. Math. Biol.* 68, 6 (2014), pp. 1371–1402.

[11] Baake, M. Recombination semigroups on measure spaces. *Monatsh. Math.* 146, 4 (2005), pp. 267–278.

[12] Baake, M., AND Baake, E. An exactly solved model for mutation, recombination and selection. *Canad. J. Math.* 55 (2003), pp. 3–41.

[13] Barton, N.H., AND Turelli, M. Natural and sexual selection on many loci. *Genetics* 127, 1 (1991), pp. 229–255.

[14] Bennett, J.H. On the theory of random mating. *Ann. Eugen.* 17, 1 (1952), pp. 311–317.

[15] Berestycki, N. Recent progress in coalescent theory. *Ensaios Mat.* 16, 1 (2009), pp. 1–193.

[16] Berge, C. *Principles of Combinatorics.* Mathematics in science and engineering, 72. Academic Press, New York (1971).

[17] Bhaskar, A., AND Song, Y.S. Closed-form asymptotic sampling distributions under the coalescent with recombination for an arbitrary number of loci. *Adv. Appl. Prob.* 44, 2 (2012), pp. 391–407.

[18] Birkhoff, G. *Lattice Theory*. 3rd ed. Colloquium publications, 25. American Mathematical Society, Providence (1967).

[19] Birkner, M., Blath, J., AND Eldon, B. An ancestral recombination graph for diploid populations with skewed offspring distribution. *Genetics* 193, 1 (2013), pp. 255–290.

[20] Bishop, D.K., AND Zickler, D. Early decision: meiotic crossover interference prior to stable strand exchange and synapsis. *Cell* 117 (2004), pp. 9–15.

[21] Bobrowski, A., AND Kimmel, M. A random evolution related to a Fisher–Wright-Moran model with mutation, recombination and drift. *Math. Methods Appl. Sci.* 26, 18 (2003), pp. 1587–1599.

[22] Bobrowski, A., Wojdyła, T., AND Kimmel, M. Asymptotic behavior of a Moran model with mutations, drift and recombination among multiple loci. *J. Math. Biol.* 61, 3 (2010), pp. 455–473.

[23] Bürger, R. *The Mathematical Theory of Selection, Recombination, and Mutation*. Wiley series in mathematical and computational biology. John Wiley & Sons, Chichester (2000).

[24] Bürger, Reinhard Evolution of genetic variability and the advantage of sex and recombination in changing environments. *Genetics* 153, 2 (1999), pp. 1055–1069.

[25] Christiansen, F.B. *Population Genetics of Multiple Loci*. Wiley series in mathematical and computational biology. John Wiley & Sons, Chichester (2000).

[26] Clifford, P., AND Sudbury, A. A sample path proof of the duality for stochastically monotone Markov processes. *Ann. Prob.* 13 (1985), pp. 558–565.

[27] Dawson, K.J. The decay of linkage disequilibrium under random union of gametes: how to calculate Bennett's principal components. *Theor. Pop. Biol.* 58, 1 (2000), pp. 1–20.

[28] Dawson, K.J. The evolution of a population under recombination: how to linearise the dynamics. *Linear Algebra Appl.* 348, 1-3 (2002), pp. 115–137.

[29] De Boer, E., Jasin, M., AND Keeney, S. Local and sex-specific biases in crossover vs. noncrossover outcomes at meiotic recombination hotspots in mouse. *Genes Dev.* 29 (2015), pp. 1721–1733.

[30] Depperschmidt, A., Pardoux, É., AND Pfaffelhuber, P. A mixing tree-valued process arising under neutral evolution with recombination. *Electron. J. Probab.* 20, 94 (2015), pp. 1–22.

[31] Donnellan, T. *Lattice Theory*. 1st ed. Pergamon Press, Oxford (1968).

[32] Donnelly, P. Dual processes in population-genetics. *Lecture Notes in Math.* 1212 (1986), pp. 94–105.

[33] Donnelly, P., AND Kurtz, T.G. A countable representation of the Fleming-Viot measure-valued diffusion. *Ann. Prob.* 24, 2 (1996), pp. 698–742.

[34] Drmota, M. *Random Trees: An Interplay between Combinatorics and Probability*. Springer, Wien (2009).

[35] Drmota, M., Iksanov, A., Möhle, M., AND Rösler, U. A limiting distribution for the number of cuts needed to isolate the root of a random recursive tree. *Random Struc. Algor.* 34, 3 (2009), pp. 319–336.

[36] Durrett, R. *Stochastic Calculus*. Probability and stochastics series. CRC Press, Boca Raton (1996).

[37] Durrett, R. *Probability Models for DNA Sequence Evolution*. 2nd ed. Probability and its applications. Springer, New York (2008).

[38] Dutheil, J.Y., Ganapathy, G., Hobolth, A., Mailund, T., Uyenoyama, M.K., AND Schierup, M.H. Ancestral population genomics: the coalescent hidden Markov model approach. *Genetics* 183, 1 (2009), pp. 259–274.

[39] Dyson, F.J. Statistical theory of energy levels of complex systems III. *J. Math. Phys.* 3, 1 (1962), pp. 166–175.

[40] Esser, M., Probst, S., AND Baake, E. Partitioning, duality, and linkage disequilibria in the Moran model with recombination. *J. Math. Biol.* 73, 1 (2016), pp. 161–197.

[41] Etheridge, A. *Some Mathematical Models from Population Genetics*. Lecture notes in mathematics, 2012 (Subseries: École d'Eté de Probabilités de Saint-Flour). Springer, Heidelberg (2011).

[42] Ethier, S.N., AND Griffiths, R.C. On the two-locus sampling distribution. *J. Math. Biol.* 29, 2 (1990), pp. 131–159.

[43] Ethier, S.N., AND Griffiths, R.C. The neutral two-locus model as a measure-valued diffusion. *Adv. Appl. Prob.* 22, 4 (1990), pp. 773–786.

[44] Ethier, S.N., AND Kurtz, T.G. *Markov Processes: Characterization and Convergence*. Reprint 2005. John Wiley & Sons, Hoboken (1986).

[45] Ethier, S.N., AND Kurtz, T.G. Fleming-Viot processes in population genetics. *SIAM J. Control Optim.* 31, 2 (1993), pp. 345–386.

[46] Ethier, S.N., AND Nagylaki, T. Diffusion approximations of the two-locus Wright-Fisher model. *J. Math. Biol.* 27, 1 (1989), pp. 17–28.

[47] Ewens, W.J. *Mathematical Population Genetics*. 2nd ed. Interdisciplinary applied mathematics, 27. Springer, New York (2004).

[48] Fearnhead, P., AND Donnelly, P. Estimating recombination rates from population genetic data. *Genetics* 159, 3 (2001), pp. 1299–1318.

[49] Filippo, J.S., Sung, P., AND Klein, H. Mechanism of eukaryotic homologous recombination. *Annu. Rev. Biochem.* 77 (2008), pp. 229–257.

[50] Foissy, L. The infinitesimal Hopf algebra and the poset of planar forests. *J. Algebraic Combin.* 30, 3 (2009), pp. 277–309.

[51] FOR Biotechnology Information (US), National Center *Genes and Disease*. `http://www.ncbi.nlm.nih.gov/books/NBK22266/`. [Online; accessed 19-May-2016]. (1998).

[52] Frucht, R., AND Rota, G.-C. La función de Möbius para particiones de un conjunto. *Scientia* 122 (1963), pp. 111–115.

[53] Geiringer, H. On the probability theory of linkage in Mendelian heredity. *Ann. Math. Stat.* 15, 1 (1944), pp. 25–57.

[54] Golding, G.B. The sampling distribution of linkage disequilibrium. *Genetics* 108, 1 (1984), pp. 257–274.

[55] Gorelick, R., AND Laubichler, M.D. Decomposing multilocus linkage disequilibrium. *Genetics* 166, 3 (2004), pp. 1581–1583.

[56] Graham, R.L., Knuth, D.E., AND Patashnik, O. *Concrete Mathematics: a Foundation for Computer Science*. 2nd ed., 22. print. Addison-Wesley, Upper Saddle River (2009).

[57] Grätzer, G.A. *General Lattice Theory*. 2nd ed., Birkhäuser, Basel (2003).

[58] Griffiths, R.C. Neutral two-locus multiple allele models with recombination. *Theor. Pop. Biol.* 19, 2 (1997), pp. 169–186.

[59] Griffiths, R.C., Jenkins, P.A., AND Lessard, S. A coalescent dual process for a Wright-Fisher diffusion with recombination and its application to haplotype partitioning. *Theor. Pop. Biol.* 112 (2016), pp. 126–138.

[60] Griffiths, R.C., AND Marjoram, R. An ancestral recombination graph. In: *Progress in Population Genetics and Human Evolution*. Springer, Berlin (1996), pp. 257–270.

[61] Griffiths, R.C., AND Marjoram, R. Ancestral inference from samples of DNA sequences with recombination. *J. Comput. Biol.* 3, 4 (1996), pp. 479–502.

[62] Hall, P. The Eulerian functions of a group. *Quart. J. Math. Oxford Ser.* 1 (1936), pp. 134–151.

[63] Hardy, G.H., AND Wright, E.M. *An Introduction to the Theory of Numbers*. 6th ed. Oxford mathematics. Oxford University Press, Oxford (2008).

[64] Hastings, A. Linkage disequilibrium, selection, and recombination at three loci. *Genetics* 106, 1 (1984), pp. 153–164.

[65] Hein, J., Schierup, M.H., AND Wiuf, C. *Gene Genealogies, Variation and Evolution: A Primer in Coalescent Theory*. Reprint 2006. Oxford University Press, Oxford (2005).

[66] Hill, W.G. Disequilibrium among several linked neutral genes in finite populations I. Mean changes in disequilibrium. *Theor. Pop. Biol.* 5, 3 (1974), pp. 366–392.

[67] Hillers, K.J. Crossover interference. *Curr. Biol.* 14 (2004), R1036–R1037.

[68] Hudson, R.R. Properties of a neutral allele model with intragenetic recombination. *Theor. Pop. Biol.* 23, 2 (1983), pp. 183–201.

[69] Hudson, R.R., AND Kaplan, N.L. Statistical properties of the number of recombination events in the history of a sample of DNA sequences. *Genetics* 111, 1 (1985), pp. 147–164.

[70] Hudson, R.R., AND Kaplan, N.L. The coalescent process in models with selection and recombination. *Genetics* 120, 3 (1988), pp. 831–840.

[71] Jansen, S., AND Kurt, N. On the notion(s) of duality for Markov processes. *Prob. Surveys* 11 (2014), pp. 59–120.

[72] Janson, S. Random cutting and records in deterministic and random trees. *Random Struct. Alg.* 29, 2 (2006), pp. 139–179.

[73] Jenkins, P.A., Fearnhead, P., AND Song, Y.S. Tractable stochastic models of evolution for loosely linked loci. *Electron. J. Probab.* 20 (2015), pp. 1–26.

[74] Jenkins, P.A., AND Griffiths, R. Inference from samples of DNA sequences using a two-locus model. *J. Comp. Biol.* 18, 1 (2011), pp. 109–127.

[75] Jenkins, P.A., AND Song, Y.S. Closed-form two-locus sampling distributions: accuracy and universality. *Genetics* 183, 3 (2009), pp. 1087–1103.

[76] Jenkins, P.A., AND Song, Y.S. An asymptotic sampling formula for the coalescent with recombination. *Ann. Appl. Prob.* 20, 3 (2010), pp. 1005–1028.

[77] Jenkins, P.A., AND Song, Y.S. Padé approximants and exact two-locus sampling distribution. *Ann. Appl. Prob.* 22, 2 (2012), pp. 576–607.

[78] Jennings, H.S. The numerical results of diverse systems of breeding, with respect to two pairs of characters, linked or independent, with special relation to the effects of linkage. *Genetics* 2, 1 (1917), pp. 97–154.

[79] Kaplan, N., AND Hudson, R.R. The use of sample genealogies for studying a selectively neutral m-loci model with recombination. *Theor. Pop. Biol.* 28, 3 (1985), pp. 382–396.

[80] Kauppi, L., Jeffreys, A.J., AND Keeney, S. Where the crossovers are: recombination distributions in mammals. *Nat. Rev. Genet.* 5, 6 (2004), pp. 413–424.

[81] Kingman, J.F.C On the genealogy of large populations. *J. Appl. Prob.* 19 (1982), pp. 27–43.

[82] Kingman, J.F.C. The coalescent. *Stoch. Process. Appl.* 13, 3 (1982), pp. 235–248.

[83] Krone, S.M., AND Neuhauser, C. Ancestral processes with selection. *Theor. Pop. Biol.* 51, 3 (1997), pp. 210–237.

[84] Lawrie, N.M., Tease, C., AND Hulten, M.A. Chiasma frequency, distribution and inference maps of mouse autosomes. *Chromosoma* 104 (1995), pp. 308–314.

[85] Lenormand, T., AND Otto, S.P. The evolution of recombination in a heterogeneous environment. *Genetics* 156, 1 (2000), pp. 423–438.

[86] Lewontin, R.C. The interaction of selection and linkage I. General considerations; heterotic models. *Genetics* 49, 1 (1964), pp. 49–67.

[87] Lewontin, R.C., AND Kojima, K. The evolutionary dynamics of complex polymorphisms. *Evolution* 14, 4 (1960), pp. 458–472.

[88] Liggett, T.M. *Interacting Particle Systems*. Reprint 2005. Classics in mathematics. Springer, Berlin (1985).

[89] Ljubič, J.I. *Mathematical Structures in Population Genetics*. Biomathematics, 22. Springer, Berlin (1992).

[90] Mancera, E., Bourgon, R., Brozzi, A., Huber, W., AND Steinmetz, L.M. High-resolution mapping of meiotic crossovers and non-crossovers in yeast. *Nature* 454 (2008), pp. 479–485.

[91] Mano, S. Duality between the two-locus Wright-Fisher diffusion model and the ancestral process with recombination. *J. Appl. Prob.* 50, 1 (2013), pp. 256–271.

[92] Martin, G., Otto, S.P., AND Lenormand, T. Selection for recombination in structured populations. *Genetics* 172, 1 (2006), pp. 593–609.

[93] Martinez, S. A probabilistic analysis of a discrete-time evolution in recombination. *arXiv preprint: 1603.07201* (2016).

[94] Martinez, S. A probabilistic analysis of a discrete-time evolution in recombination II. (On partitions). *arXiv preprint: 1604.05124* (2016).

[95] McVean, G.A.T., AND Cardin, N.J. Approximating the coalescent with recombination. *Phil. Trans. R. Soc. B* 360, 1459 (2005), pp. 1387–1393.

[96] Mehta, M.L. *Random Matrices*. 2nd ed. Academic Press, New York (1991).

[97] Meir, A, AND Moon, J.W. Cutting down random trees. *J. Austral. Math. Soc.* 11, 3 (1970), pp. 313–324.

[98] Möbius, A.F. Über eine besondere Art von Umkehrung der Reihen. *J. reine angew. Math.* 9 (1832), pp. 105–123.

[99] Möhle, M. The concept of duality and applications to Markov processes arising in neutral population genetics models. *Bernoulli* 5, 5 (1999), pp. 761–777.

[100] Möhle, M. Forward and backward diffusion approximations for haploid exchangeable population models. *Stoch. Proc. Appl.* 95 (2001), pp. 133–149.

[101] Morgan, T.H. Random segregation versus coupling in Mendelian inheritance. *Science* (1911), pp. 384–384.

[102] Myers, S., Bottolo, L., Freeman, C., McVean, G., AND Donnelly, P. A fine-scale map of recombination rates and hotspots across the human genome. *Science* 310 (2005), pp. 321–324.

[103] Ohta, T., AND Kimura, M. Linkage disequilibrium at steady state determined by random genetic drift and recurrent mutation. *Genetics* 63, 1 (1969), pp. 229–238.

[104] Ohta, T., AND Kimura, M. Linkage disequilibrium due to random genetic drift. *Genet. Res.* 13, 1 (1969), pp. 47–55.

[105] Oksendal, B. *Stochastic Differential Equations: an Introduction with Applications*. 6th ed., corr. 4. print. Universitext. Springer, Heidelberg (2007).

[106] Otto, S.P., and Barton, N.H. Selection for recombination in small populations. *Evolution* 55, 10 (2001), pp. 1921–1931.

[107] Otto, S.P., and Lenormand, T. Resolving the paradox of sex and recombination. *Nat. Rev. Genet.* 3, 4 (2002), pp. 252–261.

[108] Page, S.L., and Hawley, R.S. Chromosome choreography: the meiotic ballet. *Science* 301 (2003), pp. 785–789.

[109] Panholzer, A. Cutting down very simple trees. *Quaest. Math.* 29, 2 (2006), pp. 211–227.

[110] Pâques, F., and Haber, J.E. Multiple pathways of recombination induced by double-strand breaks on *Saccharomyces cerevisiae*. *Microbiol. Mol. Biol. Rev.* 63 (1999), pp. 349–404.

[111] Petes, T.D. Meiotic recombination hot spots and cold spots. *Nat. Rev. Genet.* 2 (2001), pp. 360–369.

[112] Polanska, J., and Kimmel, M. A simple model of linkage disequilibrium and genetic drift in human genomic SNPs: Importance of demography and SNP age. *Hum. Hered.* 60, 4 (2005), pp. 181–195.

[113] Reed, M. Algebraic structure of genetic inheritance. *Bull. Math. Biol.* 34, 2 (1997), pp. 107–130.

[114] Robbins, R.B. Some applications of mathematics to breeding problems III. *Genetics* 3, 4 (1918), pp. 375–389.

[115] Roeder, G.S. Meiotic chromosomes: it takes two to tango. *Genes Dev.* 11 (1997), pp. 2600–2621.

[116] Rosenberg, N.A., and Nordborg, M. Genealogical trees, coalescent theory and the analysis of genetic polymorphisms. *Nat. Rev. Genet.* 3 (2002), pp. 380–390.

[117] Rota, G.-C. On the foundations of combinatorial theory I. Theory of Möbius functions. *Z. Wahrscheinlichkeitstheorie* 2, 4 (1964), pp. 340–368.

[118] Sansam, C.L., and Pezza, R.J. Connecting by breaking and repairing: mechanisms of DNA strand exchange in meiotic recombination. *FEBS J.* 282 (2015), pp. 2444–2457.

[119] Schützenberger, M.-P. *Contribution aux applications statistiques de la théorie de l'information*. Vol. 3. Pupl. Inst. de Stat. Univ. Paris, (1954).

[120] Slatkin, M. On treating the chromosome as the unit of selection. *Genetics* 72, 1 (1972), pp. 157–168.

[121] Song, Y.S., and Song, J.S. Analytic computation of the expectation of the linkage disequilibrium coefficient r^2. *Theor. Pop. Biol.* 71, 1 (2007), pp. 49–60.

[122] Staab, P.R., Zhu, S., Metzler, D., and Lunter, G. SCRM: efficiently simulating along sequences using the approximated coalescent with recombination. *Bioinformatics* 31, 10 (2015), pp. 1680–1682.

[123] Stanley, R.P. *Enumerative Combinatorics*. Reprint. Vol. I. Cambridge studies in advanced mathematics, 49. Cambridge University Press, Cambridge (2002).

[124] Stumpf, M.P.H., and McVean, G.A.T. Estimating recombination rates from population-genetic data. *Nat. Rev. Genet.* 4 (2003), pp. 959–968.

[125] Von Wangenheim, U., Baake, E., and Baake, M. Single-crossover recombination in discrete time. *J. Math. Biol.* 60, 5 (2010), pp. 727–760.

[126] Wakeley, J. Recent trends in population genetics: More data! More math! Simple models? *J. Hered.* 95, 5 (2004), pp. 397–405.

[127] Wakeley, J. The limits of theoretical population genetics. *Genetics* 169, 1 (2005), pp. 1–7.

[128] Wakeley, J. *Coalescent Theory: An Introduction*. Roberts and Co., Greenwood Village (2009).

[129] Wang, Y., AND Rannala, B. Bayesian inference of fine-scale recombination rates using population genomic data. *Phil. Trans. R. Soc. B* 363 (2008), pp. 3921–3930.

[130] Weisner, L. Abstract theory of inversion of finite series. *Trans. Amer. math. Soc.* 38, 3 (1935), pp. 474–484.

[131] Wiuf, C., AND Hein, J. On the number of ancestors to a DNA sequence. *Genetics* 147, 3 (1997), pp. 1459–1468.

[132] Wiuf, C., AND Hein, J. Recombination as a point process along sequences. *Theor. Pop. Biol.* 55, 3 (1999), pp. 248–259.

[133] Wiuf, C., AND Hein, J. The ancestry of a sample of sequences subject to recombination. *Genetics* 151, 3 (1999), pp. 1217–1228.