The Moran model of population genetics: case studies with recombination and selection

Dissertation
zur Erlangung des Doktorgrades
der Mathematik

vorgelegt an der Fakultät für Mathematik
der Universität Bielefeld

von

Dipl.-Math. Thiemo Hustedt

Bielefeld im August 2012

Gedruckt auf alterungsbeständigem Papier nach ISO 9706.

Contents

Chapter 1

Introduction

In recent years, the processes of population genetics, which describe the genetic structure of populations under the influence of evolutionary forces such as mutation, selection, recombination, migration, and genetic drift, have been a rich source of fascinating probabilistic problems. More precisely, the dynamics is often well understood in the limit of infinite population size, where a law of large numbers leads to a deterministic description (in terms of discrete dynamical systems or differential equations), but great challenges ensue if the population is finite, in particular if there is interaction between individuals, such as competition (selection) or recombination (the combination of genetic material of two parents into the 'mixed' genetic type of an offspring); see [12, 17, 20].

For finite populations, there are mainly two classes of models in population genetics. One important class are branching processes where individuals reproduce independently and thus the population size varies. See [24] for an introductory overview.

At the other extreme, one may consider models where the population size is fixed. On the one hand, the Wright-Fisher model where the successive generation is drawn out of the preceding, thus models species with non-overlapping generations [17]. On the other hand, the Moran model, first formulated by Moran [42] in 1958. In its simplest formulation, one considers a population of N individuals. Each individual is either of type 0 or 1. At each time step, one individual is randomly chosen and replaces another randomly chosen one. Of course, one may extend this simple model by further evolutionary forces.

In this thesis, we consider a Moran model with recombination and mutation and a Moran model with selection and mutation.

Recombination

Recombination means the exchange and rearrangement of genetic material at sexual reproduction. It is one of the major sources of genetic variability. This induces interaction between the individuals in the model.

Interactions usually make the infinite-population model nonlinear and, often, already difficult enough to treat. In the corresponding stochastic model, they are reflected by transition rates (or probabilities) that depend nonlinearly on the current state of the system and often result in processes whose treatment provides enormous challenges. Even the relationship between the stochastic process and its deterministic counterpart, the infinite-population model is usually unclear (apart from the infinite population limit). In particular, the expectation of the stochastic process is, usually, not given by the corresponding deterministic dynamics - in general, such coincidence is reserved for populations of individuals that evolve independently.

In most processes of population genetics with interaction, even the analysis of the expectation is difficult. Its dynamics does, usually, not only depend on the current expectation, but on higher moments, whose change, in turn, depends on even higher moments. Formulating this hierarchy of dependencies is a common approach for stochastic processes arising in various applications in physics, chemistry, and biology [39, 38, 14]. Usually, this hierarchy continues indefinitely (it does not 'close'); to extract at least an approximation to the (lower) moments of interest, some method of 'moment closure' must be employed (in the simplest case, a truncation) [14]. Most approaches in this field are purely heuristic. One concept was developed by Levermore [39] where the moment closure is chosen such that the entropy is maximised.

A recent paper that provides a good overview on moment closure in the context of population dynamics is from Hausken [26] in which he systemizes the common approaches and shows that the approximations indeed improve when the 'power' of the moment closure increases.

The corresponding deterministic systems (that arise through a law of large numbers) are also often tackled via systems of moments or cumulants, see [12, Ch. V.4] for an overview.

Models of recombination take a special role between linear and nonlinear models. Although there is abundant interaction and hence nonlinearity, the deterministic system that describes the frequencies of all possible (geno)types may be (exactly) transformed into a linear one by embedding it into a higher-dimensional space (more explicitly, by adding further components that correspond to products of type frequencies). This method is known as Haldane linearisation [41]. The underlying linear structure even allows a diagonalisation and explicit solution, see [52] and references therein. In certain important special cases (notably, in so-

called single-crossover dynamics in continuous time), this solution is surprisingly simple and immediately plausible [3, 8].

Elucidating underlying linear structures in the corresponding stochastic system (more precisely, in the Moran model with recombination) has only started very recently whereas a lot of literature is available for the corresponding retrospective processes, that means backward in time, as the ancestral recombination graph. See [28] and references therein for a good introduction.

In the aforementioned single-crossover case, Bobrowski et al. [10] analysed the asymptotic behaviour in the presence of mutation. They obtain the surprising result that the stationary distribution is hardly affected by recombination rather than mutation and resampling.

Baake and Herms [6] observed that the expected type frequencies in the finite system (but without genetic drift) follow those in the deterministic model; this could be explained by the (conditional) independence of certain marginalised processes that appear as 'subsystems' of the stochastic model. This and other results now lead to the question whether in the general recombination scheme (i.e., not restricted to single crossovers) the dynamics of the expectations may be embedded into a higher but finite dimensional space, such that they are given by a finite system of differential equations. Is there an equivalent of Haldane linearisation in the sense of moments?

In Chapter 2 we will address these questions in the framework of the Moran model with recombination and mutation. In particular, we will show that the system of moments closes here after a finite number of steps, without any need for approximations, as long as there is no genetic drift. This may be considered as a stochastic analogue of Haldane linearisation. First, we present a short excursus about recombination in the field of genetic algebras and conduct Haldane linearisation in this context which originally motivated us to investigate the Moran model with recombination with respect to moment closure.

Selection

In Chapter 3, we investigate problems that arise in the Moran model with selection and mutation. Whereas in Chapter 2, our focus was on the behaviour of our model in forward time, here we concentrate on properties in backward time.

From a biological point of view, interesting questions are such as how strong is selection, what is a good choice for the parameters or how does selection interact with genetic drift (for example, a favoured mutation may get lost due to random individual reproduction nevertheless in expectation its frequency should grow exponentially [50])?

Of course, these effects shall be measured and quantified with data. Typically, one has some data of the current population but no chance to get data from all

generations needed for an evolutionarily relevant time span. Thus, one has to reconstruct the ancestry of a sample of the population probabilistically.

This problem is tackled by coalescent theory. Typical questions in coalescent theory are [5]:

- What is the ancestry of the sample?
- What is a good choice of the genetic parameters governing the process?
- When does the genealogical tree reach its root, the most recent common ancestor? What is its type?

It is often more convenient in population genetics to pass to the diffusion limit. By a proper scaling of parameters and acceleration of time, the Moran model converges weakly to a diffusion limit if the number of individuals tends to infinity [37]. The importance of this limit comes from the fact that on the one hand random fluctuations in the population are kept and thus it is a valid approximation for a finite population and on the other hand these diffusion processes are often mathematically easier to analyse. See [17, Ch. 7] for an introduction. The disadavantage of this approach is that the graphical representation, as it is naturally provided by the Moran model, gets lost.

Kingman [34, 35] 1981 described a method to construct genealogies in the neutral model, that means without selection. For a given sample without types, drawn from the diffusion limit, one considers their behaviour in the past where reproduction events lead to the merging of lines. One finally reaches a state with one single lineage, determines its type from the stationary distribution of the forward process and imposes the mutation process on the lines in forward time. The simplicity of this procedure hinges on the fact that in the neutral model reproduction of single individuals is independent of their types and of the composition of the background population. In the last thirty years, a lot of literature came up to this neutral model. For an overview see [17].

If selection is incorporated, this simple procedure is no more applicable. For a very long period of time it seemed intractable to construct a coalescent for models with selection. Not before 1997, Krone and Neuhauser [43] provided a procedure by which genealogies under the influence of selection may be constructed, the so-called ancestral selection graph. They overcome the mentioned problems by the permission of branching events in backward time that means, new lines may emerge. So, a large graph is constructed, out of which the true genealogical tree must be extracted, the remaining lines called virtual lines.

Another approach was first suggested by Kaplan, Darden and Hudson [31] in 1988. The idea is to divide the population into groups (so called 'demes') of individuals of the same type and then model their genealogy by the structured coalescent. In

this model, mutation corresponds to migration between the demes [9]. In 2004, Barton, Etheridge, Sturm [9] formalised this approach and proved existence of this process. They underline that simulations based on the structured coalescent are "essentially more efficient" compared to those based on the ancestral selection graph and that their approach applies two "very general forms of selection".

A natural question that comes up in this field is the distribution of the most recent common ancestor in these models. Whereas in the neutral model its distribution is simply given by the stationary distribution of the forward-process, in models with selection this should no more be true. It is a sensible conjecture that it is more likely that it is favoured under selection [21].

In 2002, Fearnhead [21] calculated its distribution with the help of the construction of the common ancestor process. The construction of this process is similar to the ancestral selection graph. The key idea is to look at the ancestry of a single lineage. Compared to the ancestral selection graph an essential simplification is possible due to the observation that only unfit virtual lines may affect the genealogy of this individual. In the deep past, its line must coincide with the line of the common ancestor of the whole population. Thus, its stationary distribution is the distribution of the most recent common ancestor in the ancestral selection graph. Fearnhead gives an explicit distribution in the case of purifying selection and proves its stationarity by a rather technical verification of the stationarity condition.

In 2007, Taylor [50] also considers a common ancestor process starting from the structured coalescent approach. He first considers the retrospective process which is the structured coalescent for sample size 1 and identifies the common ancestor process as its time-reversed process. To this end, he needs the common ancestor distribution which is the stationary distribution of both the forward and backward in time process. For this, he defines the conditional probability $h(x)$ which is the probability that the ancestor of the current population is fit, if the frequency of fit individuals in the current population is x. It is no surprise that the distributions from Fearnhead's and Taylor's approach coincide.

Unfortunately, their solutions are no more connected to the particle picture of the Moran model. Nevertheless, there is some remarkable interest in models that permit particle interpretations. A particle picture from the diffusion limit was provided by the look-down process in 1996 [15] for the neutral case. Here, each particle is given a level, and only particles of lower levels may be replaced at birth events. From this process, the former diffusion limit may be regained if the number of particles tends to infinity. A further convenient property is, that the genealogy of these particles is explicit and the common ancestor is the particle of the highest level. For the nonneutral model the construction of the look-down process becomes much more difficult and especially the genealogies can no more be recovered that easily [16]. See [18, Ch. 5] for an overview.

In Chapter 3 we connect the common ancestor distribution directly to the Moran model with the aim to enlighten the quantities in thei formula and to provide a particle picture behind this distribution. After a short introduction into the construction of the common ancestor process and a presentation of the results of Fearnhead and Taylor, we derive a boundary value problem that describes the distribution of the common ancestor via a first step argument directly from the Moran model without a loop way over diffusion theory or the construction of a sophisticated process as it was done by Taylor [50] and Fearnhead [21]. We are also able to solve this problem directly from the Moran model. This provides some better understanding of the process and enables us to present a new particle picture comparable to the look down process in the case of no mutation.

Chapter 2

Moment closure in a Moran model with recombination

The main results of this chapter have been published in [7], it is joint work with Ellen Baake. We follow the presentation in [7]. Beyond this, we undergo an excursus about genetic algebras. The occupation on this field motivated us to apply the moment approach to the Moran model with recombination.

2.1 Excursus about genetic algebras

In the first part of this section we give a short introduction into the theory of genetic algebras and Haldane linearisation. Detailed presentations of this topic may be found in [54, 40, 46]. Then we apply this theory to the dynamics of recombination.

Consider a collection of gametes $a_1, \ldots, a_n$. The span of these over the real numbers may be considered as a vector space together with the canonical addition and scalar multiplication. The convex combinations of these gametes may naturally be interpreted as populations with the weights as the frequencies of the corresponding gametes or types.

When two types mate, some other types may emerge. We express the mating by a multiplication:

$$a_i a_j = \sum_{k=1}^{n} \gamma_{ijk} a_k$$

with $\gamma_{ijk} \geq 0$, $\sum_{k=1}^{n} \gamma_{ijk} = 1$, being the probability that the offspring of this mating event is a_k. By this reasoning, it is clear that algebras naturally arise in the description of genetic inheritance. The standard example is the Mendelian algebra:

Example 2.1. [Gametic algebra with recombination [54, Ex.1.3]] We consider a population of gametes. Each gamete has two loci which may be occupied either by allele A or a, and B or b, respectively. Thus, four types of gametes are present:

$$a_1 := AB, \quad a_2 := Ab, \quad a_3 := aB, \quad a_4 = ab.$$

We consider these four gametes as free vectors over $\mathbb{R}$. The multiplication is explained via the following table

	a_1	a_2	a_3	a_4
a_1	a_1	$\frac{1}{2}(a_1 + a_2)$	$\frac{1}{2}(a_1 + a_3)$	$\frac{1}{2}(a_1 + a_4) - \frac{1}{2}\theta d$
a_2		a_2	$\frac{1}{2}(a_2 + a_3) + \frac{1}{2}\theta d$	$\frac{1}{2}(a_2 + a_4)$
a_3			a_3	$\frac{1}{2}(a_3 + a_4)$
a_4				a_4

where $d := a_1 - a_2 - a_3 + a_4$.

These multiplication rules reflect the following heredity scheme: if two gametes of the same type mate, the offspring will be of this type, too. If type a_1 mates with a_2, then the offspring will be of type a_1 or a_2 with the same frequency. These rules reflect simple Mendelian inheritance. A special situation is on hand for the pairing of a_1 and a_4, or a_2 and a_3, respectively. Here, recombination may interfere. Thus, with probability θ the gametes may exchange their genetic material. In the case of types a_1 and a_4 this means that the resulting gametes will be of type a_2 and a_3. So, with probability $1 - \theta$, the resulting offspring will be a_1 and a_4 with the same frequency and with probability θ it will be a_2 and a_3 with the same frequency.

In the table, we omitted the entries below the diagonal due to symmetry. It is clear that for biological reasons the multiplication should be commutative.

Hereby, we consider a convex combination of the four basis vectors as a population and θ may be interpreted as the recombination probability. The product of a population with itself is then again a population and denotes the reproductive step of the population. □

The most general definition of algebras in this field is the algebra with genetic realization or stochastic algebra.

Definition 2.2. *An algebra is called* algebra with genetic *or* stochastic *if it has a basis* $a_1, \ldots, a_n$ *such that the multiplication is given by*

$$a_i a_j = \sum_{k=1}^{n} \gamma_{ijk} a_k,$$

where the multiplication constants γ_{ijk} *fulfill the equations*

$$\gamma_{ijk} \geq 0, \quad i, j, k = 1, \ldots, n$$

and

$$\sum_{k=1}^{n} \gamma_{ijk} = 1, \quad i, j = 1, \ldots, n.$$

This basis is then called natural basis.

The algebra in Example 2.1 is stochastic.

Consider three populations P, Q and R. From a biological point of view it is obvious that it makes a difference whether P and Q mate first and their product with R or whether the product of Q and R mates with P. Mathematically this means that algebras in genetics are usually not associative.

Likewise, it is obvious that it should not matter whether P mates Q or Q mates P. Thus, algebras in genetics are usually commutative. Compare with our example.

Definition 2.3. *An algebra* $\mathcal{A}$ *over a field* k *is called* baric *if it admits a non-trivial algebra homomorphism*

$$\omega : \mathcal{A} \to k,$$

which is called the weight function.

For an algebra with genetic realisation in the natural basis $a_1, \ldots, a_n$, define an algebra homomorphism $\omega : \mathcal{A} \to k$ by $\omega(a_j) = 1$. Thus, algebras with genetic realisation are baric algebras. The reverse is not true in general.

Algebras with genetic realisation turn out to be very general objects. Schafer [48] gave a definition for genetic algebras, called *Schafer-genetic*. This definition is very useful because on the one hand most algebras with an application in genetics fulfill the conditions, and on the other hand these have some useful properties. In our context, it is an important fact that Schafer-genetic algebras are baric algebras whose weight function is unique [41]. Thus, the set of populations, which is the pre-image of 1 under the weight function, is well-defined.

Gonshor gave a definition which is equivalent to Schafer's[1] but is of more practical use because the conditions are usually easier to check.

[1] Woerz-Busekroes [54] points out that this assertion is only true if the underlying field is algebraically closed. Otherwise there are examples of Schafer-genetic algebras which are not Gonshor-genetic. In account of this subtlety, Woerz-Busekroes gives a slightly different definition than Gonshor.

Definition 2.4. [23, 46] *A commutative finite dimensional algebra $\mathcal{A}$ is called* Gonshor-genetic *if there exists a basis $\{a_0, a_1, \ldots, a_n\}$ with*

$$a_i a_j = \sum_{k=0}^{n} \gamma_{ijk} a_k,$$

such that the multiplication constants satisfy

$$\begin{aligned} \gamma_{000} &= 1, \\ \gamma_{0jk} &= 0, \quad \textit{for } k < j, \\ \gamma_{ijk} &= 0, \quad \textit{for } i, j > 0 \textit{ and } k \leq \max(i, j). \end{aligned}$$

Such a basis is called canonical basis for $\mathcal{A}$.

Example 2.5. We change the basis by the matrix

$$\begin{pmatrix} 1 & 0 & 0 & 0 \\ 1 & -1 & 0 & 0 \\ 1 & 0 & -1 & 0 \\ 1 & -1 & -1 & 1 \end{pmatrix}$$

this means the vectors $c_0 = a_1$, $c_1 = a_1 - a_2$, $c_2 = a_1 - a_3$, and $c_3 = a_1 - a_2 - a_3 + a_4$ make up the new basis. In this basis, the multiplication is given by the following table:

	c_0	c_1	c_2	c_3
c_0	c_0	$\frac{1}{2}c_1$	$\frac{1}{2}c_2$	$\frac{1}{2}(1-\theta)c_3$
c_1		0	$\frac{1}{2}\theta c_3$	0
c_2			0	0
c_3				0

Thus, this algebra is Gonshor-genetic. □

Woerz [54] gives a detailed overview on the construction of genetic algebras according to Heuch [27], Reiersol [47], and Holgate [30]. Essentially, it is the tensor product of the gametic algebras of the different alleles at the corresponding loci. For our presentation it suffices to look at the concrete algebras as we do in the second part of this excursus. Heuch [27] showed that the recombination algebra is genetic.

We are interested in the analysis of the evolution map

$$E : p \mapsto p^2.$$

A problem is to determine successive generations in terms of the initial generation. Haldane [25] gave a constructive procedure which tackled this problem. It was

later theoretically examined by Holgate [29] and McHale and Ringwood [41]. The key idea is to embed the set of populations which is the pre-image of 1 under the weight function into a (higher but) finite dimensional vector space V such that the following diagram commutes, where L is a linear map.

$$\begin{array}{ccc} \omega^{-1}(1) & \xrightarrow{D} & V \\ \Big\downarrow{\scriptstyle E} & & \Big\downarrow{\scriptstyle L} \\ \omega^{-1}(1) & \xrightarrow{D} & V \end{array} \tag{2.1}$$

In our context we call algebras *linearisable*, if this procedure succeeds for the evolution map.

Example 2.6. In the situation of our example in the canonical basis, the set of populations $\omega^{-1}(1)$ is simply given by

$$\{p = c_0 + g^1 c_1 + g^2 c_2 + g^3 c_3 \mid g^1, g^2, g^3 \in \mathbb{R}\}.$$

We consider the effect of the evolution map on such an arbitrary population $p = c_0 + g^1 c_1 + g^2 c_2 + g^3 c_3$:

$$\begin{aligned} E(p) = p^2 &= c_0^2 + 2g^1 c_0 c_1 + 2g^2 c_0 c_2 + 2g^3 c_0 c_3 + 2^1 p^2 c_1 c_2 \\ &= c_0 + g^1 c_1 + g^2 c_2 + ((1-\theta)g^3 + \theta g^1 g^2) c_3 \,. \end{aligned}$$

Thus, the coordinates are mapped in the following way:

$$g^1 \mapsto g^1, \quad g^2 \mapsto g^2, \quad g^3 \mapsto (1-\theta)g^3 + \theta g^1 g^2.$$

So the transformation is nonlinear; a product term emerges. We take this new product term as an additional basis vector in the higher dimensional vector space W. In this vector space, we may represent the evolution map by the following matrix:

$$\begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1-\theta & 0 \\ 0 & 0 & \theta & 1 \end{pmatrix}$$

and hence obtain a linear mapping. □

Holgate [29] stated that genetic algebras are linearisable.

McHale and Ringwood found necessary and sufficient conditions for linearisability which they summarized in the following theorem.

Theorem 2.7. [41] *A baric algebra $\mathcal{A}$ is Haldane linearisable if and only if the plenary power*[2] *$\mathcal{A}^{[r]}$ is Schafer-genetic for some integer r.* □

[2]Plenary powers are defined recursively: $a^{[1]} := a$, $a^{[r]} := a^{[r-1]} a^{[r-1]}$.

In our example we first go over to the canonical basis before we linearise the dynamics. Abraham [1] showed that for Schafer algebras the linearisation is independent of the basis. Hence, the vector space in which the algebra is embedded is unique. Nevertheless, for practical use it is convenient to go to the canonical basis.

Haldane Linearisation for the recombination gametic algebra

We are interested in the concrete shape of Haldane linearisation in the context of recombination algebras. So, we pursued it for algebras with two, three, and four sites [51, 53].

2 sites

In this case, the natural basis of our algebra is given by $m \cdot n$ gametes: $a_1 = [1,1], a_2 = [1,2], \ldots, a_{mn} = [m,n]$, with multiplication given by

$$[i,j] \times [k,l] = \frac{1}{2}\big([i,j] + [k,l]\big) + \frac{1}{2}\theta\big(-([i,j] + [k,l]) + [i,l] + [k,j]\big)$$

for all $1 \leq i,k \leq m$, $1 \leq j,l \leq n$, where θ again is the recombination probability.

The canonical basis is then given by

$$\begin{aligned}
w_1 &= [1,1] = a_1\,, \\
w_j &= [1,1] - [1,j] = a_1 - a_j \quad \text{for } j = 2,\ldots,n\,, \\
w_{kn+1} &= [1,1] - [k+1,1] = a_1 - a_{kn+1} \quad \text{for } k = 1,\ldots,m-1\,, \\
w_{kn+j+1} &= [1,1] - [1,j+1] - [k+1,1] + [k+1,j+1]\,, \\
&= a_1 - a_{j+1} - a_{kn+1} + a_{kn+j+1} \quad k = 1,\ldots,m-1\,, j = 1,\ldots,n-1\,.
\end{aligned}$$

and denoted by the following transformation matrix:

$$\underbrace{\underbrace{\begin{pmatrix} 1 & 0 & 0 & \cdots & 0 \\ 1 & -1 & 0 & \cdots & 0 \\ 1 & 0 & -1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & 0 & 0 & \cdots & -1 \end{pmatrix}}_{m\times m} \otimes \underbrace{\begin{pmatrix} 1 & 0 & 0 & \cdots\cdots & 0 \\ 1 & -1 & 0 & \cdots\cdots & 0 \\ 1 & 0 & -1 & \cdots\cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & 0 & 0 & \cdots\cdots & -1 \end{pmatrix}}_{n\times n}}_{(mn)\times(mn)}\,.$$

Here, $\otimes$ denotes the tensor product for matrices, also known as *Kronecker product* [22]. The new basis is Gonshor as the following equations reveal:

$$\begin{aligned} w_1 \times w_1 &= w_1\,, \\ w_1 \times w_j &= \frac{1}{2} w_j \quad \text{for } j = 2, \ldots, n\,, \\ w_1 \times w_{kn+1} &= \frac{1}{2} w_{kn+1} \quad \text{for } k = 1, \ldots, m-1\,, \\ w_1 \times w_{kn+j+1} &= \frac{1}{2}\theta w_{kn+j+1} \quad \text{for } k = 1, \ldots, m-1 \text{ and } j = 1, \ldots, n-1\,, \end{aligned}$$

while all other multiplications yield 0. The populations are then given by

$$p = w_1 + g^2 w_2 + \ldots + g^{mn} w_{mn} \quad \text{with } g^i \in \mathbb{R}\,.$$

Under the evolution map E the coordinates g^i are mapped in the following way:

$$\begin{aligned} g^j &\mapsto g^j \quad \text{for } j = 1, \ldots, n \text{ and } j = n+1, 2n+1, \ldots, (m-1)n+1 \\ g^{kn+j+1} &\mapsto (1-\theta) g^{kn+j+1} + \theta g^{j+1} g^{kn+1} \; k = 1, \ldots, m-1, j = 1, \ldots, n-1\,. \end{aligned}$$

For the linearisation we extend the vector space by $(m-1)(n-1)$ new basis vectors, namely

$$g^{j,k} = g^{j+1} \cdot g^{kn+1} \quad \text{for } k = 1, \ldots, m-1 \text{ and } j = 1, \ldots, n-1$$

and map them onto themselves. Then the dynamics can be described by a $(2mn - m - n) \times (2mn - m - n)$-matrix:

$$\mathbb{1}_{(m-1)+(n-1)} \oplus \left(\mathbb{1}_{(m-1)(n-1)} \otimes \begin{pmatrix} 1-\theta & \theta \\ 0 & 1 \end{pmatrix} \right),$$

where $\oplus$ denotes the *Kronecker sum* of two matrices [22].

The shape of this matrix can be understood in the following way. In the new Gonshor basis we have the specified type $a_1 = [1, 1]$, and all other new basis vectors are defined in respect to it. Recombination of this type with another type that carries allele 1 at site 1 or 2 does not change the types. These are covered by the identity matrix. Recombination events with other types do change the composition of the population and, of course, their dynamics is essentially the same as in Example 2.6.

3 sites

In the situation of three sites, three recombination events are possible. We introduce the probabilities θ_1, and θ_2 resp., for the recombination events $(1|23)$, and $(12|3)$ resp. The double crossover event $(13|2)$ has probability γ.

In the natural basis, which is analogue to the two-site case, $a_1 = [1,1,1]$, $a_2 = [1,1,2], \ldots, a_{n_1 n_2 n_3} = [n_1, n_2, n_3]$, the multiplication is then given by

$$\begin{aligned}
[a,b,c] \times [d,e,f] =& \frac{1}{2}\big([a,b,c] + [d,e,f]\big) \\
&+ \frac{1}{2}\theta_1\big(-[a,b,c] - [d,e,f] + [a,e,f] + [d,b,c]\big) \\
&+ \frac{1}{2}\theta_2\big(-[a,b,c] - [d,e,f] + [a,b,f] + [d,e,c]\big) \\
&+ \frac{1}{2}\gamma\big(-[a,b,c] - [d,e,f] + [a,e,c] + [d,b,f]\big)\,.
\end{aligned}$$

Again, we change the basis. The new basis is given by

$$\begin{aligned}
w_{1,1,1} &= [1,1,1]\,, \\
w_{1,1,k} &= [1,1,1] - [1,1,k] \quad \text{for } k = 2,\ldots,n_3\,, \\
w_{1,j,1} &= [1,1,1] - [1,j,1] \quad \text{for } j = 2,\ldots,n_2\,, \\
w_{i,1,1} &= [1,1,1] - [i,1,1] \quad \text{for } i = 2,\ldots n_1\,, \\
w_{i,j,1} &= [1,1,1] - [1,j,1] - [i,1,1] + [i,j,1] \quad \text{for } i = 2,\ldots,n_1 \text{ and } j = 2,\ldots,n_2\,, \\
w_{i,1,k} &= [1,1,1] - [i,1,1] - [1,1,k] + [i,1,k] \quad \text{for } i = 2,\ldots,n_1 \text{ and } k = 2,\ldots,n_3\,, \\
w_{1,j,k} &= [1,1,1] - [1,j,1] - [1,1,k] + [1,j,k] \quad \text{for } j = 2,\ldots,n_2 \text{ and } k = 2,\ldots,n_3\,, \\
w_{i,j,k} &= [1,1,1] - [1,1,k] - [1,j,1] - [i,1,1] + [1,j,k] + [i,1,k] + [i,j,1] - [i,j,k] \\
&\quad \text{for } i = 2,\ldots,n_1,\ j = 2,\ldots,n_2 \text{ and } k = 2,\ldots,n_3\,.
\end{aligned}$$

where the transformation is represented by the matrix

$$\underbrace{\begin{pmatrix} 1 & 0 & 0 & \cdots & 0 \\ 1 & -1 & 0 & \cdots & 0 \\ 1 & 0 & -1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & 0 & 0 & \cdots & -1 \end{pmatrix}}_{n_1 \times n_1} \otimes \underbrace{\begin{pmatrix} 1 & 0 & 0 \cdots & 0 \\ 1 & -1 & 0 \cdots & 0 \\ 1 & 0 & -1 \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots & \vdots \\ 1 & 0 & \cdots\cdots & -1 \end{pmatrix}}_{n_2 \times n_2} \otimes \underbrace{\begin{pmatrix} 1 & 0 & 0 & \cdots & 0 \\ 1 & -1 & 0 & \cdots & 0 \\ 1 & 0 & -1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & 0 & 0 & \cdots & -1 \end{pmatrix}}_{n_3 \times n_3}\,.$$

This basis is Gonshor.

Again, we consider the action of the evolution map onto the coordinates of an arbitrary population $p = g^{1,1,1} w_{1,1,1} + g^{1,1,2} w_{1,1,2} + \ldots + g^{n_1,n_2,n_3} w_{n_1,n_2,n_3}$:

$$
\begin{aligned}
g^{1,1,1} &\mapsto g^{1,1,1}\,,\\
g^{1,1,k} &\mapsto g^{1,1,k} \quad k=2,\dots,n_3\,,\\
g^{1,j,1} &\mapsto g^{1,j,1} \quad j=2,\dots,n_2\,,\\
g^{i,1,1} &\mapsto g^{i,1,1} \quad i=2,\dots n_1\,,\\
g^{i,j,1} &\mapsto (1-\theta_1-\gamma)g^{i,j,1}+(\theta_1+\gamma)g^{1,j,1}\cdot g^{i,1,1} \quad i=2,\dots,n_1\,, j=2,\dots,n_2\,,\\
g^{i,1,k} &\mapsto (1-\theta_1-\theta_2)g^{i,1,k}+(\theta_1+\theta_2)g^{1,1,k}\cdot g^{i,1,1} \quad i=2,\dots,n_1\,, k=2,\dots,n_3\,,\\
g^{1,j,k} &\mapsto (1-\theta_2-\gamma)g^{1,j,k}+(\theta_2+\gamma)g^{1,1,k}\cdot g^{1,j,1} \quad j=2,\dots,n_2\,, k=2,\dots,n_3\,,\\
g^{i,j,k} &\mapsto (1-\theta_1-\theta_2-\gamma)g^{i,j,k}+\theta_1\cdot g^{i,1,1}\cdot g^{1,j,k}+\theta_2\cdot g^{1,1,k}+\gamma\cdot g^{1,j,1}\cdot g^{i,1,k}\\
&\quad \text{for } i=2,\dots,n_1,\ j=2,\dots,n_2 \text{ and } k=2,\dots,n_3\,.
\end{aligned}
$$

New product terms arise. Under the evolution map, these are mapped in the following way:

$$
\begin{aligned}
g^{1,j,1}\cdot g^{i,1,1} &\mapsto g^{1,j,1}\cdot g^{i,1,1} \quad \text{for } i=2,\dots,n_1 \text{ and } j=2,\dots,n_2\,,\\
g^{1,1,k}\cdot g^{i,1,1} &\mapsto g^{1,1,k}\cdot g^{i,1,1} \quad \text{for } i=2,\dots,n_1 \text{ and } k=2,\dots,n_3\,,\\
g^{1,1,k}\cdot g^{1,j,1} &\mapsto g^{1,1,k}\cdot g^{1,j,1} \quad \text{for } j=2,\dots,n_2 \text{ and } k=2,\dots,n_3\,,\\
g^{i,1,1}\cdot g^{1,j,k} &\mapsto g^{i,1,1}\big((1-\theta_2-\gamma)g^{1,j,k}+g^{1,1,k}\cdot g^{1,j,1}(\theta_2+\gamma)\big)\\
&\quad \text{for } i=2,\dots,n_1,\ j=2,\dots,n_2 \text{ and } k=2,\dots,n_3\,,\\
g^{1,1,k}\cdot g^{i,j,1} &\mapsto g^{1,1,k}\big((1-\theta_1-\gamma)g^{i,j,1}+g^{1,j,1}\cdot g^{i,1,1}(\theta_1+\gamma)\big)\\
&\quad \text{for } i=2,\dots,n_1,\ j=2,\dots,n_2 \text{ and } k=2,\dots,n_3\,,\\
g^{1,j,1}\cdot g^{i,1,k} &\mapsto g^{1,j,1}\big((1-\theta_1-\theta_2)g^{i,1,k}+g^{1,1,k}\cdot g^{i,1,1}(\theta_1+\theta_2)\big)\\
&\quad \text{for } i=2,\dots,n_1,\ j=2,\dots,n_2 \text{ and } k=2,\dots,n_3\,,
\end{aligned}
$$

.

A further product term emerges, it is mapped onto itself:

$$
\begin{aligned}
g^{i,1,1}\cdot g^{1,1,k}\cdot g^{1,j,1} &\mapsto g^{i,1,1}\cdot g^{1,1,k}\cdot g^{1,j,1}\\
&\quad \text{for } i=2,\dots,n_1,\ j=2,\dots,n_2 \text{ and } k=2,\dots,n_3.
\end{aligned}
$$

Thus, the dynamics can be linearised by adding these

$$(n_1-1)(n_2-1)+(n_1-1)(n_3-1)+(n_2-1)(n_3-1)+4(n_1-1)(n_2-1)(n_3-1)$$

additional parameters and the resulting linear mapping is represented by the

following matrix:

$$\mathbb{1}_{(n_1-1)+(n_2-1)+(n_3-1)} \oplus \mathbb{1}_{(n_1-1)(n_2-1)} \otimes \begin{pmatrix} 1-\theta_1-\gamma & \theta_1+\gamma \\ 0 & 1 \end{pmatrix}$$
$$\oplus\, \mathbb{1}_{(n_1-1)(n_3-1)} \otimes \begin{pmatrix} 1-\theta_1-\theta_2 & \theta_1+\theta_2 \\ 0 & 1 \end{pmatrix}$$
$$\oplus\, \mathbb{1}_{(n_2-1)(n_3-1)} \otimes \begin{pmatrix} 1-\theta_2-\gamma & \theta_2+\gamma \\ 0 & 1 \end{pmatrix} \oplus$$
$$\mathbb{1}_{(n_1-1)(n_2-1)(n_3-1)} \otimes \begin{pmatrix} 1-\theta_1-\theta_2-\gamma & \theta_1 & \theta_2 & \gamma & 0 \\ 0 & 1-\theta_2-\gamma & 0 & 0 & \theta_2+\gamma \\ 0 & 0 & 1-\theta_1-\gamma & 0 & \theta_1+\gamma \\ 0 & 0 & 0 & 1-\theta_1-\theta_2 & \theta_1+\theta_2 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix}$$

Again, the structure of the matrix can be explained by the same reasoning as in the two sites case. Types that carry allele 1 at two sites correspond to the identity matrix, those which carry allele 1 at one site reveal essentially the same dynamics as in the two-site case and those that differ from the specified type a_1 at all of their sites belong to the building block with the 4×4-matrix.

4 sites

With four sites, we have six recombination events and thus six parameters. The multiplication rule is

$$\begin{aligned}
[a,b,c,d] \times [e,f,g,h] = {} & \frac{1}{2}\big([a,b,c,d] + [e,f,g,h]\big) \\
& + \frac{1}{2}\theta_1\big(-[a,b,c,d] - [e,f,g,h] + [a,f,g,h] + [e,b,c,d]\big) \\
& + \frac{1}{2}\theta_2\big(-[a,b,c,d] - [e,f,g,h] + [a,b,g,h] + [e,f,c,d]\big) \\
& + \frac{1}{2}\theta_3\big(-[a,b,c,d] - [e,f,g,h] + [a,b,c,h] + [e,f,g,d]\big) \\
& + \frac{1}{2}\gamma_1\big(-[a,b,c,d] - [e,f,g,h] + [a,f,c,d] + [e,b,c,d]\big) \\
& + \frac{1}{2}\gamma_2\big(-[a,b,c,d] - [e,f,g,h] + [a,b,g,d] + [e,f,c,h]\big) \\
& + \frac{1}{2}\gamma_3\big(-[a,b,c,d] - [e,f,g,h] + [a,f,g,d] + [e,b,c,h]\big) \\
& + \frac{1}{2}\sigma\big(-[a,b,c,d] - [e,f,g,h] + [a,f,c,h] + [e,b,g,d]\big).
\end{aligned}$$

Again, we specify one type $[1,1,1,1]$ for the change of the basis. It works following the same procedure as in the previous cases. It becomes clear that it is comparable to the Moebius transformation [2]

$$[i,j,k,l] \mapsto \sum_{U \subset \mathcal{P}(\{i,j,k,l\})} (-1)^{|U|}[\mathbf{1},U],$$

where the notation $[\mathbf{1},U]$ means that the type is made up of the alleles in U at the corresponding sites and 1 elsewhere. The corresponding matrix is clear. This basis is also Gonshor.

Again, we consider the effect of the evolution map onto an arbitrary population and add emerging product terms of the coordinates as new basis vectors. The scheme that we could observe in the two and three site cases is approved. Each possible product of coordinates that belongs to one or several recombination events of an arbitrary type with the specified type occurs. For example, the new coordinates that emerge together with the coordinate g^{ijkl} are listed in the first column of the following table.

g^{ijkl}	$1-\theta_1-\theta_2-\theta_3-\gamma_1-\gamma_2-\gamma_3-\sigma$
$g^{i111}g^{1jkl}$	$1-\theta_2-\theta_3-\gamma_1-\gamma_2-\gamma_3-\sigma$
$g^{ij11}g^{11kl}$	$(1-\theta_1-\gamma_1-\gamma_3-\sigma)(1-\theta_3-\gamma_2-\gamma_3-\sigma)$
$g^{ijk1}g^{111l}$	$1-\theta_1-\theta_2-\gamma_1-\gamma_2-\gamma_3-\sigma$
$g^{i1kl}g^{1j11}$	$1-\theta_1-\theta_2-\theta_3-\gamma_2-\gamma_3-\sigma$
$g^{ij1l}g^{11k1}$	$1-\theta_1-\theta_2-\theta_3-\gamma_1-\gamma_3-\sigma$
$g^{i11l}g^{1jk1}$	$(1-\theta_1-\theta_2-\theta_3-\sigma)(1-\theta_2-\gamma_1-\gamma_2-\sigma)$
$g^{i1k1}g^{1j1m}$	$(1-\theta_1-\theta_2-\gamma_2-\gamma_3)(1-\theta_2-\theta_3-\gamma_1-\gamma_3)$
$g^{i111}g^{1j11}g^{11kl}$	$1-\theta_3-\gamma_2-\gamma_3-\sigma$
$g^{i111}g^{1jk1}g^{111l}$	$1-\theta_2-\gamma_1-\gamma_2-\sigma$
$g^{i111}g^{1j1l}g^{11k1}$	$1-\theta_2-\gamma_1-\gamma_2-\sigma$
$g^{ij11}g^{11k1}g^{111l}$	$1-\theta_2-\theta_3-\gamma_1-\gamma_3$
$g^{i1k1}g^{1j11}g^{111l}$	$1-\theta_1-\theta_2-\gamma_2-\gamma_3$
$g^{i11l}g^{1j11}g^{11k1}$	$1-\theta_1-\theta_2-\theta_3-\sigma$
$g^{i111}g^{1j11}g^{11k1}g^{111l}$	1

The building blocks of the matrix that represents the evolution map are thus 15×15 matrices and we do not present them in this thesis. These are upper triangular, too. The eigenvalues are listed in the second column of the table above.

Thus, for some basis vectors, the scheme of the dynamics that could be found in the previous cases remains. However, for some vectors, new product terms emerge, and we could not give a proper heuristic interpretation of these. So, the dynamics becomes somewhat cumbersome. Thus, a generalisation for more sites is not obvious, and the dynamics does not permit an obvious heuristic interpretation anymore.

Conclusion

Our interest lied in the dynamics of recombination in the framework of genetic algebras. We wanted to determine the solution of the Haldane linearisation procedure explicitly and stated it for the first four sites. Unfortunately, the out-coming eigenvalues did not give some more insight into the dynamics.

The dimension of the algebra depends on the number of alleles and the solutions become very large for a wide type space. Another disadvantage of this approach lies in the fact that the concrete solution is only available after the change of basis.

An ansatz that does not have these disadvantages is discussed in detail in [52, 53] in the case of single-crossover recombination.

Nevertheless, the basic idea of Haldane linearisation which works successfully in the field of genetic algebras, which is adding nonlinear terms as additional basis vectors, motivated us to tackle the dynamics of the Moran model with recombination with the moment closure approach.

2.2 Moran model with recombination

We consider a population of N individuals. Each of them is endowed with the set $S = \{1, \ldots, n\}$ of sites. These can be interpreted as nucleotide positions in a string of DNA or as gene loci on a chromosome. For each site i there is a finite set X_i of alleles that may occur at site i. A string of alleles is then called a type, $X := \bigtimes_{i=1}^{n} X_i$ is the type space.

We are interested in modeling recombination, which means the rearrangement of genetic material in sexually reproducing populations. It may occur during meiosis, the creation of gametes, that is egg cells or sperm. Homologous chromosomes may cross over at some points and exchange the genetic material in between (see Figure 2.1).

In the following, we will assign recombination events to subsets of sites in a natural way. Let $G \subset S$. Then the corresponding recombination event between two individuals is the following: the alleles at the sites given by G remain at their positions, whereas the alleles at the sites in $\bar{G}$, the complement of G, are exchanged (see Figure 2.2).

We define the mappings $p_G : X \times X \to X$, $G \subset S$ by

$$p_G(x, y) = : \Big(\bigtimes\nolimits_{i \in G} \{x_i\} \Big) \times \Big(\bigtimes\nolimits_{i \in \bar{G}} \{y_i\} \Big) : , \tag{2.2}$$

where $: \cdots :$ means that the coordinates are ordered as in X. So, $p_G(x, y)$ and $p_{\bar{G}}(x, y)$ are the new types resulting from the recombination event corresponding

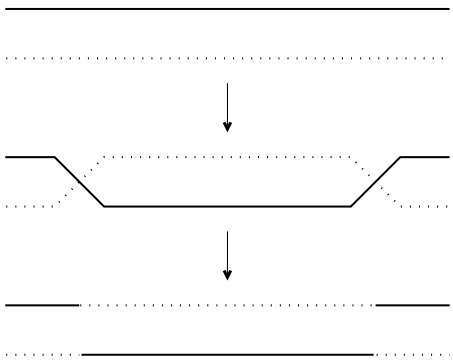

Figure 2.1: Recombination

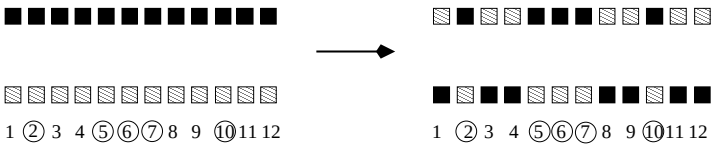

Figure 2.2: Recombination event defined by the set G (circled sites)

to G between the types x and y. Obviously, $p_G(x, y) = p_{\bar{G}}(y, x)$. So, G and $\bar{G}$ essentially correspond to the same recombination event.

We now define a Moran model with recombination and mutation. Each individual undergoes recombination events corresponding to $G \subset S$ at rate $\varrho_G/4 \geq 0$ for all $G \subset S$. The recombination partner is chosen out of the whole population (including the opening individual itself). Then they exchange their genetic material according to the recombination event corresponding to G (see Figure 2.3). To keep things well-defined, the recombination rates ϱ_G have the properties $\varrho_G = \varrho_{\bar{G}}$ and $\varrho_\varnothing = \varrho_S = 0$.

Furthermore, mutation events may occur. An allele $x_i \in X_i$ at site i mutates into allele $y_i \in X_i$ with rate $\mu^i_{x_i y_i} \geq 0$. Thus, the mutation rate depends on both the parental and the offspring allele.

Additionally, we introduce birth events or, more precisely, resampling. Each individual produces an offspring at rate $b/2 \geq 0$. The offspring inherits the parent's type and replaces another individual, randomly chosen from the entire population (again including the parent individual).

In the following we are interested in the composition of the population, so we define the stochastic process $(Z_t)_{t\geq 0}$ with state space

$$E := \{\omega \text{ counting measure on } X \text{ with } \omega(X) = N\},$$

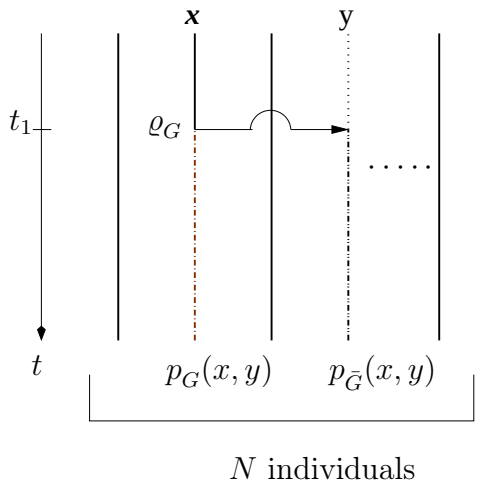

Figure 2.3: Moran model with recombination. At time t_1 the second individual, which is of type x, undergoes a recombination event corresponding to G and chooses its partner randomly, here the fourth individual, which is of type y; from that time on, the individuals are of type $p_G(x,y)$ and $p_{\bar{G}}(x,y)$.

by

$$Z_t(\{x\}) := \text{number of individuals of type } x.$$

In the following we will use shorthands like $Z_t(x)$, $z(x)$ instead of $Z_t(\{x\})$, $z(\{x\})$. Recombination, mutation and resampling events induce the following transitions if $Z_t = z$:

$$\begin{aligned} &z \to z + v_{U,x,y} \quad \text{with} \quad v_{U,x,y} := -\delta_x - \delta_y + \delta_{p_G(x,y)} + \delta_{p_{\bar{G}}(x,y)} \\ &\text{at rate} \quad \frac{1}{N}\varrho_G z(x)z(y) \quad \text{for} \quad x,y \in X, \quad G \subset S, \end{aligned} \tag{2.3}$$

$$z \to z - \delta_{(x_1,\ldots,x_i,\ldots,x_n)} + \delta_{(x_1,\ldots,y_i,\ldots,x_n)} \quad \text{at rate} \quad \mu^i_{x_i y_i} z(x), \tag{2.4}$$

$$z \to z + \delta_x - \delta_y \quad \text{at rate} \quad \frac{b}{2N} z(x)z(y). \tag{2.5}$$

The rate in (2.3) is determined in the following way: An individual of type x recombines at rate $\frac{1}{4}\varrho_G z(x)$ and chooses one individual of type y with probability $\frac{z(y)}{N}$. This leads to the rate $\frac{1}{4N}\varrho_G z(x)z(y)$ which needs to be multiplied by 4 to account for the fact that the recombination could be initiated by an individual of type y and that recombination according to G is the same as recombination according to $\bar{G}$.

A brief comment on the model is in order. We consider recombination and reproduction as independent events whereas, in true biology, recombination is coupled to reproduction. We use the decoupled version here because it is simpler, and because it allows to clearly separate the effects of random recombination from those of random reproduction. This version is also used elsewhere [45], on the argument that recombination events are rare.

For a subset $G \subset S$ we define $X_G := \bigtimes_{i\in G} X_i$ and the mapping $\pi_G : X \to X_G$ as the canonical projection. Let ω be a (signed) measure on X. We define the pullback $\pi_G.$ by $\pi_G.\omega := \omega \circ \pi_G^{-1}$. So, $\pi_G.$ maps a measure on X onto its corresponding marginal measure on X_G.

In the following, marginal processes of Z_t will play a crucial role. The following proposition states that these are Markov chains, too. It is an extension of Lemma 1 in [6].

Proposition 2.8. *Let $I \subset S$, and let $(Z_t)_{t\geq 0}$ be the recombination process as defined by equations* (2.3)-(2.5)*. Then $(\pi_I.Z_t)_{t\geq 0}$ is a Markov process with state space $E_I := \{\omega$ counting measure on X_I, $\omega(X_I) = N\}$.*

Proof. Obviously, $(\pi_I.Z_t)_{t\geq 0}$ is a stochastic process on E_I.

We must show that the transition rates of $(\pi_I.Z_t)_{t\geq 0}$ only depend on the current state of the process. A recombination event induces the following transition:

$$\pi_I.z \to \pi_I.(z + v_{U,x,y}),$$

with

$$\pi_I.v_{U,x,y} = \delta_{\pi_I(p_G(x,y))} + \delta_{\pi_I(p_{\bar{G}}(x,y))} - \delta_{\pi_I(x)} - \delta_{\pi_I(y)} \tag{2.6}$$

and $\pi_I(p_G(x,y)) =: \Big(\bigtimes_{i\in G\cap I}\{x_i\}\Big) \times \Big(\bigtimes_{i\in \bar{G}\cap I}\{y_i\}\Big)$: in line with (2.2).

Consider now any nonzero jump. If it comes from a recombination event, it must be of the form (2.6). That means there are types $x_I, y_I \in X_I$ and a subset H of I such that $(\pi_I.z)(x_I)$ and $(\pi_I.z)(y_I)$ both decrease by one and the frequencies of the marginal types arising in the recombination event corresponding to H increase. The rate for this transition is then given by the sum of all transitions of the original process that induce this transition in the marginal process:

$$\begin{aligned}\sum_{\substack{G\subset S:\\ G\cap I=H}} \sum_{\substack{x\in X:\\ \pi_I(x)=x_I}} \sum_{\substack{y\in X:\\ \pi_I(y)=y_I}} \frac{\varrho_G}{N} z(x)z(y) &= \sum_{\substack{G\subset S:\\ G\cap I=H}} \frac{\varrho_G}{N}\big(\pi_I.z\big)\big(x_I\big)\cdot\big(\pi_I.z\big)\big(y_I\big) \\ &= \frac{\varrho_H^{(I)}}{N}\big(\pi_I.z\big)(x_I)\cdot\big(\pi_I.z\big)(y_I),\end{aligned} \tag{2.7}$$

with $\varrho_H^{(I)} := \sum_{G\subset S: G\cap I=H} \varrho_G$. So, this last term depends only on the current state of the marginal process $(\pi_I.Z_t)_{t\geq 0}$.

A mutation event of an individual of type x at site i from allele x_i to allele y_i induces the following transition of $\pi_I.Z_t$:

$$\pi_I.z \to \pi_I.z + \pi_I.\delta_{(\dots,y_i,\dots)} - \pi_I.\delta_{(\dots,x_i,\dots)}.$$

This jump is zero if $i \notin I$. Obviously, the transition rate is $\mu^i_{x_i y_i}(\pi_I.z)(\pi_I(x))$ and depends merely on the current state of $\pi_I.Z_t$, too. The case of resampling is treated analogously. □

This proof is an example for the so-called lumping procedure for Markov chains, compare [13, 32] for the general context or [4] for the sequence context considered here.

Remark 2.9. A comparison between (2.3) and (2.7) shows that the marginal process $(\pi_I.Z_t)_{t\geq 0}$ can itself be considered as a recombination process on the sites I. So, assertions about Z_t will also hold for all derived marginal processes.

2.3 Recombination alone

In this Section we restrict ourselves to the case *without mutation and resampling*, that means with $\mu^i_{x_i y_i} = b = 0$ for all $i \in S$ and $x_i, y_i \in X_i$.

Since $v_{U,x,y} = 0$ for some $x, y \in X$, there are 'empty' recombination events at positive rate, but including these redundancies makes the rates in (2.3) so simple. The rates become considerably more complicated if only 'true jumps' are considered. This is already visible in the projection onto a single type. Let $x \in X$ and $Z_t = z$. In order to figure out the rate for the transition $z(x) \to z(x)+1$, we first determine the set of all pairs of types $\tilde{x}, \tilde{y} \in X$ such that for a given $G \subset S$ the jump $v_{G,\tilde{x},\tilde{y}}(x)$ equals 1:

$$\begin{aligned}
&\{\{\tilde{x},\tilde{y}\} \subset X : v_{G,\tilde{x},\tilde{y}}(x) = 1\} \\
&\quad = \{\{\tilde{x},\tilde{y}\} \subset X : \pi_G(\tilde{x}) = \pi_G(x), \pi_{\bar{G}}(\tilde{y}) = \pi_{\bar{G}}(x), \tilde{x} \neq x, \tilde{y} \neq x\} \\
&\quad = \{\{\tilde{x},\tilde{y}\} \subset X : \tilde{x} \in \pi_G^{-1}\big(\pi_G(x)\big) \setminus \{x\}, \tilde{y} \in \pi_{\bar{G}}^{-1}\big(\pi_{\bar{G}}(x)\big) \setminus \{x\}\}.
\end{aligned}$$

This leads to the transition rate

$$\sum_{G\subset S} \frac{\varrho_G}{2N} \big[\big(\pi_G.z\big)\big(\pi_G(x)\big) - z(x)\big] \big[\big(\pi_{\bar{G}}.z\big)\big(\pi_{\bar{G}}(x)\big) - z(x)\big]. \tag{2.8}$$

The transition rate for $z(x) \to z(x)-1$ can be figured out analogously; $v_{G,\tilde{x},\tilde{y}}(x) = -1$ iff $\tilde{x} = x$ and $\tilde{y}$ is any type which is neither in $\pi_G^{-1}\big(\pi_G(x)\big)$ nor in $\pi_{\bar{G}}^{-1}\big(\pi_{\bar{G}}(x)\big)$. Since $\pi_G^{-1}\big(\pi_G(x)\big) \cap \pi_{\bar{G}}^{-1}\big(\pi_{\bar{G}}(x)\big) = \{x\}$ one has the rate

$$\sum_{G\subset S} \frac{\varrho_G}{2N} z(x) \big[N - \big(\pi_G.z\big)\big(\pi_G(x)\big) - \big(\pi_{\bar{G}}.z\big)\big(\pi_{\bar{G}}(x)\big) + z(x)\big]. \tag{2.9}$$

Our aim is now to reformulate the process with the help of additional random variables, so that the transition rates become simpler, in particular, unaffected by empty events. To this end, we define two new counting measures derived from $(Z_t)_{t\geq 0}$, namely $(U_t)_{t\geq 0}$ by $U_0(x) = 0$ and

$$U_t(x) = \text{number of events at which } x\text{-individuals are created until time } t$$

and $(V_t)_{t\geq 0}$ by $V_0(x) = 0$ and

$$V_t(x) = \text{number of events at which } x\text{-individuals are broken up until time } t.$$

These processes also count events at which Z_t does not change, namely the case that individuals of type x are created and broken up at the same time. This may happen when an individual of type x recombines according to G with an individual of type y with $\pi_G(y) = \pi_G(x)$. Whenever this occurs, both counters increase but their difference remains unchanged. Altogether, we thus have

$$Z_t = Z_0 + U_t - V_t \tag{2.10}$$

with the transition rates of U_t and V_t unaffected by 'empty' events: For $U_t(x) = u$, $u \to u + 1$ happens at rate

$$\sum_{G\subset S} \frac{\varrho_G}{2N} \big(\pi_G.z\big)\big(\pi_G(x)\big) \cdot \big(\pi_{\bar{G}}.z\big)\big(\pi_{\bar{G}}(x)\big),$$

and for $V_t(x) = v$, the transition $v \to v + 1$ happens at rate

$$\sum_{G\subset S} \frac{\varrho_G}{2} z(x).$$

In the following, marginal processes will emerge frequently. We introduce a shorthand, symbolic notation similar to the one described in [3]. Fix an arbitrary $x \in X$ and define for a subset $G = \{g_1, \ldots, g_{|G|}\}$ of sites

$$[G]_t := [g_1, \ldots, g_{|G|}]_t := \big(\pi_G.Z_t\big)\big(\pi_G(x)\big).$$

$[G]_t$ is the number of individuals that are identical to x at the sites corresponding to G, at time t. Again, we use shorthands $[g_1, \ldots, g_{|G|}]_t$ instead of $[\{g_1, \ldots, g_{|G|}\}]_t$. Note that we suppress the dependence on x in $[G]_t$ for ease of notation. Analogously, we define for the processes U_t and V_t:

$$\langle G\rangle_t := \big(\pi_G.U_t\big)\big(\pi_G(x)\big),$$

$$(G)_t := \big(\pi_G.V_t\big)\big(\pi_G(x)\big).$$

By Remark 2.9, we can now consider $[G]_t$ as a recombination process on $|G|$ sites evaluated at the type $(x_{g_1}, \ldots, x_{g_{|G|}})$.

For $|G| = 2$, the distribution of $\langle g_1, g_2 \rangle_t$ can be given explicitly because the transition rates

$$\sum_{H \subset S : |H \cap G| = 1} \frac{\varrho_H}{2N} [g_1]_t [g_2]_t =: a \tag{2.11}$$

are constant in time because all 1-site marginals are constant in time. So, $\langle g_1, g_2 \rangle_t$ follows a Poisson distribution with parameter at.

2.3.1 Analysis of the expectation

Since we will use it frequently, we want to recall an elementary fact concerning the dynamics of the mean of a continuous-time Markov chain with a finite state space, which is often used implicitly. The proof is a straightforward exercise that can be found in [6, Fact 1], for example.

Lemma 2.10. *Let $(Z_t)_{t \geq 0}$ be a Markov process with finite state space $E \subset \mathbb{Z}^d$ with transition rates $q(z, z+v)$ for transitions from z to $z+v$ for $z \in E$, $v \neq 0$ (let $q(z, z+v) = 0$ if $z + v \notin E$). Then the following equation holds for all $t \geq 0$*

$$\frac{d}{dt} \mathbb{E}(Z_t) = \mathbb{E}(F(Z_t)),$$

where F is defined as

$$F(z) := \sum_{v \in \mathbb{Z}^d} v q(z, z+v).$$

□

Lemma 2.10 together with the representation of Z_t in (2.10) gives us the dynamics of the mean:

$$\frac{d}{dt} \mathbb{E}\Big[[1, \ldots, n]_t\Big] = \sum_{G \subset S} \mathbb{E}\Big[\frac{\varrho_G}{2N} ([G]_t [\bar{G}]_t - N \cdot [1, \ldots, n]_t)\Big]. \tag{2.12}$$

The motivation for this comes from the well-understood special case of single crossovers [6]. Here, all recombination rates that are attached to multiple crossover recombination events vanish. This affects all ϱ_G with G that either do not contain 1 or n, or have gaps.

In this case, the induced marginal processes are conditionally independent of each other and so moment closure is immediate [6, Lemma 1 and Theorem 1]:

$$\frac{d}{dt}\mathbb{E}\Big[[1,\dots,n]_t\Big] = \sum_{G\subset S}\mathbb{E}\Big[\frac{\varrho_G}{2N}([G]_t[\bar{G}]_t - N\cdot[1,\dots,n]_t)\Big]$$
$$= \sum_{G\subset S}\frac{\varrho_G}{2N}(\mathbb{E}\Big[[G]_t\Big]\mathbb{E}\Big[[\bar{G}]_t\Big] - N\cdot\mathbb{E}\Big[[1,\dots,n]_t\Big]).$$

We obtain a finite nonlinear system of differential equations, whose solution is known in closed form [3].

The independence relies on two properties. First, a single crossover recombination event induces a pair of marginal processes $[G]_t, [\bar{G}]_t$ for which $\{G, \bar{G}\}$ is an *ordered* partition of S. Second, a single crossover recombination event only affects one of the induced processes while leaving the other one constant.

With general recombination both these properties are violated. First, marginal processes arise that are given by *non-ordered* partitions, so even single-crossover recombination events may affect both processes at the same instant. Second, a multiple-crossover recombination event may affect the frequency of a pair of marginals that are given by an ordered partition. So, the independence of the induced marginal processes is violated in two ways.

Let us now look at (2.12) again. On the right-hand side, an expectation of products emerges. This is what one may expect due to the inherent nonlinearity of the recombination process. Nevertheless, we see that no site arises more than once, so the arising products are described by a partition of sites. This leads us to the following question: Given an arbitrary partition of sites, what is the dynamics of the mean of the product of the induced marginal processes? Theorem 2.12 below answers this. For its formulation we need the following definition.

Definition 2.11. *Let $\{A_j\}_{j\in J}$ be a collection of sets with $A_i \cap A_j = \varnothing,\ i \neq j$. Define $A_J := \bigcup_{j\in J} A_j$. Then, $G \subset A_J$ disrupts $\{A_j\}_{j\in J}$, denoted by $G|\{A_j\}_{j\in J}$, if $G \cap A_j \neq \varnothing$ and $G \not\supseteq A_j$ for all $j \in J$. For $|J| = 1$, we simply write $G|A_j$.*

Note that for a collection of pairwise disjoint subsets of sites $\{A_j\}_{j\in J}$ disrupted by G, in a recombination event corresponding to G between individuals of marginal types $\pi_G(x)$ and $\pi_{A_J\setminus G}(x)$, the processes $\langle A_j\rangle_t,\ j \in J$, increase. Similarly, in the recombination event corresponding to G between individuals of marginal types $\pi_{A_K}(x)$ and $\pi_{A_J\setminus A_K}(x)$ for $K \subset J$, the processes $(A_j)_t,\ j \in J$, increase.

With these preparations, we are now ready to state

Theorem 2.12. *Let $m \leq n$, $M := \{1, \dots, m\}$ and $\mathcal{A} := \{A_1, \dots, A_m\}$ be a partition of $\{1, \dots, n\}$. Define $\mathcal{P}$ as the set of all triples (I, J, K), where $\{I, J, K\}$ is a partition of M. Then the derivative of $\mathbb{E}\Big[\prod_{\ell \in M} [A_\ell]_t\Big]$ is*

$$\mathbb{E}\Big[\sum_{\substack{(I,J,K)\in\mathcal{P}\\ I\neq M}} \prod_{i\in I}[A_i]_t \sum_{\tilde{K}\subset K} \sum_{\substack{G\subset A_J\\ G|\{A_j\}_{j\in J}}} \frac{\varrho^I_{K,G}}{4N}(-1)^{|K|}[A_{\tilde{K}} \cup G]_t [A_{K\setminus\tilde{K}} \cup G^c]_t\Big], \tag{2.13}$$

where G^c is the complement of G in A_J and $\varrho^I_{K,G}$ is defined as

$$\varrho^I_{K,G} := \sum_{D\subset A_I} \sum_{\substack{H\subset A_K\\ H|\{A_k\}_{k\in K}}} \varrho_{H\cup D\cup G}. \tag{2.14}$$

Remark 2.13. The right-hand side of (2.13) may be read in the following way. The set I indicates the parts of the genome that remain unchanged under the corresponding recombination event, the sets J and K indicate sets for which the derived processes U_t and V_t, respectively, increase. So the splitting of Z_t into U_t and V_t does not only simplify the calculation but also shows up in the result.

Proof of Theorem 2.12. Define

$$\langle A_\ell\rangle^t_{\delta t} := \langle A_\ell\rangle_{t+\delta t} - \langle A_\ell\rangle_t$$

and

$$(A_\ell)^t_{\delta t} := (A_\ell)_{t+\delta t} - (A_\ell)_t,$$

then

$$[A_\ell]_{t+\delta t} = [A_\ell]_t + \langle A_\ell\rangle^t_{\delta t} - (A_\ell)^t_{\delta t}$$

and $\prod_{\ell\in M}[A_\ell]_{t+\delta t}$ reads

$$\prod_{\ell\in M}[A_\ell]_{t+\delta t} = \sum_{(I,J,K)\in\mathcal{P}} (-1)^{|K|} \prod_{i\in I}[A_i]_t \prod_{j\in J}\langle A_j\rangle^t_{\delta t} \prod_{k\in K}(A_k)^t_{\delta t}.$$

Let $t+\delta t$ be the time of the first recombination event after time t.
Then, a summand $\prod_{i\in I}[A_i]_t \prod_{j\in J}\langle A_j\rangle^t_{\delta t} \prod_{k\in K}(A_k)^t_{\delta t}$ may evaluate to:

- zero if there is any $j \in J$ or $k \in K$ such that $\langle A_j\rangle^t_{\delta t} = 0$ or $(A_k)^t_{\delta t} = 0$;
- $(-1)^{|K|}\prod_{i\in I}[A_i]_t$ otherwise, that means if $\langle A_j\rangle^t_{\delta t} = (A_k)^t_{\delta t} = 1$ for all $j \in J$, $k \in K$.

The latter transition comes from recombination events that correspond to the union of some G disrupting $\{A_j\}_{j\in J}$ and H disrupting $\{A_k\}_{k\in K}$ and any subset D of A_I. At such recombination events, the recombining individuals must be of the following form: x-alleles at G, G^c resp., x-alleles at $A_k,\ k\in K$, whereas the particular A_k may be arbitrarily distributed across the two individuals (but the individual sets may not be disrupted!). Thus, the complete rate reads

$$r(I,J,K) := \sum_{\tilde{K}\subset K} \sum_{\substack{G\subset A_J \\ G|\{A_j\}_{j\in J}}} \frac{\varrho^I_{K,G}}{4N}[A_{\tilde{K}}\cup G]_t[A_{K\setminus\tilde{K}}\cup G^c]_t, \tag{2.15}$$

with G^c and $\varrho^I_{K,G}$ as defined above. This is the rate of the event that the terms corresponding to J and K increase, that means it is the rate of all recombination events such that $A_j,\ j\in J$, bindings arise, and such that $A_k,\ k\in K$, bindings break.

Thus,

$$\frac{d}{dt}\mathbb{E}\Big[\prod_{\ell\in M}[A_\ell]_t\Big] = \mathbb{E}\Big[\sum_{\substack{(I,J,K)\in\mathcal{P}\\ I\neq M}}(-1)^{|K|}\prod_{i\in I}[A_i]_t r(I,J,K)\Big],$$

which is the assertion of the theorem.

□

Let us now consider the implication of the theorem for the moment-closure problem. The theorem tells us that the dynamics of the mean of a product of marginal processes defined by a partition of sites can be described by the mean of another product of marginal processes defined by a (finer) partition of sites. Since the number of sites is finite and so is the number of partitions of sites, the moment closure approach (for the mean) directly leads to a finite and linear system of ODE's. We have thus proved

Corollary 2.14. *For the Moran model with recombination alone, the moment approach closes.*

Example 2.15. For three sites, the whole system of differential equations arising in the moment approach is still not too complex.

$$\begin{aligned}\frac{d}{dt}\mathbb{E}\Big[[1,2,3]_t\Big] =& \mathbb{E}\Big[\frac{\varrho_1}{N}[1]_t[2,3]_t + \frac{\varrho_2}{N}[2]_t[1,3]_t + \frac{\varrho_3}{N}[3]_t[1,2]_t\Big]\\ &- \mathbb{E}\Big[\frac{\varrho_1+\varrho_2+\varrho_3}{N}[1,2,3]_t\Big].\end{aligned}$$

$$\frac{d}{dt}\mathbb{E}\Big[[12]_t\Big] = \frac{\varrho_1+\varrho_2}{N}\mathbb{E}\Big[[1]_t[2]_t - [12]_t\Big].$$

$$\frac{d}{dt}\mathbb{E}\Big[[13]_t\Big] = \frac{\varrho_1+\varrho_3}{N}\mathbb{E}\Big[[1]_t[3]_t - [13]_t\Big].$$

$$\frac{d}{dt}\mathbb{E}\Big[[23]_t\Big] = \frac{\varrho_2+\varrho_3}{N}\mathbb{E}\Big[[2]_t[3]_t - [23]_t\Big].$$

$$\frac{d}{dt}\mathbb{E}\Big[[1]_t\Big] = \frac{d}{dt}\mathbb{E}\Big[[2]_t\Big] = \frac{d}{dt}\mathbb{E}\Big[[3]_t\Big] = 0.$$

□

The size of these systems explodes with the number of sites. Nevertheless, there is much redundancy in the concrete calculation of particular means. For example, in the analysis of $\mathbb{E}\big[[1,2,3,4]_t\big]$, marginal processes on three sites emerge. According to Remark 2.9, these can be treated as recombination processes on three sites, such that by a proper summation of the recombination rates, one can easily determine their solutions given the solution of the three-sites recombination process.

2.3.2 Comparison with the deterministic dynamics

We now want to compare the result of Theorem 2.12 to the corresponding deterministic dynamics. To this end, let $\mathcal{M}(X)$ be the space of all measures on X. For $G \subset S$ define the recombinator[3] R_G

by

$$R_G(\omega) := \frac{1}{|\omega|}(\pi_G.\omega) \otimes (\pi_{\bar{G}}.\omega)$$

with $R_G(0) = 0$. Consider the following dynamical system on $\mathcal{M}(X)$:

$$\dot{\omega} = \sum_{G\subset S} \frac{\varrho_G}{2}\big(R_G - \mathbb{1}\big)\omega. \tag{2.16}$$

This is the infinite population limit of the recombination process (without and with resampling) in the following sense. If we consider $\hat{Z}_t^N := \frac{1}{N}Z_t$ and let $\lim_{N\to\infty} \hat{Z}_0^N = p_0$, then

$$\lim_{N\to\infty} \sup_{s\le t} |\hat{Z}_s^N - p_s| = 0 \tag{2.17}$$

with probability 1, where p_s is the solution of the initial value problem (2.16) with $\omega_0 = p_0$. This is shown in [6] for the special case of single-crossovers, but it is obvious that the proof, which is based on the general law of large numbers

[3]This is a generalisation of the recombinator in [3]. Note that notational similarity is deceptive because G denotes sites here rather than 'links' (the bonds between sites) as in [3].

by Ethier and Kurtz ([19, Thm. 11.2.1], see also [37]), may be generalised to the case of multiple crossovers.

We are now interested in the relationship between (2.13) and the deterministic dynamics. If ω_t follows (2.16), then a (tensor) product of marginal measures $(\pi_{A_1}.\omega_t) \otimes \cdots \otimes (\pi_{A_m}.\omega_t)$ given by a partition of sites as in Theorem 2.12 exhibits the following dynamics:

$$\begin{aligned}
&\frac{d}{dt}\Big((\pi_{A_1}.\omega_t) \otimes \cdots \otimes (\pi_{A_m}.\omega_t)\Big) \\
&\quad = \Big(\pi_{A_1}.\big(\sum_{G\subset S} \frac{\varrho_G}{2}(R_G - \mathbb{1})\big)(\omega_t)\Big) \otimes \big(\pi_{A_2}.\omega_t\big) \otimes \cdots \otimes \big(\pi_{A_m}.\omega_t\big) \\
&\qquad + \big(\pi_{A_1}.\omega_t\big) \otimes \Big(\pi_{A_2}.\big(\sum_{G\subset S} \frac{\varrho_G}{2}(R_G - \mathbb{1})\big)(\omega_t)\Big) \otimes \big(\pi_{A_3}.\omega_t\big) \otimes \cdots \otimes \big(\pi_{A_m}.\omega_t\big) \\
&\qquad + \cdots + \big(\pi_{A_1}.\omega_t\big) \otimes \cdots \otimes \big(\pi_{A_{m-1}}.\omega_t\big) \otimes \Big(\pi_{A_m}.\big(\sum_{G\subset S} \frac{\varrho_G}{2}(R_G - \mathbb{1})\big)(\omega_t)\Big) \\
&\quad = \sum_{j=1}^{m} \sum_{B\subset A_j} \varrho_B\big(\pi_{A_1}.\omega_t\big) \otimes \cdots \otimes \big[\frac{1}{|\omega_t|}\big(\pi_B.\omega_t\big) \otimes \big(\pi_{A_j\setminus B}.\omega_t\big) - \big(\pi_{A_j}.\omega_t\big)\big] \\
&\qquad \otimes \cdots \otimes \big(\pi_{A_m}.\omega_t\big),
\end{aligned} \tag{2.18}$$

with $\varrho_B := \sum_{\substack{H\subset S \\ H\cap A_j = B}} \varrho_H$.

Compare this to (2.13), and only consider summands where $|J| = 1$ and $|K| = 0$ or $|J| = 0$ and $|K| = 1$. According to Remark 2.13, we can understand the corresponding transitions as 'uncorrelated' events at which only one marginal process changes at a given instant. We get the following terms on the right-hand side of (2.13):

- $J = \{j\}, K = \varnothing$:

$$\mathbb{E}\Big[\prod_{i:i\neq j} [A_i]_t \sum_{\substack{G\subset A_j \\ G|A_j}} \frac{\varrho^{M\setminus\{j\}}_{\varnothing,G}}{4N} [G]_t [A_j \setminus G]_t\Big], \tag{2.19}$$

- $J = \varnothing, K = \{j\}$:

$$\mathbb{E}\Big[\prod_{i:i\neq j} [A_i]_t \frac{\varrho^{M\setminus\{j\}}_{\{j\},\varnothing}}{4N}(-1)[A_j]_t N\Big], \tag{2.20}$$

with (cf. (2.14))

$$\varrho^{M\setminus\{j\}}_{\{j\},\varnothing} = \sum_{D\subset A_I} \sum_{\substack{H\subset A_j \\ H|A_j}} \varrho_{H\cup D} = \sum_{\substack{H\subset A_j \\ H|A_j}} \sum_{D\subset A_I} \varrho_{H\cup D} = \sum_{\substack{H\subset A_j \\ H|A_j}} \varrho^{M\setminus\{j\}}_{\varnothing,H}.$$

Adding all terms of kind (2.19) and (2.20), one obtains the analogue of the right-hand side of (2.18). According to Remark 2.13, the summands with $|J \cup K| \geq 2$ correspond to 'correlated' events at which two or more marginal processes change simultaneously.

We may thus conclude that the uncorrelated events correspond to the deterministic equation. We will now show that the correlated events are of lower order and thus tend to zero in the limit $N \to \infty$. To this end, look at (2.12). The right-hand side consists of the terms $\frac{1}{N}[G]_t[\bar{G}]_t$ and $[1, \ldots, n]_t$. They are both of order N, since each individual term $[\ldots]_t$ is of order N (which follows, for example, from (2.17)). Let us look at the derivative of the mean of $\frac{1}{N}[G]_t[\bar{G}]_t$ (cf. (2.13)). The terms with $|I| = 1$ (those belonging to 'uncorrelated' events) will be of order N again, whereas the terms with $I = \varnothing$ are of order 1. By differentiating terms such as $\mathbb{E}\left[\frac{1}{N^2}[A_1]_t[A_2]_t[A_3]_t\right]$ and beyond, the same observation applies: the order of summands belonging to 'correlated' events is less or equal 1, so for the relative frequencies $([1, \ldots, n]_t/N)$ the dynamics of the mean tends to the dynamics of the deterministic model.

2.3.3 Two sites, arbitrary moments

In the case of two sites, the recombination process is rather simple. This mainly relies on the fact that the transition rate for $\langle 1, 2\rangle_t$ is constant, as we have already seen in (2.11). Furthermore, the set of partitions of two sites is trivial. In this special case we can easily show moment closure for arbitrary moments. The simplicity of the setting permits to look at $[1, 2]_t$ itself without considering $\langle 1, 2\rangle_t$ and $(1, 2)_t$. The process $[1, 2]_t^m$ has the following possible transitions (cf. (2.8), (2.9)):

$$[1, 2]_t^m \to ([1, 2]_t + 1)^m \quad \text{at rate} \quad \frac{\varrho_1}{N}([1]_t - [1, 2]_t)([2]_t - [1, 2]_t)$$

and

$$[1, 2]_t^m \to ([1, 2]_t - 1)^m \quad \text{at rate} \quad \frac{\varrho_1}{N}[1, 2]_t(N - [1]_t - [2]_t + [1, 2]_t).$$

Using the binomial theorem and eliminating empty transitions, we obtain for the

derivative of the m-th moment:

$$\begin{aligned}
&\frac{\varrho_1}{N}\sum_{k=0}^{m-1}\mathbb{E}\Big[\binom{m}{k}[1,2]_t^k([1]_t-[1,2]_t)([2]_t-[1,2]_t)\Big]\\
&+\frac{\varrho_1}{N}\sum_{k=0}^{m-1}(-1)^{m-k}\mathbb{E}\Big[\binom{m}{k}[1,2]_t^k[1,2]_t(N-[1]_t-[2]_t+[1,2]_t)\Big]\\
=&\sum_{k=0}^{m-2}\mathbb{E}\Big[\frac{\varrho_1}{N}\binom{m}{k}[1,2]_t^k\{[1]_t[2]_t-2\delta^{(2)}_{m-k}[1,2]_t c+2\delta^{(2)}_{m-k+1}[1,2]_t^2+(-1)^{m-k}[1,2]_t N\}\Big]\\
&+\frac{\varrho_1 m}{N}\mathbb{E}\Big[[1,2]^{m-1}([1]_t[2]_t-[1,2]_t N)\Big],
\end{aligned}$$

with $c := ([1]_t + [2]_t)$ and

$$\delta^{(2)}_{m-k} := \begin{cases} 1 & \text{if} \quad m-k \equiv 0 \bmod 2 \\ 0 & \text{otherwise.} \end{cases}$$

So all emerging terms are moments of order m or less.

2.4 Recombination and Mutation

We now want to add mutation to our process. Let us first look at the process with mutation alone, e.g. $b = \varrho_G = 0$. By Lemma 2.10, the derivative of $\mathbb{E}\Big[[1,\dots,n]_t\Big]$ is:

$$\sum_{j\in S}\Big(\sum_{y\in X_j}\mu^j_{yx_j}[1,\dots,j-1,j+1,\dots,n]_t-\sum_{y\in X_j\setminus\{x_j\}}\mu^j_{x_jy}[1,\dots,n]_t\Big).$$

So, it only consists of linear terms and marginal processes. When we consider a product of marginal processes given by a partition of sites as in the previous section, we have, due to the fact that mutation only acts on single sites independently of others:

$$\begin{aligned}
\frac{d}{dt}\mathbb{E}\Big[\prod_{\ell\in M}[A_\ell]_t\Big] =&\mathbb{E}\Big[\sum_{\ell\in M}\Big(\prod_{i\in M\setminus\{\ell\}}[A_i]_t\sum_{j\in A_\ell}\sum_{y\in X_j}\mu^j_{yx_j}[A_\ell\setminus\{j\}]_t\Big)\\
&-\sum_{\ell\in M}\Big(\prod_{i\in M\setminus\{\ell\}}[A_i]_t\sum_{j\in A_\ell}\sum_{y\in X_j\setminus\{x_j\}}\mu^j_{x_jy}[A_\ell]_t\Big)\Big].
\end{aligned}$$

What happens when we add recombination? Let F_M and F_R, respectively, be the 'mean rate of change functions' from Lemma 2.10 for $\prod_{\ell\in M}[A_\ell]_t$ from the

process with solely mutation and recombination, respectively. Since mutation and recombination proceed independently, the respective function F_{RM} for the recombination-mutation process is then just $F_R + F_M$ and according to Lemma 2.10 we have for $\frac{d}{dt}\mathbb{E}\Big[\prod_{\ell\in M}[A_\ell]_t\Big]$:

$$\begin{aligned}\mathbb{E}\Bigg[&\sum_{\substack{(I,J,K)\in\mathcal{P}\\ I\neq M}}\prod_{i\in I}[A_i]_t(-1)^{|K|}\sum_{\tilde{K}\subset K}\sum_{\substack{G\subset A_J\\ G|\{A_j\}_{j\in J}}}\frac{\varrho^I_{K,G}}{4N}[A_{\tilde{K}}\cup G]_t[A_{K\setminus\tilde{K}}\cup G^c]_t\\ &+\sum_{\ell\in M}\Big(\prod_{i\in M\setminus\{\ell\}}[A_i]_t\sum_{j\in A_\ell}\sum_{y\in X_j}\mu^j_{yx_j}[A_\ell\setminus\{j\}]_t\Big)\\ &-\sum_{\ell\in M}\Big(\prod_{i\in M\setminus\{\ell\}}[A_i]_t\sum_{j\in A_\ell}\sum_{y\in X_j\setminus\{x_j\}}\mu^j_{x_jy}[A_\ell]_t\Big)\Bigg].\end{aligned}$$

So, the arising terms are the same as in the pure recombination process plus linear terms. It is therefore clear that we have moment closure here as well.

2.5 Recombination and Resampling

In this section, we set $b > 0$ and the mutation rates zero again, so we look at the Moran model with recombination and resampling only. At first glance, one may think that resampling has no effect on the expectation since the process with resampling alone has a constant mean. Indeed, the first derivative of the mean looks the same as in the pure recombination case:

$$\frac{d}{dt}\mathbb{E}\big[[1,2]_t\big] = \frac{\varrho_1}{N}\mathbb{E}\big[[1]_t[2]_t - [1,2]_tN\big]. \tag{2.21}$$

However, due to resampling, the one-site marginal processes are no longer constant, so we do not have instantaneous moment closure any more (cf. (2.5): at a resampling event, the frequency of alleles may change). The derivative of their product is obtained after an elementary but lengthy calculation:

$$\frac{d}{dt}\mathbb{E}\big[[1]_t[2]_t\big] = \frac{b}{N}\mathbb{E}\big[[1,2]_tN - [1]_t[2]_t\big]. \tag{2.22}$$

We obtain a finite linear system of differential equations, namely (2.21) and (2.22). In particular, we obtain

$$\frac{d}{dt}\mathbb{E}\big[[1,2]_tN - [1]_t[2]_t\big] = -\big(\frac{\varrho_1}{N}+\frac{b}{N}\big)\mathbb{E}\big[[1,2]_tN - [1]_t[2]_t\big] \tag{2.23}$$

with the obvious exponential solution. The term $[1,2]_t N - [1]_t[2]_t$ is a correlation function, a so called linkage disequilibrium, which is widely used in population genetics. We see that both, recombination and resampling, reduce correlations between sites.

For more than two sites, exact moment closure can no longer be established. To make this plausible, we will only present the derivative of $\mathbb{E}\big[[1]_t[2]_t[3]_t\big]$ (again, the calculation is elementary but lengthy), which is a term that emerges in the derivatives of the process on three sites due to recombination:

$$\begin{aligned}\frac{d}{dt}\mathbb{E}\big[[1]_t[2]_t[3]_t\big] = \frac{b}{N}\mathbb{E}\big[&[1]_t[2,3]_t N + [2]_t[1,3]_t N + [3]_t[1,2]_t N + [1]_t[1,2]_t[1,3]_t \\ &+ [2]_t[1,2]_t[2,3]_t + [3]_t[1,3]_t[2,3]_t - 3[1]_t[2]_t[3]_t \\ &- [1]_t^2[1,2,3]_t - [2]_t^2[1,2,3]_t - [3]_t^2[1,2,3]_t\big].\end{aligned}$$

The last three terms are quadratic and it is clear that further differentiating will lead to terms such as $[1]_t^3[2]_t[3]_t$ whose derivative will contain moments of even higher order. Thus, the interaction between recombination and resampling destroys moment closure.

2.6 Conclusion

In this chapter, we have extended the single-crossover Moran model from [6] to include general recombination. The dynamics of the expectation under general recombination becomes significantly more complicated. In particular, it now deviates from the dynamics in the infinite population model. The reason is the loss of independence of certain marginal processes.

As is usual with nonlinear processes, the dynamics of a given moment requires higher moments. Nevertheless, in this case after a finite number of steps no additional terms emerge. This is due to the fact that the arising processes may in each step be described by a partition of sites. When mutation is included, this exact moment closure persists, but the arising processes can no longer be described by a partition of sites. Altogether, we have an exception to the rule that the dynamics of the moments of nonlinear processes lead to infinite hierarchies of ODE's.

This exact moment closure gets lost when we extend the model to include genetic drift (i.e., resampling). This is, of course, disappointing since the Moran model with recombination alone is mathematically interesting, but of limited biological value. Nevertheless, the resulting hierarchy of moments might be interesting to analyse with respect to the various possibilities of approximate moment closure. Furthermore, the arising terms such as $\mathbb{E}\big[\prod_{\ell\in M}[A_\ell]_t\big]$ are of considerable interest in population genetics beyond this moment closure procedure since they are the

building blocks of the linkage disequilibria [12] that are so important in population genetics (compare (2.23) for the simplest example).

Chapter 3

Towards a particle picture behind the common ancestor distribution

In this chapter, we consider a Moran model with selection and mutation. We first present some results of Fearnhead and Taylor that we will prove in another way. Since this provides us with a better understanding of the Moran model we may then propose a new particle model in the case without mutation. In opposite to the situation in the previous chapter, here, the integration of mutation makes the situation substantially more complicated.

3.1 Moran model with selection and genealogies

We consider a population of N individuals of type $i \in S = \{0, 1\}$. Individuals of type 1 reproduce at rate 1, individuals of type 0 are favoured by selection, they reproduce at rate $1 + s_N$, $s_N \geq 0$. In a reproduction event, the new individual replaces a randomly chosen other one.

Additionally, mutation events may occur. An individual mutates at rate u_N. With probability ν_0, it will then be of type 0, with probability $\nu_1 := 1 - \nu_0$, it will be of type 1. So, u_N may be considered as the overall mutation rate. Note, that the mutation is independent of the original type. Obviously, the population size N remains constant.

This model is called a Moran model due to its properties continuous time and constant population size. Besides the Wright-Fisher model, which is formulated in discrete time, it is one of the standard models in population genetics. It can be considered as an interacting particle system and allows for a graphical representation (see Fig. 3.1). Another important class of models in population genetics are branching processes. In contrast to the Moran model and the Wright Fisher model, the population size varies, see [24].

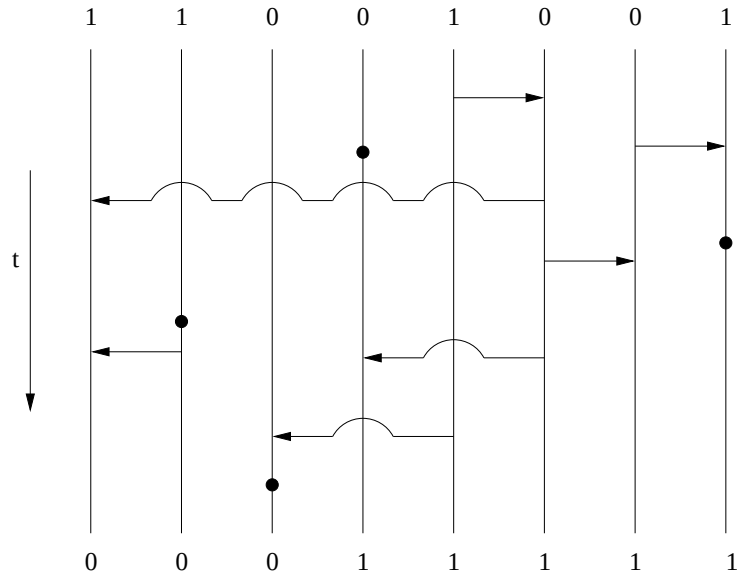

Figure 3.1: Moran model with selection. The lines represent single individuals, their initial types are denoted at the top, their final types at the bottom. The origin of an arrow marks the birth of an individual, whereas the individual at the tip is replaced. Mutation events are marked with dots.

Consider the process $(Y_t^N)_{t\geq 0}$ where Y_t^N is the number of individuals of type 0 at time t. It is a Markov jump process. For $Y_t^N = k$, $k = 0, \ldots, N$, the following transitions are possible:

$$\begin{aligned} &k \to k+1 \quad \text{at rate } (N-k)\big(u_N\nu_0 + (1+s_N)\frac{k}{N}\big), \\ &k \to k+1 \quad \text{at rate } k\big(u_N\nu_1 + \frac{N-k}{N}\big). \end{aligned} \tag{3.1}$$

A standard technique in population genetics is to consider the diffusion limit. Hereby, the population size tends to infinity and additionally the time is rescaled by the population size. Since stochastic fluctuations remain possible in the diffusion limit, it is a reasonable approximation for finite populations and permits the application of results for diffusion processes. To this purpose, we define the normalised process $(X_t^N)_{t\geq 0}$ by

$$X_t^N := \frac{1}{N} Y_{Nt}^N. \tag{3.2}$$

The parameters u_N and s_N are assumed to be scaled by the population size such that

$$Nu_N \to \theta \quad \text{and}$$

$$Ns_N \to \sigma,$$

for some nonnegative θ and σ.

Then the sequence of processes $(X_t^N)_{t\geq 0}$ converges to the diffusion limit $(X_t)_{t\geq 0}$ for $N \to \infty$ in distribution [37]. It is characterised by its generator [17, 37]

$$A\phi(x) = x(1-x)\phi''(x) + \big(\theta\nu_0(1-x) - \theta\nu_1 x + \sigma x(1-x)\big)\phi'(x). \tag{3.3}$$

where $\phi \in \mathcal{C}^2([0,1])$.

The diffusion limit has a well-known stationary distribution, its density is given by Wright's formula

$$f(x) := \frac{1}{C} x^{\theta\nu_0 - 1}(1-x_1)^{\theta\nu_1 - 1} e^{\sigma x}, \tag{3.4}$$

with normalising constant C.

For an overview of diffusion theory in the context of population genetics, see [17, 20].

Remark 3.1. As usual in the relevant literature, we will omit the index N in the parameters of the Moran model in the latter, e.g. write s instead of s_N.

Remark 3.2. The stationary distribution (3.4) may be deduced from the diffusion process. Alternatively, one may derive (3.4) directly from the stationary distribution of the Moran model which is also known explicitly [17, Ch. 7 and 8].

Remark 3.3. A finite particle representation in the diffusion limit is provided by the construction of the look-down process (originated by Donnelly and Kurtz [15], also see [18, Ch. 5] for a review). This construction relies on assigning levels to each particle. The analogue to birth-death events are look-down events. In the absence of selection, only particles from higher levels may be replaced by particles from lower levels. Thus, the particle of the lowest level will be ancestral to all particles after some finite time.

When selection is incorporated, one can no more conserve the property that the death always takes place in the higher level because the death rates depend on the type and the background. The ancestral particle may be found anyway but the procedure needs to enlarge the state space if selective events take place. This idea of the construction to be used is similar to the one of the ancestral selection graph defined by Krone and Neuhauser [36] which will be discussed later.

Remark 3.4. In the formulation of the Moran model, the mutation rate often contains a factor 2. This factor is included to make sure that the Moran model shares the same diffusion limit as the Wright-Fisher model. Since we do not consider the Wright-Fisher model in this thesis, we do not stick to this convention.

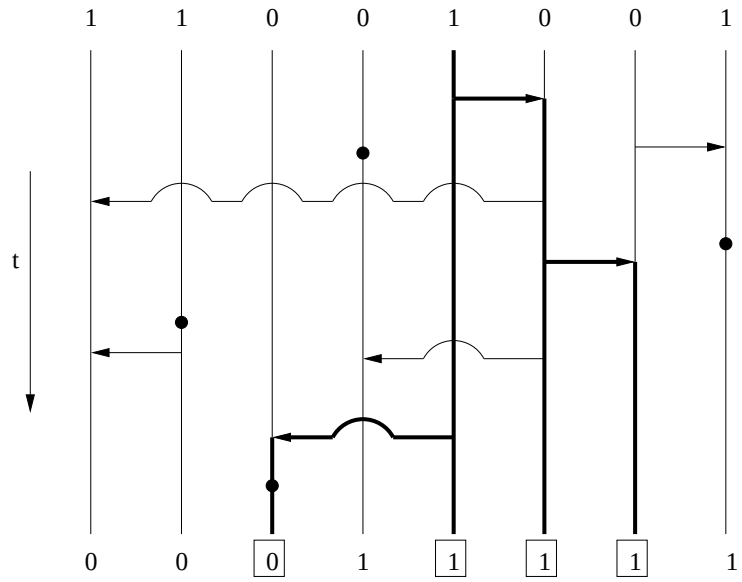

Figure 3.2: We grab a sample out of the population (boxed individuals) and follow their genealogy backwards in time (fat lines). The fifth individual is the most recent common ancestor.

3.1.1 Genealogies

This section aims to introduce some theory that is necessary to understand the issues handled in this chapter and to present the main previous results. The presentation is mainly a summary of a review paper [5]. First, we want to explain the construction of genealogies.

We are given a realisation of the Moran model. We take a sample of individuals and are interested in their joint history. We can construct a *genealogical tree* in the following way: we trace the lines of the individuals from the sample backwards in time. Whenever a line is hit by the tip of an arrow we look at the origin of the arrow. If it belongs to the sample the lines merge. We call this a *coalescent event*. In this way, we can extract a *genealogy*, which is the genealogical tree together with the types along the branches (see Figure 3.1.1). The root of the tree is the *most recent common ancestor (MRCA)* of the sample, the individual which is ancestral to the whole sample.

But how can we sample genealogies for a sample of n individuals without a given realisation of the Moran model? We address this question in this section.

The neutral case

A simple situation is present when selection is not incorporated in the model, that means, when we set $s = 0$. In this neutral case, individuals reproduce independently. Thus, we may first construct the genealogical tree of n individuals while ignoring their types and then introducing the types afterwards.

The leaves are obviously given by the sample individuals. But when does the first coalescent event occur? It emerges when a sample individual reproduces and its offspring replaces another sample individual. Since each individual reproduces at rate 1, n individuals are present and a sample individual is hit at probability $(n-1)/N$, the rate is

$$\frac{n(n-1)}{N}.$$

So, after some exponentially distributed waiting time a coalescent event happens and two arbitrary lines merge. This procedure is continued until there is only one line left. In account of a scaling of time by factor N, this procedure corresponds to the well-known *Kingman coalescent* [34, 35].

In the second step, we assign types along the branches. For this, we assume that the population is *stationary*. We draw the type of the MRCA out of the stationary distribution which is also known in the finite population case which in the neutral case is obviously simply given by the mutation rates. Usually, one constructs these genealogies for an underlying infinite population in the diffusion limit. The stationary distribution is then given by (3.4). Whenever a line branches, the individuals inherit the parental type. In addition, mutation events happen along the branches in forward time at rates $u\nu_i$, and $\theta\nu_i$, $i = 0, 1$, in the rescaled version, respectively.

3.1.2 The ancestral selection graph

The construction of the coalescent in the neutral case relies on the fact that the individuals reproduce independently of their types. So, we can introduce types *after* the construction of the tree topology. This is crucial since offspring individuals must be of the parental type and thus only the merge of edges of the same type may be allowed.

In the nonneutral case the emergence of branches in the tree topology is no more independent of the type of the corresponding individual. Krone and Neuhauser [36, 43] developed a construction that copes with this problem. The key idea is to differ artificially between neutral and selective events, and to decide after the introduction of types, whether selective branches maintain or not. Again, one splits the mutational from the reproduction process. The procedure goes in three steps:

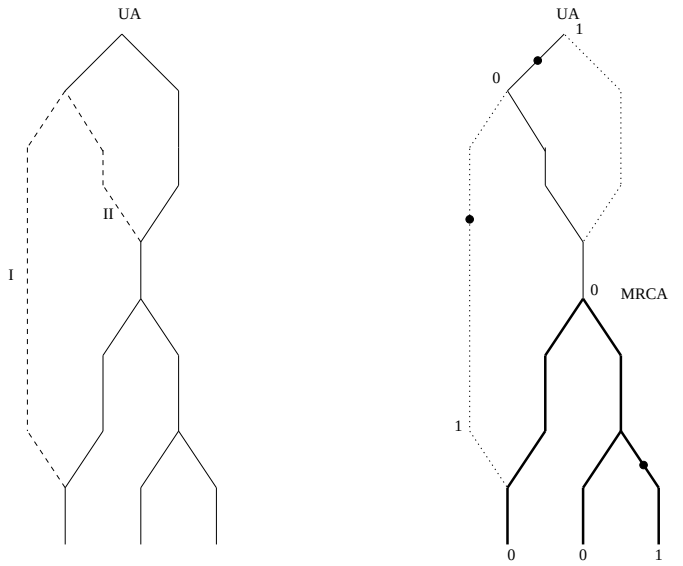

Figure 3.3: Construction of the ASG. Left: Graph without types. Dashed branches are incoming. Right: Introduction of types and mutation events. Branch I is unfit and thus virtual. Branch II is fit and so real. The MRCA deviates from the UA.

- Construction of the graph topology without types. We start with n individuals. Again, two branches coalesce at rate $n(n-1)$ due to the same reasoning as in the neutral case. Additionally, we consider selection. At this step, selective events are possible in all lines and they cause the splitting of a branch. It happens at rate σn. We label one branch the incoming branch and the other the continuing branch. After some finite time the sample size will almost surely be one. At this point of time we stop and call this individual the ultimate ancestor.

- We draw the type of the ultimate ancestor from the stationary distribution (3.4) and run the mutation process along the graph. At a coalescent event, the offspring inherits the parental type as in the neutral case. At branching events, we must decide, which line is the the parental one. We proceed according to the following rule: If the incoming branch is fit, then it is parental. If it is unfit, the continuing line is parental. We call the parental line the real line, the other line is called virtual line.

- We extract the final genealogy by taking all real lines and rejecting all virtual lines.

We briefly want to motivate the transition rates of the birth-death process that constructs the graph topology. The death rate and the mutation rates are ex-

plained in the same way as in the neutral case. Birth events arise at selective events. In the Moran model, the rate that a sample individual undergoes a selective event and hits an individual that does not yet belong to the graph is $ns(N-n)$. Since the probability that a sample individual hits another line in the graph is zero in an infinite population, we neglect this event. Then, in the diffusion limit the birth rate is the limit for $N \to \infty$.

Remark 3.5. The construction of the topology of the graph in step one is independent of the types of the respective particles. Thus, the type of the first single particle is not distinguished in any way. This is the reason that it is appropriate to determine the type of the ultimate ancestor from the stationary distribution (3.4) of the forward process. One should keep in mind that the ultimate ancestor is not necessarily the most recent common ancestor and so, the distributions may deviate. For a rigorous reasoning see [16, Lemma 8.1 and Thm. 8.2].

Remark 3.6. A central aim of this chapter is the derivation of the stationary distribution of the type of the common ancestor. In the neutral case this is simply the stationary distribution of the forward process.

Things become more complicated when selection takes place. It is a reasonable conjecture that the most recent common ancestor of a sample or the whole population is fitter than a randomly chosen individual because it has descendants in all generations. An individual that is an ancestor of the most recent common ancestor is obviously ancestral to the whole sample and population, respectively. Thus, the stationary distribution of a process which describes the type of the common ancestor describes the distribution of the most recent common ancestor, too [21].

3.2 Previous results

3.2.1 Graphical representation

The ancestry of a single individual

Following the reasoning of Remark 3.6, we want to give a procedure how to construct the genealogy of one individual of *given* type which was presented by Fearnhead [21]. Stephens and Donnelly also developed an algorithm that gives the genealogy of known samples of arbitrary (finite) size [49, Algorithm 3.2], but in the context of this chapter we are only interested in the simpler case of sample size 1.

Exchangeability and notation

We describe samples from our model by a summary statistics, where a vector (n_0, n_1) means that our sample consists of n_0 individuals of type 0 and n_1 individuals of type 1, respectively. Nevertheless, this representation ignores any kind of ordering of the sample. Due to exchangeability, the probability of each ordered sample with the same summary statistics is equal. Though some abuse of notation, we describe with $p(\mathbf{n} = (n_0, n_1))$ the probability of an *ordered* sample from a stationary population. It is given by

$$p(\mathbf{n}) = \frac{1}{C} \int_0^1 x^{n_0+\theta\nu_0-1}(1-x)^{n_1+\theta\nu_1-1} \exp(\sigma x) dx.$$

Within this notation, we always imply that a sample $\mathbf{n}$ consists of n individuals. We write

$$p(i|\mathbf{n}) = \frac{p(\mathbf{n}+\mathbf{e}_i)}{p(\mathbf{n})}$$

for the conditional probability that the $n+1$st individual in the sample is of type i given that the first n individuals have the summary statistics $\mathbf{n}$.

Transitions and rates

We follow the line of one individual backwards in time. Thus, in the first step we draw its type from Wright's formula, $p(\mathbf{e}_i)$. Then we follow its ancestry backwards in time. The mutation processes run backward in time. Since branching events may occur, and thus the number of lines may rise, we must describe the time evolution of a *sample*. Our states are in

$$E = \{\big(i, (n_0, n_1)\big),\ i \in \{0, 1\},\ n_0, n_1 \in \mathbb{N}_0\},$$

where the first entry denotes the type of the real branch and (n_0, n_1) are the multiplicities of virtual branches with the corresponding types.

The following events may occur. Fearnhead [21] gives the rates which are calculated via time reversal ([44, Ch. 3.7], [11, Ch.8.5]. See [5] for a detailed presentation. If the current state is $\big(i,(n_0,n_1)\big)$, then the possible events and their rates are

- Coalescence of two branches of type j at rate $(n_j+\delta_{ij})(n_j+\delta_{ij}-1)p(\mathbf{n}+\mathbf{e}_i-\mathbf{e}_j)/p(\mathbf{n}+\mathbf{e}_i)$. The new state is $(i,\mathbf{n}-\mathbf{e}_j)$.
- Mutation of the real branch at rate $\theta\nu_i p(\mathbf{n}+\mathbf{e}_j)/p(\mathbf{n}+\mathbf{e}_i)$[1] with new state $(j,\mathbf{n})$.
- Mutation of a virtual branch at rate $n_j\theta\nu_j p(\mathbf{n}+\mathbf{e}_i-\mathbf{e}_j+\mathbf{e}_k)/p(\mathbf{n}+\mathbf{e}_i)$ with new state $(i,\mathbf{n}+\mathbf{e}_k-\mathbf{e}_j)$.
- Branching to an unfit incoming branch with new state $(i,\mathbf{n}+\mathbf{e}_1)$ at rate $(n_j+\delta_{ij})\sigma p(\mathbf{e}_i+\mathbf{e}_1)/p(\mathbf{n}+\mathbf{e}_i)$. Here, the continuing type is parental.
- Branching to a fit incoming branch with new state $(i,\mathbf{n}+\mathbf{e}_k)$ at rate $(n_0+\delta_{i0})\sigma p(\mathbf{n}+\mathbf{e}_i+\mathbf{e}_k)/p(\mathbf{n}+\mathbf{e}_i)$. Here, the incoming branch is parental.

The common ancestor process and its stationary distribution

The procedure above serves to construct a genealogical graph, coming from a sample of size one. Since we are only interested in the distribution of the common ancestor, we do not necessarily need the full genealogy. Indeed, a simplification is possible. The Markov property is still maintained when we remove all fit virtual lines. This is due to the following reasoning. Take a closer look at the possible transitions above. The rates of transition 3 and 5 may be rewritten as follows

$$\frac{n_j\theta\nu_j p(\mathbf{n}+\mathbf{e}_i-\mathbf{e}_j)}{p(\mathbf{n}+\mathbf{e}_i)}p(k|\mathbf{n}+\mathbf{e}_i-\mathbf{e}_j)$$

and

$$(n_0+\delta_{i0})\sigma p(k|\mathbf{n}+\mathbf{e}_i).$$

In both cases the new type is drawn from the conditional probability given the composition of the remaining population. But this means that the type of these individuals cannot affect the genealogy of the remaining sample, otherwise there should be more dependence. This reasoning relies on [21, Theorem 1]. Thus we may remove transitions 3 and 5 which are the only transitions where fit virtual lines may emerge. We obtain the following dynamics for this so called *common ancestor process*. See [21, 5] for a detailed presentation.

[1] This rate deviates from the one in [21] by a factor 2, see Remark 3.4.

Again, given we are in state $\big(i,(n_0,n_1)\big)$, then the possible events and their rates are

- Mutation of the real branch at rate $\theta\nu_i p(\mathbf{n}+\mathbf{e}_j)/p(\mathbf{n}+\mathbf{e}_i)$ with new state $(j,\mathbf{n})$.
- Removal of a virtual branch at a coalescence event or by mutation at rate $((n_1+\delta_{i1})(n_1+\delta_{i1}-1)+\theta\nu_1 n_1)p(\mathbf{n}+\mathbf{e}_i-\mathbf{e}_1)/p(\mathbf{n}+\mathbf{e}_i)$ with new state $(i,\mathbf{n}-\mathbf{e}_1)$.
- Branching at rate $\sigma(n_1+1)p(\mathbf{n}+\mathbf{e}_i+\mathbf{e}_1)/p(\mathbf{n}+\mathbf{e}_i)$.

Starting in $(i,(0,0))$ with probability $p(\mathbf{e}_i)$, Fearnhead defines this process as the *common ancestor process*. Its stationary distribution is given by the following theorem.

Theorem 3.7 ([21], Lemma 1, Theorem 3)**.** *Let $\theta>0$, $\sigma\geq 0$, $\nu_0>0$ and $0<\nu_1<1$. For $k\in\mathbb{N}$ define $\lambda_1^{(k)},\dots,\lambda_{k+1}^{(k)}$ by $\lambda_{k+1}^{(k)}=0$ and*

$$\lambda_{i-1}^{(k)}=\frac{\sigma}{i+\theta+\sigma-(i+\theta\nu_1)\lambda_i^{(k)}}. \tag{3.5}$$

The limits

$$\lambda_i=\lim_{k\to\infty}\lambda_i^{(k)} \tag{3.6}$$

exist and satisfy $0\leq\lambda_i\leq 1$ for all i and the stationary distribution of the common ancestor process is

$$\pi_F(j,\mathbf{n})=\begin{cases}\Big(\prod_{i=1}^{n_1}\lambda_i\Big)p(\mathbf{n}+\mathbf{e_0}), & \text{if } j=0,\\ \Big(\prod_{i=1}^{n_1}\lambda_i\Big)(1-\lambda_{n+1})p(\mathbf{n}+\mathbf{e_1}), & \text{if } j=1.\end{cases} \tag{3.7}$$

Summing over the frequencies of virtual lines then gives the distribution (π_0,π_1) of the common ancestor:

$$\pi_0=\sum_{\mathbf{n}}\pi_F(0,\mathbf{n}). \tag{3.8}$$

Fearnhead proves this theorem by a verification of the stationarity condition '$\pi Q=0$'. We will find a different way to this result which gives some insight into the particle picture behind the scenery.

3.2.2 The diffusion approximation - structured coalescent

Another approach to the common ancestor distribution was pursued by Taylor [50]. He considers a diffusion process arising from a Wright-Fisher model. This model is closely related to the Moran model. One considers a population of N individuals, each individual is of type 0 or 1. The model is in discrete time, the next generation is drawn from the current. Each new individual is drawn independently of the others from the current generation, where individuals of type 1 have weight 1, individuals of type 0 have weight $1+s$. Then, the new individual is a nonmutant copy at probability $1-u$, with probability $u\nu_0$ it is of type 0, and with probability $u\nu_1$ of type 1 respectively.

In the diffusion limit the frequency of fit individuals is then described by a Markov process with generator[2]

$$\hat{A}\phi(x) = \frac{1}{2}x(1-x)\phi''(x) + \big(\theta\nu_0(1-x) - \theta\nu_1 x + \sigma x(1-x)\big)\phi'(x).$$

where $\phi \in \mathcal{C}^2([0,1])$. This is basically the same generator as in (3.3), they only differ by a factor $1/2$. This is due to the fact that the Moran model coalesces twice as often as the Wright-Fisher model, cf [17, Sect. 1.5], Remark 3.4.

Corresponding to this diffusion, Taylor [50] considers the coalescent process introduced by Kaplan et al [31]. It is called the *structured coalescent* and characterized by its generator [9]

$$\begin{aligned}\tilde{G}\phi(n_0,n_1,x) = &\binom{n_0}{2}\frac{1}{x}\big(\phi(n_0-1,n_1,x) - \phi(n_0,n_1,x)\big)\\ &+ \binom{n_1}{2}\frac{1}{1-x}\big(\phi(n_0,n_1-1,x) - \phi(n_0,n_1,x)\big)\\ &+ n_0\theta\nu_0\frac{1-x}{x}\big(\phi(n_0-1,n_1+1,x) - \phi(n_0,n_1,x)\big)\\ &+ n_1\theta\nu_1\frac{x}{1-x}\big(\phi(n_0+1,n_1-1,x) - \phi(n_0,n_1,x)\big) + \hat{A}\phi(n_0,n_1,x),\end{aligned}$$

where $\phi(n_0,n_1,\cdot) \in \mathcal{C}^2([0,1])$ for each $n_0,n_1 \in \mathbb{N}$. In [9], Barton et al. derive the generator from a Moran model (that additionally incorporates recombination) because in opposite to the Wright-Fisher model the frequency of fit alleles in the Moran model has a reversible invariant measure which makes the calculations significantly easier. Nevertheless, the genealogies of a sample from the population given by the two models coincide in the diffusion limit because the diffusion

[2]Taylor even considers density-dependent selection, this means, the selection coefficient is a function of x. The general results about the uniqueness of the solution are valid under this generalization. However, he does not give an explicit formula for the stationary distribution in this general case.

approximations for both models are the same and in both models the particles are not ordered. Thus, the generator derived from the Moran model is also valid for the structured coalescent coming from a Wright-Fisher diffusion (under consideration of the factor 2).

They show that the stochastic process with this generator exists and is unique. Hereby, the existence is not obvious since the coalescence and migration rates may become unbounded if the frequency of fit alleles tends to zero or one (Chapter 4 in [9] and Lemma 2.1 in [50]).

Only interested in samples of size 1, one can consider a bivariate process in the space $E = (\{0\} \times (0,1]) \cup \{1\} \times [0,1))$, where the first coordinate gives the type of the ancestral lineage. The generator of the structured coalescent with sample size 1 then is

$$\begin{aligned} \tilde{G}\phi(0,x) &= \theta\nu_0 \frac{1-x}{x}\big(\phi(1,x) - \phi(0,x)\big) + \hat{A}\phi(0,x) \\ \tilde{G}\phi(1,x) &= \theta\nu_1 \frac{x}{1-x}\big(\phi(0,x) - \phi(1,x)\big) + \hat{A}\phi(1,x) \end{aligned} \tag{3.9}$$

for functions ϕ that are twice continuously differentiable on E and have compact support.

Taylor examines the same question as Fearnhead, that is the distribution of the common ancestor. To this end, he introduces the conditional probability h, where $h(x)$ is the probability that in a population with a frequency of x fit individuals, the common ancestor is fit. The stationary measure of the structured coalescent as in (3.9) is then given by

$$\begin{aligned} \pi_T(0,dx) &= \pi(0,x)dx = h(x)f(x)dx \\ \pi_T(1,dx) &= \pi(1,x)dx = (1-h(x))f(x)dx, \end{aligned}$$

where f is the density of the stationary measure from (3.4). Taylor shows that h is given by a boundary value problem

$$\begin{aligned} &\hat{A}h(x) - \Big(\theta\nu_0 \frac{1-x}{x} + \theta\nu_1 \frac{x}{1-x}\Big)h(x) = -\theta\nu_1 \frac{x}{1-x}, \\ &h(0) = 0, \quad h(1) = 1. \end{aligned} \tag{3.10}$$

He proves the following lemma concerning existence, uniqueness and smoothness of the solution of this boundary value problem.

Lemma 3.8 ([50], Lemma 2.3.)**.** *There exists a unique solution to the bvp* (3.10) *which is holomorphic on* $(0,1)$ *and whose first derivative can be continuously extended to* $[0,1]$. □

Furthermore, Taylor shows that the stationary distribution of the retrospective process is unique.

Taylor defines the *common ancestor process* as the time-reversal of the retrospective process. Its generator is then given by the adjoint of $\tilde{G}$ with respect to $\pi(z, x)$. He shows that the solution of the corresponding martingale problem exists and is unique [50, Prop. 2.7].

Taylor states the following lemma which connects his diffusion and Fearnhead's graphical representation.

Lemma 3.9 ([50], Lemma 4.1)**.** *Let* $(\lambda_n)_{n\geq 1}$ *be the sequence defined in* (3.5) *and* (3.6)*. Then,*

$$h(x) = x + x\sum_{n=1}^{\infty}(1-x)^n\Big(\prod_{i=1}^{n}\lambda_i\Big). \tag{3.11}$$

□

Thus, the marginal distributions of Taylor's and Fearnhead's solutions coincide:

$$\begin{aligned}\pi_0 &= \int_0^1 h(x)f(x)dx = \int_0^1 xf(x)dx + \int_0^1\sum_{n=1}^{\infty}x(1-x)^n f(x)dx\Big(\prod_{i=1}^{n}\lambda_i\Big)\\ &= \sum_{n_1=0}^{\infty}\Big(\prod_{i=1}^{n_1}\lambda_i\Big)p(\mathbf{n}+\mathbf{e}_0) = \sum_{n_1=0}^{\infty}\pi_F(0,\mathbf{n}).\end{aligned}$$

This coincidence of both approaches is intuitively clear because they are derived from the same basic models.

3.3 Towards the particle model behind the diffusion

3.3.1 Recursion for h

In this section, we want to find a new approach to the conditional probability h, and $(h_k^N)_{k=0,\dots,N}$, respectively. The latter is defined as the probability that after some finite time an individual of type 0 will be the ancestor of the whole population given that there are k individuals of type 0 in a current population of size N. To clarify the presentation, we will omit the upper index N in the further course. We follow the presentation in [33]. The ansatz for the calculation of h_k is a first step analysis argument. The following lemma which is often used implicitly clarifies this approach [11].

Lemma 3.10. [First-Step Analysis] *Let $X = (X_t)_{t\geq 0}$ be a time-homogeneous Markov process with countable state space E and $A \subset E$ closed (in the sense of Markov processes) subset in E. Let T_x be the residence time in state $x \in E$. Then*

$$\begin{aligned}&\mathbb{P}(X \text{ absorbs in } A \mid X_0 = x)\\ &\quad = \sum_{y\neq x} \mathbb{P}(X_{T_x} = y \mid X_0 = x) \cdot \mathbb{P}(X \text{ absorbs in } A \mid X_0 = y).\end{aligned} \tag{3.12}$$

□

We apply this lemma to the probability h_k. To this end, we follow the offspring of a sample.

Definition 3.11. [[33], offspring process] *Consider a given sample of n_0 individuals of type 0 and n_1 individuals of type 1, respectively. Let $\mathcal{O}_0 = (n_0, n_1)$ and $\mathcal{O}_t$ be the composition of the offspring of the sample at time $t > 0$.*

Then, we call $(\mathcal{O}_t, Y_t)_{t\geq 0}$ with Y_t as defined in (3.1) the offspring process which is a Markov process with state space

$$E_{\mathcal{O}} = \{((m_0, m_1), k),\ m_0, m_1 \in \mathbb{N},\ m_0 + m_1 \leq N,\ 0 \leq k \leq N\}.$$

Remark 3.12. Since in the nonneutral model the reproduction rates depend on the frequency of fit alleles, it is necessary to record the background frequency of fit alleles in order to obtain the Markov property.

This process has one absorbing state, namely the extinction of the sample. Furthermore, all states of the offspring process with offspring size N form a closed set $\mathcal{A}$:

$$\mathcal{A} := \{((m_0, m_1), k) \in E,\ m_0 + m_1 = N\}.$$

So, let $(\mathcal{O}_0, Y_0) = ((k, 0), k)$. Obviously, the probability that the offspring process gets trapped in $\mathcal{A}$ is exactly the conditional probability that after some finite time a fit individual will be ancestral to the whole population, that is h_k.

Thus, we are in the framework of Lemma 3.10 and may apply it to the the offspring process. We consider all possible transitions given we are in state $((k,0),k)$, calculate their rates and the absorption probability at the new state [33].

1. Replacement of an individual of type 1: $((k,0),k) \to ((k+1,0),k)$ happens at rate $(1+s)\frac{k(N-k)}{N}$. The absorption probability in this state is h_{k+1}.
2. Replacement of an individual of type 0: $((k,0),k) \to ((k-1,0),k-1)$ happens at rate $\frac{k(N-k)}{N}$, the absorption probability is h_{k-1}.
3. Mutation of an individual of type 0: $((k,0),k) \to ((k-1,0),k-1)$ happens at rate $u\nu_1 k$. The probability that any single individual of type 1 is after some finite time the common ancestor of the entire population is $1-h_{k-1}$ given the process is in state $((k-1,0),k-1)$. For a selected individual this probability is then $\frac{1-h_{k-1}}{N-k+1}$. So, the absorption probability for the sample individuals then is
$$h_{k-1} + \frac{1-h_{k-1}}{N-k+1}.$$
4. Mutation of an individual of type 1: $((k,0),k) \to ((k,0),k+1)$ happens at rate $u\nu_0(N-k)$. The absorption probability of the sample is the probability that any individual of type 0 is the common ancestor of the population in the future apart from the selected individual which does not belong to the sample, this is $(1-\frac{1}{k+1})h_{k+1}$.

With this and Lemma 3.10 we obtain

Theorem 3.13. [33] *The probability $(h_k)_{k=0,\dots,N}$ fulfills the following recursion for $k = 1,\dots,N-1$*

$$\begin{aligned} Rh_k =& (1+s)\frac{k(N-k)}{k}h_{k+1} + \frac{k(N-k)}{N}h_{k-1} \\ &+ u\nu_0(N-k)\Big(h_{k+1} - \frac{h_{k+1}}{k+1}\Big) + u\nu_1 k\Big(h_{k-1} + \frac{1-h_{k-1}}{N-k+1}\Big) \end{aligned} \tag{3.13}$$

with $R = \Big((2+s)\frac{k(N-k)}{N} + u\nu_0(N-k) + u\nu_1 k\Big)$ and boundary conditions $h_0 = 0$, $h_N = 1$. □

In the neutral case, $h(x) = x$, and $h_k = \frac{k}{N}$, respectively. So, just by neutrality, there is always a positive fixation probability. We want to quantify the additional advantage given by selection and so define

$$\psi(x) := h(x) - x,$$

and

$$\psi_k := h_k - \frac{k}{N},$$

respectively.

These quantities can also be characterized by a recursion, which one may derive by a first-step-analysis. Of course, one may derive these formulas directly from equation (3.13) [33].

Proposition 3.14. *The quantities ψ_k, $k = 1, \dots, N-1$ follow the recursion*

$$\begin{aligned} NR\psi_k =& \Big((1+s)k(N-k) + \frac{k}{k+1}(N-k)Nu\nu_0\Big)\psi_{k+1} \\ &+ \Big(k(N-k) + \frac{N-k}{N+1-k}iNu\nu_1\Big)\psi_{k-1} + k(N-k)\frac{s}{N}, \end{aligned} \tag{3.14}$$

with $\psi_0 = \psi_N = 0$. □

With these results, we will directly derive a boundary value problem which describes the conditional probability h as it was done in [33]. However, the reasoning in [33] works under the assumption that the quantities h_k converge towards h in an adequate sense. In the following two sections we will prove the validity of this assumption. This proof is rather tedious. For this we first need an explicit solution of recursion 3.14, which we derive iin the following section.

3.3.2 Solution of the recursion

We can give an explicit solution for the recursion (3.14). The following Lemma turns out to be helpful

Lemma 3.15. *Let ψ_j, $j = 0, \dots, N$ be the quantities defined in* (3.14)*. Then the following relation holds for $k = 1, \dots, N$:*

$$(1 + \frac{Nu\nu_1}{k})\psi_{N-k} = (1 + Nu\nu_1)\psi_{N-1} + (1 + s + \frac{Nu\nu_0}{N-k+1})\psi_{N-k+1} - (k-1)\frac{s}{N}. \tag{3.15}$$

Proof. Recursion (3.14) is equivalent to:

$$\begin{aligned} \Big((2+s) + \frac{N}{k}u\nu_0 + \frac{N}{N-k}u\nu_1\Big)\psi_k =& \Big((1+s) + \frac{N}{k+1}u\nu_0\Big)\psi_{k+1} \\ &+ \Big(1 + \frac{N}{N+1-k}u\nu_1\Big)\psi_{k-1} + \frac{s}{N}. \end{aligned} \tag{3.16}$$

Summation over the first $k-1$ expressions of (3.16) gives

$$\sum_{j=1}^{k-1}(2+s+\frac{Nu\nu_0}{N-j}+\frac{Nu\nu_1}{j})\psi_{N-j} = \sum_{j=1}^{k-1}(1+s+\frac{Nu\nu_0}{N-j+1})\psi_{N-j+1} + \sum_{j=1}^{k-1}(1+\frac{Nu\nu_1}{j+1})\psi_{N-j-1}+\frac{(k-1)s}{N},$$

which is equivalent to the assertion of the Lemma. □

The advantage that equation (3.15) has over (3.31) is that it gives a formula in which ψ_{N-k} only depends on ψ_{N-k+1}, the preceding value. Evaluating this formula for some little k suggests the solution of recursion (3.14).

Theorem 3.16. *The solution of recursion* (3.14) *is given by*

$$\psi_{N-k} = \sum_{\ell=1}^{k} A_\ell^k B_\ell^{k-1}\big[(1+Nu\nu_1)\psi_{N-1} - \frac{(\ell-1)s}{N}\big], \tag{3.17}$$

for $k = 0,\dots,N$ *with*

$$A_\ell^{\ell+m} := \prod_{j=0}^{m}\frac{\ell+j}{\ell+j+Nu\nu_1}$$

and

$$B_\ell^{\ell+m} := \prod_{j=0}^{m}(1+s+\frac{Nu\nu_0}{N-\ell-j}).$$

Proof. Obviously, $\psi_N = 0$ holds. For $k \geq 1$, we must check that ψ_{N-k} as it is

given by (3.17) complies with (3.15):

$$\begin{aligned}
\psi_{N-k} =& \frac{k}{k+Nu\nu_1}\Big[(1+Nu\nu_1)\psi_{N-1} + (1+s+\frac{Nu\nu_0}{N-k+1})\psi_{n-k+1} - (k-1)\frac{s}{N}\Big] \\
=& \frac{k}{k+Nu\nu_1}\Big[(1+Nu\nu_1)\psi_{N-1} - (k-1)\frac{s}{N} \\
&+ (1+s+\frac{Nu\nu_0}{N-k+1})\sum_{\ell=1}^{k-1} A_\ell^{k-1}B_\ell^{k-2}[(1+Nu\nu_1)\psi_{N-1} + (\ell-1)\frac{s}{N}]\Big] \\
=& \frac{k}{k+Nu\nu_1}\Big[(1+Nu\nu_1)\psi_{N-1} - (k-1)\frac{s}{N} \\
&+ \sum_{\ell=1}^{k-1} A_\ell^{k-1}B_\ell^{k-1}[(1+Nu\nu_1)\psi_{N-1} + (\ell-1)\frac{s}{N}]\Big] \\
=& A_k^k\Big((1+Nu\nu_1)\psi_{N-1} - (k-1)\frac{s}{N}\Big) \\
&+ \sum_{\ell=1}^{k-1} A_\ell^k B_\ell^{k-1}[(1+Nu\nu_1)\psi_{N-1} + (\ell-1)\frac{s}{N}] \\
=& \sum_{\ell=1}^{k} A_\ell^k B_\ell^{k-1}[(1+Nu\nu_1)\psi_{N-1} + (\ell-1)\frac{s}{N}].
\end{aligned}$$

□

Obviously, we may dispose of ψ_{N-1} in (3.17) by using the boundary condition $\psi_0 = 0$. This leads to

$$\psi_{N-1} = \frac{s}{1+Nu\nu_1}\frac{\sum_{\ell=0}^{N-1}\frac{N-1-\ell}{N}A_{N-\ell}^N B_{N-\ell}^{N-1}}{\sum_{\ell=0}^{N-1}A_{N-\ell}^N B_{N-\ell}^{N-1}}.$$

So, we have

Corollary 3.17.

$$\psi_{N-k} = \sum_{\ell=1}^{k} A_\ell^k B_\ell^{k-1} s\Big[\frac{\sum_{\ell=0}^{N-1}\frac{N-1-\ell}{N}A_{N-\ell}^N B_{N-\ell}^{N-1}}{\sum_{\ell=0}^{N-1}A_{N-\ell}^N B_{N-\ell}^{N-1}} - \frac{(\ell-1)}{N}\Big]. \tag{3.18}$$

□

In the following section we use this solution to prove the convergence of h_k towards h.

3.3.3 Convergence of h_k towards h

In Theorem 3.13 we give a recursion that describes the quantities $(h_k)_{k=0,\dots,N}$ as defined in Section 3.3.1. Taylor gives an explicit solution of the boundary value problem for h [50, Eq (23)]. We also found a slightly different way to this solution in Section 3.3.6. With this solution at hand, we may prove the convergence of the h_k towards h in the following sense: Let $x \in [0,1]$ and let $(k_N)_{N\in\mathbb{N}}$ with $0 \le k_N \le N$ and $k_N/N \to x$. Then, $h_{k_N}/N \to h(x)$ which is simply the convergence of real numbers.

We show the convergence in this sense for the quantities ψ_k, which are defined by $\psi_k := h_k - k/N$, $\psi(x) := h(x) - x$ respectively. By the definition, it is clear that their convergence assures the convergence of the h_k, too.

Lemma 3.18. *The ψ_{N-1}, given by (3.14), have the following asymptotic behaviour*

$$\lim_{N\to\infty} N\psi_{N-1} = \frac{\sigma}{1+\theta\nu_1}(1-\tilde{p}),$$

where

$$\tilde{p} := \frac{\int_0^1 e^{\sigma y}(1-y)^{\theta\nu_1} y^{\theta\nu_0+1} dy}{\int_0^1 e^{\sigma y}(1-y)^{\theta\nu_1} y^{\theta\nu_0} dy}$$

.

Proof. The ψ_k are given by (3.17) which is

$$\psi_{N-k} = \sum_{\ell=1}^{k} A_\ell^k B_\ell^{k-1}\big[(1+Nu\nu_1)\psi_{N-1} - \frac{(\ell-1)s}{N}\big].$$

For $k = N$ we use the boundary condition $\psi_0 = 0$ and obtain

$$\psi_{N-1} = \frac{s}{1+Nu\nu_1} \frac{\sum_{\ell=1}^{N} A_\ell^N B_\ell^{N-1} \frac{\ell-1}{N}}{\sum_{\ell=1}^{N} A_\ell^N B_\ell^{N-1}}. \tag{3.19}$$

We consider the coefficients A_ℓ^N and B_ℓ^{N-1}. With $\ell > 1$, we have for the coeffi-

cients A_ℓ^N:

$$\begin{aligned}
A_\ell^N &= \prod_{j=\ell}^{N} \frac{j}{j+Nu\nu_1} = \prod_{j=\ell}^{N} (1+\frac{Nu\nu_1}{j})^{-1} \\
&= \exp\Bigg[-\sum_{j=\ell}^{N} \log(1+\frac{Nu\nu_1}{j}) \Bigg] \\
&= \exp\Bigg[-Nu\nu_1 \sum_{j=\ell}^{N} \frac{1}{j} - \sum_{j=\ell}^{N} R_1(\frac{Nu\nu_1}{j}) \Bigg] \\
&= \exp\Bigg[-Nu\nu_1 \Big(\sum_{j=1}^{N} \frac{1}{j} - \sum_{j=1}^{\ell-1} \frac{1}{j} \Big) \Bigg] \exp\Bigg[-\sum_{j=\ell}^{N} R_1(\frac{Nu\nu_1}{j}) \Bigg] \\
&= \exp\Bigg[-Nu\nu_1 \Big(\log(N) - \log(\ell-1) \Big) \Bigg] \\
&\quad \times \exp\Bigg[-\sum_{j=\ell}^{N} R_1(\frac{Nu\nu_1}{j}) - Nu\nu_1 R_2(N) + Nu\nu_1 R_2(\ell-1) \Bigg] \\
&= \left(\frac{\ell-1}{N} \right)^{Nu\nu_1} \exp\Bigg[-\sum_{j=\ell}^{N} R_1(\frac{Nu\nu_1}{j}) - Nu\nu_1 R_2(N) + Nu\nu_1 R_2(\ell-1) \Bigg],
\end{aligned} \tag{3.20}$$

where the functions R_1 and R_2 are defined by

$$R_1(x) := \log(1+x) - x$$

and

$$R_2(n) := \sum_{i=1}^{n} \frac{1}{i} - \gamma - \log(n),$$

and γ is the Euler-Mascheroni constant. For R_1, we have the estimate $|R_1(x)| \leq \frac{x^2}{2}$ for $x \geq 0$ and $R_2(n) \in \mathcal{O}(\frac{1}{n})$ for $n \to \infty$. We must keep in mind, that this transformation cannot be done for $\ell = 1$.

Similarly, we obtain for B_ℓ^{N-1} with $\ell < N$:

$$\begin{aligned} B_\ell^{N-1} &= \prod_{j=1}^{N-\ell}(1+s+\frac{Nu\nu_0}{j}) = \exp\left[\sum_{j=1}^{N-\ell}(s+\frac{Nu\nu_0}{j})\right]\exp\left[\sum_{j=1}^{N-\ell}R_1(s+\frac{Nu\nu_0}{j})\right] \\ &= \exp((N-\ell)s)\exp\Big[Nu\nu_0\log(N-\ell)\Big] \\ &\quad\times\exp\left[\sum_{j=1}^{N-\ell}R_1(s+\frac{Nu\nu_0}{j})+Nu\nu_0(\gamma+R_2(N-\ell))\right] \\ &= \exp((N-\ell)s)(N-\ell)^{Nu\nu_0} \\ &\quad\times\exp\left[\sum_{j=1}^{N-\ell}R_1(s+\frac{Nu\nu_0}{j})+Nu\nu_0(\gamma+R_2(N-\ell))\right]. \end{aligned} \tag{3.21}$$

In the case $\ell = N$, we have by definition $B_N^{N-1} = 1$.
Define step functions $T_1^N : [0,1] \to \mathbb{R}$ by

$$T_1^N(y) = \begin{cases} 0, \quad y \le \frac{1}{N}; \\ \exp(sN\frac{N-\ell}{N})\Big(\frac{\ell-1}{N}\Big)^{Nu\nu_1+1}\Big(\frac{N-\ell}{N}\Big)^{Nu\nu_0}\exp\Big[R(N,\ell)\Big], \quad \frac{\ell-1}{N} < y \le \frac{\ell}{N}, \\ \quad \ell = 2,\dots,N-1; \\ \Big(\frac{N-1}{N}\Big)^{Nu\nu_1+1}\Big(\frac{1}{N}\Big)^{Nu\nu_0} \\ \times\exp\Big[-R_1(\frac{Nu\nu_1}{N})-Nu\nu_1R_2(N)+Nu\nu_1R_2(N-1)\Big], \quad y > \frac{N-1}{N}, \end{cases}$$

where

$$\begin{aligned} R(N,\ell) := &-\sum_{j=\ell}^{N}R_1(\frac{Nu\nu_1}{j})-Nu\nu_1R_2(N)+Nu\nu_1R_2(\ell-1) \\ &+\sum_{j=1}^{N-\ell}R_1(s+\frac{Nu\nu_0}{j})+Nu\nu_0(R_2(N-\ell)+\gamma). \end{aligned}$$

Again, we have a look at (3.19):

$$\begin{aligned} N\psi_{N-k} &= \frac{Ns}{1+Nu\nu_1}\frac{\sum_{\ell=1}^{N}A_\ell^N B_\ell^{N-1}\frac{\ell-1}{N}}{\sum_{\ell=1}^{N}A_\ell^N B_\ell^{N-1}} \\ &= \frac{Ns}{1+Nu\nu_1}\frac{\frac{1}{N}\sum_{\ell=1}^{N}A_\ell^N B_\ell^{N-1}N^{-Nu\nu_0}\frac{\ell-1}{N}}{\frac{1}{N}\sum_{\ell=1}^{N}A_\ell^N B_\ell^{N-1}N^{-Nu\nu_0}}. \end{aligned}$$

Then, the numerator is just the integral of T_1^N with respect to the Lebesgue measure. Let $y \in (0,1)$ and define $\ell_N^y \in \mathbb{N}$ by $\ell_N^y - 1/N < y \leq \ell_N^y/N$. Then, $\ell_N^y \to \infty$ and $N - \ell_N^y \to \infty$ as $N \to \infty$. Then, $R(N, \ell_N^y) \to \theta\gamma + C$ as $N \to \infty$, where $C := \lim_{N\to\infty} \sum_{j=1}^{N} R_1(s + \frac{Nu\nu_0}{j})$, which is finite since $|R_1(x)| \leq x^2/2$ and $s \in \mathcal{O}(1/N)$. Thus,

$$T_1^N(y) \to g(y) := \exp(\sigma(1-y))y^{\theta\nu_1+1}(1-y)^{\theta\nu_0}\exp(\gamma + C) \quad \text{as} \quad N \to \infty.$$

Furthermore, there is a $K > 0$ with $T_1^N \leq K$ for all $N \in \mathbb{N}$, since for arbitrary $\ell = 2, \ldots, N-1$:

$$\begin{aligned}
&\exp(sN\frac{N-\ell}{N})\Big(\frac{\ell-1}{N}\Big)^{Nu\nu_1+1}\Big(\frac{N-\ell}{N}\Big)^{Nu\nu_0}\exp\Big[R(N,\ell)\Big] \\
&\quad\leq \exp(sN)\exp\Big[R(N,\ell)\Big] \\
&\quad\leq \exp(sN)\exp\Bigg[\sum_{j=1}^{N}|R_1(\frac{Nu\nu_1}{j})| + Nu\nu_1|R_2(N)| + Nu\nu_1|R_2(\ell-1)| \\
&\qquad\qquad + \sum_{j=1}^{N}|R_1(s+\frac{Nu\nu_0}{j})| + |Nu\nu_0 R_2(N-\ell)| + Nu\nu_0\gamma\Bigg]
\end{aligned}$$

Since the series $\sum_{j=1}^{N}|R_1(\frac{Nu\nu_1}{j})|$ and $\sum_{j=1}^{N}|R_1(s+\frac{Nu\nu_0}{j})|$ converge, the sums are less than some $K_1 > 0$. Furthermore, the maximum over all $|R_2(n)|$, which we denote by K_2, exists because R_2 vanishes for large arguments. Trivially, there are upper bounds K_3, K_4 and K_5, for Ns, $Nu\nu_0$ and $Nu\nu_1$, such that we finally have

$$|T_1^N(y)| \leq \exp(K_3)\exp(2K_1 + K_4\gamma + 2K_5K_2 + K_4K_2) =: K$$

for $\frac{1}{N} < y \leq \frac{N-1}{N}$. For $y \leq \frac{1}{N}$, T_1^N is zero, and for $y > \frac{N-1}{N}$, it is obviously bounded, too. So, we have

$$\lim_{N\to\infty}\int_0^1 T_1^N(y)dy = \int_0^1 \big(\lim_{N\to\infty} T_1^N(y)\big)dy.$$

Of course, we can do the same procedure for the denominator of (3.19) since it deviates from the numerator just by the factor $(\ell-1)/N$, that means, we may define step functions T_2^N which are bounded and converge almost everywhere to

$\exp(\sigma(1-y))y^{\theta\nu_1}(1-y)^{\theta\nu_0}\exp(\theta\gamma+C)$. Thus, we have

$$\begin{aligned}
\lim_{N\to\infty} N\psi_{N-1} &= \frac{\sigma}{1+\theta\nu_1}\frac{\int_0^1 \exp(\sigma(1-y))y^{\theta\nu_1+1}(1-y)^{\theta\nu_0}\exp(\theta\gamma+C)dy}{\int_0^1 \exp(\sigma(1-y))y^{\theta\nu_1}(1-y)^{\theta\nu_0}\exp(\theta\gamma+C)dy} \\
&= \frac{\sigma}{1+\theta\nu_1}\frac{\int_0^1 \exp(\sigma(1-y))y^{\theta\nu_1+1}(1-y)^{\theta\nu_0}dy}{\int_0^1 \exp(\sigma(1-y))y^{\theta\nu_1}(1-y)^{\theta\nu_0}dy} \\
&= \frac{\sigma}{1+\theta\nu_1}\frac{\int_0^1 e^{\sigma y}(1-y)^{\theta\nu_1+1}y^{\theta\nu_0}dy}{\int_0^1 e^{\sigma y}(1-y)^{\theta\nu_1}y^{\theta\nu_0}dy} \\
&= \frac{\sigma}{1+\theta\nu_1}\left(1-\frac{\int_0^1 e^{\sigma y}(1-y)^{\theta\nu_1}y^{\theta\nu_0+1}dy}{\int_0^1 e^{\sigma y}(1-y)^{\theta\nu_1}y^{\theta\nu_0}dy}\right) \\
&= \frac{\sigma}{1+\theta\nu_1}(1-\tilde{p}),
\end{aligned}$$

where we used in the fourth step that

$$\int_0^1 e^{\sigma y}(1-y)^{\theta\nu_1+1}y^{\theta\nu_0}dy + \int_0^1 e^{\sigma y}(1-y)^{\theta\nu_1}y^{\theta\nu_0+1}dy = \int_0^1 e^{\sigma y}(1-y)^{\theta\nu_1}y^{\theta\nu_0}dy.$$

□

Theorem 3.19. *Let $x \in [0,1]$, $(k_N)_{N\in\mathbb{N}}$ a sequence of natural numbers with $k_N \leq N$ and $\lim_{N\to\infty}\frac{k_N}{N} = x$. Then, the sequence $(\psi_{k_N})_{N\in\mathbb{N}}$ converges to $\psi(x)$ as it is given in* (3.36)*, the unique solution of the boundary value problem* (3.26)*.*

Proof. From Theorem 3.16 we have

$$\psi_k = \frac{1}{N}\sum_{\ell=1}^{N-k} A_\ell^{N-k} B_\ell^{N-k-1}[(1+Nu\nu_1)N\psi_{N-1} - Ns\frac{\ell-1}{N}] \tag{3.22}$$

Similar to our approach in the previous Lemma, we first consider A_ℓ^{N-k} and B_ℓ^{N-k-1} and use the same notation for the error functions R_1 and R_2 that emerge in the approximation of the harmonic series and of the series expansion of the logarithm.

Let $\ell > 1, k > 0$, then

$$\begin{aligned}
A_\ell^{N-k} &= \exp\Big[-\sum_{j=\ell}^{N-k} \frac{Nu\nu_1}{j} - \sum_{j=\ell}^{N-k} R_1(\frac{Nu\nu_1}{j})\Big] \\
&= \exp\Big[-Nu\nu_1\Big(\sum_{j=1}^{N-k}\frac{1}{j} - \sum_{j=1}^{\ell-1}\frac{1}{j}\Big) - \sum_{j=\ell}^{N-k} R_1(\frac{Nu\nu_1}{j})\Big] \\
&= \exp\Big[-Nu\nu_1(\log(N-k) - \log(\ell-1))\Big] \\
&\exp\Big[-\sum_{j=\ell}^{N-k} R_1(\frac{Nu\nu_1}{j}) - Nu\nu_1(R_2(N-k) - R_2(\ell-1))\Big] \\
&= \Big(\frac{\ell-1}{N}\Big)^{Nu\nu_1}\Big(\frac{N-k}{N}\Big)^{-Nu\nu_1} \\
&\exp\Big[-\sum_{j=\ell}^{N-k} R_1(\frac{Nu\nu_1}{j}) - Nu\nu_1(R_2(N-k) - R_2(\ell-1))\Big]
\end{aligned}$$

For B_ℓ^{N-k-1} with $k > 0$ and $\ell < N - k$, we have

$$\begin{aligned}
B_\ell^{N-k-1} &= \prod_{j=k+1}^{N-\ell}(1 + s + \frac{Nu\nu_0}{j}) \\
&= \exp\Big[\sum_{j=k+1}^{N-\ell}(s + \frac{Nu\nu_0}{j}) + \sum_{k+1}^{N-\ell} R_1(s + \frac{Nu\nu_0}{j})\Big] \\
&= \exp(s(N-\ell-k))\exp\Big[Nu\nu_0(\log(N-\ell) - \log(k))\Big] \\
&\quad \exp\Big[\sum_{j=k+1}^{N-\ell} R_1(s + \frac{Nu\nu_0}{j}) + Nu\nu_0(R_2(N-\ell) - R_2(k))\Big] \\
&= \exp(s(N-k-\ell))\Big(\frac{N-\ell}{N}\Big)^{Nu\nu_0}\Big(\frac{k}{N}\Big)^{-Nu\nu_0} \\
&\quad \exp\Big[\sum_{j=k+1}^{N-\ell} R_1(s + \frac{Nu\nu_0}{j}) + Nu\nu_0(R_2(N-\ell) - R_2(k))\Big].
\end{aligned}$$

For the error, we define a function R by

$$\begin{aligned}
R(N,k,\ell) := &-\sum_{j=\ell}^{N-k} R_1(\frac{Nu\nu_1}{j}) - Nu\nu_1(R_2(N-k) - R_2(\ell-1)) \\
&+ \sum_{j=k+1}^{N-\ell} R_1(s + \frac{Nu\nu_0}{j}) + Nu\nu_0(R_2(N-\ell) - R_2(k)).
\end{aligned}$$

Now, let $x \in (0,1)$ and $(k_N)_{N\in\mathbb{N}}$ a sequence with $0 \le k_N \le N$ and $k_N/N \to x$ as $N \to \infty$. Since we are interested in the limit for $N \to \infty$, we may assume, that $1 \le k_N \le N-1$. Define a step function $T_{k_N} : [0,1] \to \mathbb{R}$ by

$$T_{k_N}(y) = \begin{cases} 0, \quad \text{if } \ y = 0; \\ A_1^{N-k_N} B_1^{N-k_N-1}[(1+Nu\nu_1)N\psi_{N-1}], \quad \text{if } \ 0 < y \le \frac{1}{N}; \\ e^{-sk_N}\Big(\frac{k_N}{N}\Big)^{-Nu\nu_0}\Big(\frac{N-k_N}{N}\Big)^{-Nu\nu_1} e^{s(N-\ell)}\Big(\frac{N-\ell}{N}\Big)^{Nu\nu_0}\Big(\frac{\ell-1}{N}\Big)^{Nu\nu_1} \\ \quad \times[(1+Nu\nu_1)N\psi_{N-1} - Ns\frac{\ell-1}{N}]\exp[R(N,k_N,\ell)], \\ \quad \text{if } \ \frac{\ell-1}{N} < y \le \frac{\ell}{N}, \quad \ell = 2,\dots,N-k_N-1; \\ A_{N-k_N}^{N-k_N}[(1+Nu\nu_1)N\psi_{N-1} - Ns\frac{N-k_N-1}{N}], \quad \text{if } \ \frac{N-k_N-1}{N} < y \le \frac{N-k_N}{N}; \\ 0, \quad \text{if } \ y > \frac{N-k_N}{N}. \end{cases}$$

Then the right-hand side of (3.22) is just the integral of T_{k_N}. Let $y \in (0,1)$ and ℓ_N^y such that $(\ell_N^y - 1)/N < y \le \ell_N^y/N$. Then, ℓ_N^y, $N - \ell_N^y$, k_N, $N - k_N$ tend to infinity as $N \to \infty$ and thus $R(N, k_N, \ell_N^y)$ vanishes for large N. For $y \in \{0,1\}$, we have $T_{k_N}(y) = 0$ for large N. So,

$$\begin{aligned} g(y) := \lim_{N\to\infty} T_{k_N}(y) = &\mathbb{1}_{[0,1-x]}(y) e^{-\sigma x} x^{-\theta\nu_0}(1-x)^{-\theta\nu_1} \\ &e^{\sigma(1-y)}(1-y)^{\theta\nu_0} y^{\theta\nu_1}[(1+\theta\nu_1)\frac{\sigma}{1+\theta\nu_1}(1-\tilde{p}) - \sigma y] \end{aligned}$$

almost everywhere. In analogy to the proof of the previous Lemma, we can see that the functions T_{k_N} are bounded[3], and that we may also apply dominated convergence to obtain that

$$\lim_{N\to\infty} \psi_{k_N} = \lim_{N\to\infty} \int_0^1 T_{k_N}(y)dy = \int_0^1 g(y)dy.$$

[3]Maybe, it is not quite obvious, that the terms for $y \le \frac{1}{N}$ are bounded, too. Trivially, $A_1^{N-k_N} \le 1$. Furthermore,

$$B_1^{N-k_N-1} = e^{s(N-k_N-1)}\Big(\frac{N-1}{k_N}\Big)^{Nu\nu_0} \exp\Big[\sum_{j=k_N+1}^{N-1} R_1(s+\frac{Nu\nu_0}{j}) + Nu\nu_0(R_2(N-1) - R_2(k_N))\Big],$$

which is also bounded since $(N-1)/k_N \to \frac{1}{x}$.

Via a substitution, we finally see that this limit is indeed $\psi(x)$ as given in (3.36):

$$\begin{aligned}\int_0^{1-x} e^{\sigma(1-y)}(1-y)^{\theta\nu_0}y^{\theta\nu_1}\sigma(1-\tilde{p}-y)dy &= \int_x^1 e^{\sigma y}y^{\theta\nu_0}(1-y)^{\theta\nu_1}\sigma(y-\tilde{p})dy\\ &=\sigma\int_0^1 e^{\sigma y}y^{\theta\nu_0}(1-y)^{\theta\nu_1}(y-\tilde{p})dy\\ &\quad+\sigma\int_0^x e^{\sigma y}y^{\theta\nu_0}(1-y)^{\theta\nu_1}(\tilde{p}-y)dy\\ &=\sigma\int_0^x e^{\sigma y}y^{\theta\nu_0}(1-y)^{\theta\nu_1}(\tilde{p}-y)dy,\end{aligned}$$

where we used that

$$\sigma\int_0^1 e^{\sigma y}y^{\theta\nu_0}(1-y)^{\theta\nu_1}(y-\tilde{p})dy = 0$$

due to the definition of $\tilde{p}$. The proof is not complete yet. We must check the convergence at the boundaries, that means

$$\lim_{N\to\infty}\psi_{k_N} = 0$$

for arbitrary sequences k_N with

$$\lim_{N\to\infty}\frac{k_N}{N} = 0 \quad \text{or} \quad \lim_{N\to\infty}\frac{k_N}{N} = 1.$$

We first consider the case $k_N/N \to 1$. Since $\psi_N = 0$, we may assume that $k_N < N$. From the previous reasoning we have

$$\begin{aligned}\psi_{k_N} =& \exp(-Ns\frac{k_N}{N})\Big(\frac{k_N}{N}\Big)^{-Nu\nu_0}\Big(\frac{N-k_N}{N}\Big)^{-Nu\nu_1}\\ &\times\frac{1}{N}\sum_{\ell=2}^{N-k_N-1}\Big[e^{s(N-\ell)}\Big(\frac{N-\ell}{N}\Big)^{Nu\nu_0}\Big(\frac{\ell-1}{N}\Big)^{Nu\nu_1}[(1+Nu\nu_1)N\psi_{N-1}-Ns\frac{\ell-1}{N}]\\ &\qquad\times\exp(R(N,k_N,\ell))\Big]\\ &+\frac{1}{N}A_1^{N-k_N}B_1^{N-k_N-1}[(1+Nu\nu_1)N\psi_{N-1}]\\ &+\frac{1}{N}A_{N-k_N}^{N-k_N}[(1+Nu\nu_1)N\psi_{N-1}-Ns\frac{N-k_N-1}{N}].\end{aligned}$$

We obviously may find upper bounds for $\exp(-Ns\frac{k_N}{N})$, $\Big(\frac{k_N}{N}\Big)^{-Nu\nu_0}$, $\exp(s(N-\ell))$, $\Big(\frac{N-\ell}{N}\Big)^{Nu\nu_0}$, $[(1+Nu\nu_1)N\psi_{N-1}-Ns\frac{\ell-1}{N}]$, $R(N,k_N,\ell)$,$A_1^{N-k_N}B_1^{N-k_N-1}$ and $A_{N-k_N}^{N-k_N}$, such that we obtain

$$
\begin{aligned}
|\psi_{k_N}| \leq & C_1 \Big(\frac{N-k_N}{N}\Big)^{-Nu\nu_1} \frac{1}{N} \sum_{\ell=2}^{N-k_N-1} \Big(\frac{\ell-1}{N}\Big)^{Nu\nu_1} + \frac{C_2}{N} \\
\leq & C_1 \frac{1}{N} \sum_{\ell=2}^{N-k_N-1} \Big(\frac{N-k_N-1}{N-k_N}\Big)^{Nu\nu_1} + \frac{C_2}{N} \\
\leq & C_1 \frac{1}{N}(N-k_N-2) + \frac{C_2}{N} \to 0 \quad \text{as} \quad N \to \infty.
\end{aligned}
$$

It is not obvious how to obtain the same result for $k_N/N \to 0$ from this solution of the recursion. It becomes apparent that it is very simple again, when we consider the solution of the recursion starting from the bottom. See the following Lemma. We will also use (3.23) and (3.24). Here, we assume $k_N > 0$. With this, we have

$$
\begin{aligned}
\psi_{k_N} = & e^{-k_N s} (\frac{k_N}{N})^{-Nu\nu_0} (\frac{N-k_N}{N})^{-Nu\nu_1} \\
& \times \frac{1}{N} \sum_{\ell=2}^{k_N-1} \Big[e^{\ell s} (\frac{\ell-1}{N})^{Nu\nu_0} (\frac{N-\ell}{N})^{Nu\nu_1} [(1+s+Nu\nu_0)N\psi_1 - \frac{\ell-1}{N} Ns] \\
& \qquad \times \exp[R(N,k_N,\ell)] \Big] \\
& + \frac{1}{N} \tilde{A}_1^{k_N} \tilde{B}_1^{k_N} (1+s+Nu\nu_0) N\psi_1 \\
& + \frac{1}{N} \tilde{A}_{k_N}^{k_N} [(1+s+Nu\nu_0)N\psi_1 - (k_N-1)s],
\end{aligned}
$$

where

$$
\begin{aligned}
R(N,k_N,\ell) := & - Nu\nu_0 R_2(k_N) + Nu\nu_0 R_2(\ell-1) - \sum_{j=\ell}^{k_N} R_1(s + \frac{Nu\nu_0}{j}) \\
& + Nu\nu_1 (R_2(N-\ell) - R_2(N-k_N)) + \sum_{j=N-k_N+1}^{N-\ell} R_1(\frac{Nu\nu_1}{j}).
\end{aligned}
$$

As in the previous case with limit 1, we may find upper bounds for all factors apart from $(k_N/N)^{-Nu\nu_0}$ and finally obtain

$$
|\psi_{k_N}| \leq C_3 \frac{1}{N} \sum_{\ell=2}^{k_N-1} (\frac{\ell-1}{k_N})^{Nu\nu_0} + \frac{C_4}{N} \leq C_3 \frac{1}{N}(k_N-2) + \frac{C_4}{N} \to 0 \quad \text{as} \quad N \to \infty.
$$

□

Lemma 3.20. *A solution of recursion* (3.14) *is given by*

$$\psi_k = \sum_{\ell=1}^{k} \tilde{A}_\ell^k \tilde{B}_\ell^{k-1}[(1+s+Nu\nu_0)\psi_1 - \frac{(\ell-1)s}{N}],$$

where

$$\tilde{A}_\ell^k := \prod_{j=\ell}^{k} \frac{j}{j(1+s)+Nu\nu_0}$$

and

$$\tilde{B}_\ell^{k-1} := \prod_{j=N-k+1}^{N-\ell} (1+\frac{Nu\nu_1}{j}).$$

Furthermore,

$$\lim_{N\to\infty} N\psi_1 = \frac{\sigma}{1+\theta\nu_0}\tilde{p}.$$

Proof. For the first assertion, we sum the first $k-1$ expressions of (3.14):

$$\begin{aligned}\sum_{j=1}^{k-1}(2+s+\frac{Nu\nu_0}{j}+\frac{Nu\nu_1}{N-j})\psi_j =& \sum_{j=1}^{k-1}(1+s+\frac{Nu\nu_0}{j+1})\psi_{j+1} \\ &+\sum_{j=1}^{k-1}(1+\frac{Nu\nu_1}{N+1-j})\psi_{j-1}+(k-1)\frac{s}{N},\end{aligned}$$

which leads to

$$(1+s+\frac{Nu\nu_0}{k})\psi_k = (1+s+Nu\nu_0)\psi_1 + (1+\frac{Nu\nu_1}{N-k+1})\psi_{k-1} - (k-1)\frac{s}{N}.$$

We check the assertion by induction and use the equation above:

$$\begin{aligned}\psi_k =& \frac{k}{k(1+s)+Nu\nu_0}[(1+s+Nu\nu_0)\psi_1 - (k-1)\frac{s}{N} \\ &+\frac{N-k+1+Nu\nu_1}{N-k+1}\sum_{\ell=1}^{k-1}\tilde{A}_\ell^{k-1}\tilde{B}_\ell^{k-2}[(1+s+Nu\nu_0)\psi_1-\frac{(\ell-1)s}{N}]] \\ =& \tilde{A}_k^k\Bigg[(1+s+Nu\nu_0)\psi_1-(k-1)\frac{s}{N} \\ &+\sum_{\ell=1}^{k-1}\tilde{A}_\ell^{k-1}\tilde{B}_\ell^{k-1}[(1+s+Nu\nu_0)\psi_1-\frac{(\ell-1)s}{N}]\Bigg] \\ =& \sum_{\ell=1}^{k}\tilde{A}_\ell^k\tilde{B}_\ell^{k-1}[(1+s+Nu\nu_0)\psi_1-\frac{(\ell-1)s}{N}].\end{aligned}$$

So, we proceed with the second assertion. The first assertion gives for $k = N$, where we use $\psi_N = 0$:

$$N\psi_1 = \frac{Ns}{1+s+Nu\nu_0} \frac{\frac{1}{N}\sum_{\ell=1}^{N} \frac{\ell-1}{N} \tilde{A}_\ell^N \tilde{B}_\ell^{N-1}}{\frac{1}{N}\sum_{\ell=1}^{N} \tilde{A}_\ell^N \tilde{B}_\ell^{N-1}}.$$

As we did in Lemma 3.18, we first consider the coefficients $\tilde{A}_\ell^k$ and $\tilde{B}_\ell^{k-1}$. For $\ell > 1$,

$$\begin{aligned}
\tilde{A}_\ell^k =& \prod_{j=\ell}^{k} \Big(\frac{j(1+s)+Nu\nu_0}{j} \Big)^{-1} \\
=& \exp\Big[-\sum_{j=\ell}^{k} \log(1+s+\frac{Nu\nu_0}{j}) \Big] \\
=& \exp\Big[-\sum_{j=\ell}^{k} (s+\frac{Nu\nu_0}{j}) - \sum_{j=\ell}^{k} R_1(s+\frac{Nu\nu_0}{j}) \Big] \\
=& \exp\Big[-(k-\ell)s - Nu\nu_0 \sum_{j=\ell}^{k} \frac{1}{j} - \sum_{j=\ell}^{k} R_1(s+\frac{Nu\nu_0}{j}) \Big] \\
=& e^{(\ell-k)s} \exp\Big[-Nu\nu_0 \log(k) + Nu\nu_0 \log(\ell-1) - Nu\nu_0 R_2(k) + Nu\nu_0 R_2(\ell-1) \\
& \qquad - \sum_{j=\ell}^{k} R_1(s+\frac{Nu\nu_0}{j}) \Big] \\
=& e^{(\ell-k)s} \Big(\frac{\ell-1}{k} \Big)^{Nu\nu_0} \exp\Big[-Nu\nu_0 R_2(k) + Nu\nu_0 R_2(\ell-1) - \sum_{j=\ell}^{k} R_1(s+\frac{Nu\nu_0}{j}) \Big].
\end{aligned} \tag{3.23}$$

For $\tilde{B}_\ell^{k-1}$ with $k < N$ and $\ell < k$, we have

$$\begin{aligned}\tilde{B}_\ell^{k-1} &= \prod_{j=N-k+1}^{N-\ell}(1+\frac{Nu\nu_1}{j}) = \exp\Big[\sum_{j=N-k+1}^{N-\ell}\log(1+\frac{Nu\nu_1}{j})\Big] \\ &= \exp\Big[Nu\nu_1\sum_{j=N-k+1}^{N-\ell}\frac{1}{j}+\sum_{j=N-k+1}^{N-\ell}R_1(\frac{Nu\nu_1}{j})\Big] \\ &= \exp\Big[Nu\nu_1(\log(N-\ell)-\log(N-k))+Nu\nu_1(R_2(N-\ell)-R_2(N-k)) \\ &\qquad\qquad +\sum_{j=N-k+1}^{N-\ell}R_1(\frac{Nu\nu_1}{j})\Big] \\ &= \Big(\frac{N-\ell}{N-k}\Big)^{Nu\nu_1}\exp\Big[Nu\nu_1(R_2(N-\ell)-R_2(N-k))+\sum_{j=N-k+1}^{N-\ell}R_1(\frac{Nu\nu_1}{j})\Big]\end{aligned} \tag{3.24}$$

for $k < N$, and for $k = N$:

$$\tilde{B}_\ell^{N-1} = (N-\ell)^{Nu\nu_1}\exp\Big[Nu\nu_1(R_2(N-\ell)+\gamma)+\sum_{j=1}^{N-\ell}R_1(\frac{Nu\nu_1}{j})\Big].$$

By following a completely analogue reasoning as in Lemma 3.18, we obtain in the limit

$$\lim_{N\to\infty}N\psi_1 = \frac{\sigma}{1+\theta\nu_0}\frac{\int_0^1 e^{\sigma y}y^{\theta\nu_0+1}(1-y)^{\theta\nu_1}dy}{\int_0^1 e^{\sigma y}y^{\theta\nu_0}(1-y)^{\theta\nu_1}dy} = \frac{\sigma}{1+\theta\nu_0}\tilde{p},$$

which is the second assertion.

□

3.3.4 Derivation of the boundary value problem

As mentioned at the end of Section 3.3.1, we may derive the boundary value problem directly from the recursion in Theorem 3.13.

Theorem 3.21. *The conditional probability h fulfills the following boundary value problem*

$$\begin{aligned} 0 =& x(1-x)h''(x) + (\sigma x(1-x) + \theta\nu_0(1-x) - \theta\nu_1 x)h'(x) \\ & - \theta\Big(\nu_0\frac{1-x}{x} + \nu_1\frac{x}{1-x}\Big)h(x) + \theta\nu_1\frac{x}{1-x}, \qquad h(0)=0, \quad h(1)=1. \end{aligned} \tag{3.25}$$

Proof. From Theorem 3.19 we know that the probability $(h_k)_{0\le k\le N}$ converges to the conditional probability h. Equation (3.13) is equivalent to

$$\begin{aligned} 0 =& \frac{k(N-k)}{N^2}N^2\Big(h_{k+1} - 2h_k + h_{k-1}\Big) + Ns\frac{k(N-k)}{N}N\Big(h_{k+1} - h_k\Big) \\ & + Nu\nu_0\frac{N-k}{N}N\Big(h_{k+1} - h_k\Big) + Nu\nu_1\frac{k}{N}N\Big(h_{k-1} - h_k\Big) \\ & - Nu\nu_0\frac{N-k}{N}\frac{Nh_{k+1}}{k+1} + Nu\nu_1\frac{k}{N}\frac{N(1-h_{k-1})}{N-k+1} \end{aligned}$$

and yields for N tends to infinity, $\frac{k}{N}\to x$, $x\in(0,1)$ and due to the fact that h is holomorphic in $(0,1)$ (see Lemma 3.8) the assertion of the theorem. □

So, we can derive the same boundary value problem for the conditional probability h as in [50] directly from the Moran model. (3.25) directly gives the boundary value problem for ψ [33]:

Proposition 3.22. *The function ψ follows the following boundary value problem on $[0,1]$:*

$$\begin{aligned} 0 =& x(1-x)\psi''(x) + \sigma x(1-x)\psi'(x) + \theta\nu_0(1-x)\Big(\psi'(x) - \frac{\psi(x)}{x}\Big) \\ & - \theta\nu_1 x\Big(\psi'(x) + \frac{\psi(x)}{1-x}\Big) + \sigma x(1-x), \quad \psi(0)=\psi(1)=0. \end{aligned} \tag{3.26}$$

□

3.3.5 Solution of the boundary value problem

The aim of this section is to give an explicit solution of the boundary value problem (3.10). As in [50], we consider the series expansion

$$h(x) = x + x\sum_{j=1}^{\infty} a_j(1-x)^j, \tag{3.27}$$

and define for further analysis the function v

$$v(x) := \frac{h(x)-x}{x} = \frac{\psi(x)}{x} = \sum_{j\geq 1} a_j (1-x)^j, \quad x \in (0,1].$$

In terms of v, the coefficients a_j can be written as

$$a_j = (-1)^j \frac{v^{(j)}(1)}{j!}, \quad j \in \mathbb{N}. \tag{3.28}$$

From here on Taylor proceeds with a verification of Fearnheads solution using the λ_n-coefficients from (3.5) and (3.6).

We proceed in another way and derive the recursion for the coefficients a_j directly from the first step analysis result for h_k, (3.13). [4] To this end, we first analyse the connection between the functions v and ψ.

Lemma 3.23. *The $n-th$ derivative of v is*

$$v^{(n)}(x) = \sum_{i=0}^{n} (-1)^i \frac{n!}{(n-i)!} \frac{\psi^{(n-i)}(x)}{x^{i+1}}.$$

Proof. The proposition follows from induction:

$$\begin{aligned}
v^{(n+1)}(x) =& \frac{d}{dx} v^{(n)}(x) \\
=& \frac{d}{dx} \sum_{i=0}^{n} (-1)^i \frac{n!}{(n-i)!} \frac{\psi^{(n-i)}(x)}{x^{i+1}} \\
=& \sum_{i=0}^{n} (-1)^i \frac{n!}{(n-i)!} \Big[\frac{\psi^{(n+1-i)}(x)}{x^{i+1}} - (i+1) \frac{\psi^{(n-i)}(x)}{x^{i+2}} \Big] \\
=& \sum_{i=0}^{n} (-1)^i \frac{n!}{(n-i)!} \frac{\psi^{(n+1-i)}(x)}{x^{i+1}} + \sum_{i=0}^{n} (-1)^{i+1} (i+1) \frac{n!}{(n-i)!} \frac{\psi^{(n-i)}(x)}{x^{i+2}} \\
=& \sum_{i=0}^{n} (-1)^i \frac{n!}{(n-i)!} \frac{\psi^{(n+1-i)}(x)}{x^{i+1}} + \sum_{i=1}^{n+1} (-1)^i i \frac{n!}{(n+1-i)!} \frac{\psi^{(n+1-i)}(x)}{x^{i+1}} \\
=& \frac{\psi^{(n+1)}(x)}{x} + \sum_{i=1}^{n} (-1)^i \frac{n!}{(n+1-i)!} (n+1-i+i) \frac{\psi^{(n+1-i)}(x)}{x^{i+1}} \\
& + (-1)^{n+1} \frac{\psi(x)}{x^{n+2}} \\
=& \sum_{i=0}^{n+1} (-1)^i \frac{(n+1)!}{(n+1-i)!} \frac{\psi^{(n+1-i)}(x)}{x^{i+1}}.
\end{aligned}$$

[4] This approach was already described in [33], but not conducted till its end. She stops with the derivation of the second derivative of ψ in 1.

□

From Lemma 3.23 follows a recursion for a_j.

Proposition 3.24. *For a_j, $j \in \mathbb{N}$, from (3.27), the following recursion holds*

$$a_j = (-1)^j \frac{\psi^{(j)}(1)}{j!} + a_{j-1}$$

with $a_1 = -v'(1) = \psi'(1)$.

Proof. This follows directly from Lemma 3.23:

$$\begin{aligned}
a_j =& (-1)^j \frac{v^{(j)}(1)}{j!} \\
=& (-1)^j \frac{\psi^{(j)}(1)}{j!} + \frac{(-1)^j}{j!} \sum_{i=1}^{j} (-1)^i \frac{j!}{(j-i)!} \psi^{(j-i)}(1) \\
=& (-1)^j \frac{\psi^{(j)}(1)}{j!} + \frac{(-1)^j}{j!} \sum_{i=0}^{j-1} (-1)^{i+1} \frac{j!}{(j-1-i)!} \psi^{(j-1-i)}(1) \\
=& (-1)^j \frac{\psi^{(j)}(1)}{j!} + \frac{(-1)^{j-1}}{(j-1)!} \sum_{i=0}^{j-1} (-1)^i \frac{(j-1)!}{(j-1-i)!} \psi^{(j-1-i)}(1) \\
=& (-1)^j \frac{\psi^{(j)}(1)}{j!} + a_{j-1}.
\end{aligned}$$

□

As a trivial consequence of Proposition 3.24, we can give the a_j in terms of derivatives of ψ:

$$a_j = \sum_{i=0}^{j} (-1)^i \frac{\psi^{(i)}(1)}{i!}.$$

Thus, we reduce the calculation of the a_j to the calculation of the derivatives of ψ in 1. For this purpose, we will use recursion (3.14) for ψ_k.

Derivatives of ψ

We are interested in $\frac{d^n}{dp^n}\psi(1) = \psi^{(n)}(1)$. We may derive it from the quantities ψ_k with the help of the formula

$$\psi^{(n)}(1) = \lim_{N \to \infty} N^n \sum_{j=0}^{n} (-1)^k \binom{n}{j} \psi_{N-j}. \tag{3.29}$$

Remark 3.25. Due to the fact that recursion (3.14) is of degree two, this approach cannot directly help to calculate $\psi'(1)$. Furthermore, we will only be able to give recursions for higher derivatives in terms of lower derivatives and these thus still depend on an explicit value for $\psi'(1)$. Nevertheless, we already determined a_1 in Lemma 3.18. Another approach towards a_1 will be ginven in Section 3.3.6.

The second derivative

For $i = N-1$, recursion (3.14) gives

$$\psi_{N-2} = \frac{2}{2+Nu\nu_1}\Big[(2+s+\frac{N}{N-1}u\nu_0 + Nu\nu_1)\psi_{N-1} - \frac{s}{N}\Big].$$

So, for the second derivative we have

$$\begin{aligned} N^2\Big[\psi_N - 2\psi_{N-1} + \psi_{N-2}\Big] &= N^2\Big[-2\psi_{N-1} + \psi_{N-2}\Big] \\ &= \frac{2N^2}{2+Nu\nu_1}\Big[-(2+Nu\nu_1)\psi_{N-1} + \frac{2+Nu\nu_1}{2}\psi_{N-2}\Big] \\ &= \frac{2N^2}{2+Nu\nu_1}\Big[(-2-Nu\nu_1+2+s+\frac{N}{N-1}u\nu_0 + N\nu_1)\psi_{N-1} - \frac{s}{N}\Big] \\ &= \frac{2}{2+Nu\nu_1}\Big[-(\frac{N}{N-1}Nu\nu_0 + Ns)(-N\psi_{N-1}) - Ns\Big]. \end{aligned}$$

Thus,

$$\psi''(1) = \frac{2}{2+\theta\nu_1}\Big[-(\theta\nu_0+\sigma)\psi'(1) - \sigma\Big].$$

The third derivative

For higher derivatives it is useful to replace ψ_{N-n} according to the recursion, as we did for the second derivative. For $k = N-n+1$ we have from (3.16)

$$\begin{aligned} \psi_{N-n} =& \frac{n}{n+Nu\nu_1}\Big[(2+s+\frac{N}{N-n+1}u\nu_0 + \frac{N}{n-1})\psi_{N-n+1} \\ &- (1+s+\frac{N}{N-n+2}u\nu_0)\psi_{N-n+2} - \frac{s}{N}\Big]. \end{aligned} \tag{3.30}$$

This equation holds for $n = 2, \ldots, N$.

The crucial simplification comes from the fact that the term $\frac{s}{N}$ does not depend on k. An equivalent expression to (3.14) is

$$\begin{aligned}\frac{s}{N} =&(2+s+\frac{N}{k}u\nu_0+\frac{N}{N-k}u\nu_1)\psi_k-(1+s+\frac{N}{k+1}u\nu_0)\psi_{k+1}\\ &-(1+\frac{N}{N+1-k}u\nu_1)\psi_{k-1}.\end{aligned}$$

We evaluate it for $k=N-n+2$ and plug it into (3.30):

$$\begin{aligned}\psi_{N-n}=\frac{n}{n+Nu\nu_1}\Big[&(3+s+\frac{N}{N-n+1}u\nu_0+\frac{2N}{n-1}u\nu_1)\psi_{N-n+1}\\ &-(3+2s+\frac{2N}{N-n+2}u\nu_0+\frac{N}{n-2}u\nu_1)\psi_{N-n+2}\\ &+(1+s+\frac{N}{N-n+3}u\nu_0)\psi_{N-n+3}\Big],\end{aligned}\tag{3.31}$$

which holds for $n=3,\dots,N$.

For the calculation of the third derivative we evaluate (3.31) for $n=3$ and plug it into (3.29):

$$\begin{aligned}&N^3[\psi_N-3\psi_{N-1}+3\psi_{N-2}-\psi_{N-3}]\\ &\quad=\frac{3N^3}{3+Nu\nu_1}\Big[(3+Nu\nu_1)(-\psi_{N-1}+\psi_{N-2})-\frac{3+Nu\nu_1}{3}\psi_{N-3}\Big]\\ &\quad=\frac{3N^3}{3+Nu\nu_1}\Big[(3+Nu\nu_1)(-\psi_{N-1}+\psi_{N-2})\\ &\qquad+(3+2s+2\frac{N}{N-1}u\nu_0+Nu\nu_1)\psi_{N-1}-(3+s+\frac{N}{N-2}u\nu_0+Nu\nu_1)\psi_{N-2}\Big]\\ &\quad=\frac{3}{3+Nu\nu_1}\Big[N^2(-2\psi_{N-1}+\psi_{N-2})(-Ns-\frac{N}{N-2}Nu\nu_0)\\ &\qquad-\frac{2N^2}{(N-1)(N-2)}Nu\nu_0N\psi_{N-1}\Big].\end{aligned}$$

Thus,

$$\psi^{(3)}(1)=\frac{3}{3+\theta\nu_1}\Big[-(\sigma+\theta\nu_0)\psi''(1)+2\theta\nu_0\psi'(1)\Big].$$

Higher derivatives

For the $n-$th derivative we can proceed in the same way as we did for the third derivative: replace ψ_{N-n} by formula (3.31), and sort all terms in the right manner. This step must be executed several times. The following lemma gives a formula for the product after the $k-$th step.

Lemma 3.26. *Let $0 \le k \le n \le N$ and ψ_j, $0 \le j \le N$, as given by (3.14) and $\psi_0 = \psi_N = 0$, then*

$$\begin{aligned}
\sum_{j=0}^{n}(-1)^j\binom{n}{j}\psi_{N-j} = \frac{n}{n+Nu\nu_1}\Big[&\frac{n+Nu\nu_1}{n}\sum_{j=1}^{n-k}(-1)^j\binom{n}{j}\psi_{N-j}\\
&- s\sum_{j=n-k+1}^{n-1}(-1)^j\binom{n-1}{j}\psi_{N-j}\\
&- u\nu_0\sum_{\ell=n-k+1}^{n-1}\Big((-1)^{n-\ell-1}\frac{(n-1)!N(N-\ell-1)!}{\ell!(N-n+1)!}\sum_{j=n-k+1}^{\ell}(-1)^j\binom{\ell}{j}\psi_{N-j}\Big)\\
&+ (-1)^{n-k+1}\binom{n-3}{k-1}(1+\frac{Nu\nu_1}{n-k+1})\psi_{N-n+k-1}\\
&- (-1)^{n-k}\binom{n-3}{k-2}(3+2s+\frac{2Nu\nu_0}{N-n+k}+\frac{Nu\nu_1}{n-k})\psi_{N-n+k}\\
&+ (-1)^{n-k+1}\binom{n-3}{k-3}(1+s+\frac{Nu\nu_0}{N-n+k})\psi_{N-n+k}\\
&+ (-1)^{n-k}\binom{n-3}{k-2}(1+s+\frac{Nu\nu_0}{N-n+k+1})\psi_{N-n+k+1}\Big].
\end{aligned} \tag{3.32}$$

Proof. The proof is per induction, the calculations are straightforward. For the basis we replace ψ_{N-n} via equation (3.31) and transform the terms in the right way, the inductive step then is done the same way. We only present the inductive step beginning from the right-hand side of (3.32).

$$\begin{aligned}
&\frac{n+Nu\nu_1}{n}\sum_{j=1}^{n-k}(-1)^j\binom{n}{j}\psi_{N-j} - s\sum_{j=n-k+1}^{n-1}(-1)^j\binom{n-1}{j}\psi_{N-j}\\
&- u\nu_0\sum_{\ell=n-k+1}^{n-1}\Big((-1)^{n-\ell-1}\frac{(n-1)!N(N-\ell-1)!}{\ell!(N-n+1)!}\sum_{j=n-k+1}^{\ell}(-1)^j\binom{\ell}{j}\psi_{N-j}\Big)\\
&+ (-1)^{n-k+1}\binom{n-3}{k-1}(1+\frac{Nu\nu_1}{n-k+1})\psi_{N-n+k-1}\\
&- (-1)^{n-k}\binom{n-3}{k-2}(3+2s+\frac{2Nu\nu_0}{N-n+k}+\frac{Nu\nu_1}{n-k})\psi_{N-n+k}\\
&+ (-1)^{n-k+1}\binom{n-3}{k-3}(1+s+\frac{Nu\nu_0}{N-n+k})\psi_{N-n+k}\\
&+ (-1)^{n-k}\binom{n-3}{k-2}(1+s+\frac{Nu\nu_0}{N-n+k+1})\psi_{N-n+k+1}
\end{aligned} \tag{3.33}$$

$$
\begin{aligned}
=&\frac{n+Nu\nu_1}{n}\sum_{j=1}^{n-k-1}(-1)^j\binom{n}{j}\psi_{N-j}-s\sum_{j=n-k+1}^{n-1}(-1)^j\binom{n-1}{j}\psi_{N-j}\\
&-u\nu_0\sum_{\ell=n-k+1}^{n-1}(-1)^{n-\ell-1}\frac{(n-1)!}{\ell!}\frac{N}{(N-n+1)\cdots(N-\ell)}\sum_{j=n-k+1}^{\ell}(-1)^j\binom{\ell}{j}\psi_{N-j}\\
&+(-1)^{n-k}\frac{1}{n}\binom{n}{n-k}\psi_{N-n+k}(n+Nu\nu_1)\\
&+(-1)^{n-k+1}\binom{n-3}{k-1}(3+s+u\nu_0\frac{N}{N-n+k}+u\nu_1\frac{2N}{n-k})\psi_{N-n+k}\\
&-(-1)^{n-k+1}\binom{n-3}{k-1}(3+2s+u\nu_0\frac{2N}{N-n+k+1}+u\nu_1\frac{N}{n-k-1})\psi_{N-n+k+1}\\
&+(-1)^{n-k+1}\binom{n-3}{k-1}(1+s+u\nu_0\frac{N}{N-n+k+2})\psi_{N-n+k+2}\\
&-(-1)^{n-k}\binom{n-3}{k-2}(3+2s+2u\nu_0\frac{N}{N-n+k}+u\nu_1\frac{N}{n-k})\psi_{N-n+k}\\
&+(-1)^{n-k+1}\binom{n-3}{k-3}(1+s+u\nu_0\frac{N}{N-n+k})\psi_{N-n+k}\\
&+(-1)^{n-k}\binom{n-3}{k-2}(1+s+u\nu_0\frac{N}{N-n+k+1})\psi_{N-n+k+1}\\
=&\frac{n+Nu\nu_1}{n}\sum_{j=1}^{n-k-1}(-1)^j\binom{n}{j}\psi_{N-j}-s\sum_{j=n-k+1}^{n-1}(-1)^j\binom{n-1}{j}\psi_{N-j}\\
&-u\nu_0\sum_{\ell=n-k+1}^{n-1}(-1)^{n-\ell-1}\frac{(n-1)!}{\ell!}\frac{N}{(N-n+1)\cdots(N-\ell)}\sum_{j=n-k+1}^{\ell}(-1)^j\binom{\ell}{j}\psi_{N-j}\\
&+(-1)^{n-k}(n+Nu\nu_1)\frac{(n-1)!}{k!(n-k)!}\psi_{N-n+k}\\
&+(-1)^{n-k}(s+u\nu_0\frac{N}{N-n+k})\left(-\binom{n-3}{k-1}-2\binom{n-3}{k-2}-\binom{n-3}{k-3}\right)\psi_{N-n+k}\\
&+(-1)^{n-k-1}\left(\binom{n-3}{k-1}(3+u\nu_1\frac{2N}{n-k})+\binom{n-3}{k-2}(3+u\nu_1\frac{N}{n-k})\right.\\
&\qquad\left.+\binom{n-3}{k-3}\right)\psi_{N-n+k}\\
&-(-1)^{n-k+1}\binom{n-3}{k-1}(3+2s+u\nu_0\frac{2N}{N-n+k+1}+u\nu_1\frac{N}{n-k-1})\psi_{N-n+k+1}\\
&+(-1)^{n-k}\binom{n-3}{k-2}(1+s+u\nu_0\frac{N}{N-n+k+1})\psi_{N-n+k+1}\\
&+(-1)^{n-k-1}\binom{n-3}{k-1}(1+s+u\nu_0\frac{N}{N-n+k+2})\psi_{N-n+k+2}
\end{aligned}
$$

$$
\begin{aligned}
=&\frac{n+Nu\nu_1}{n}\sum_{j=1}^{n-k-1}(-1)^j\binom{n}{j}\psi_{N-j}-s\sum_{j=n-k}^{n-1}(-1)^j\binom{n-1}{j}\psi_{N-j}\\
&-u\nu_0\sum_{\ell=n-k}^{n-1}(-1)^{n-\ell-1}\frac{(n-1)!}{\ell!}\frac{N}{(N-n+1)\cdots(N-\ell)}\sum_{j=n-k}^{\ell}(-1)^j\binom{\ell}{j}\psi_{N-j}\ (*)\\
&+(-1)^{n-k}\binom{n-3}{k}(1+\frac{u\nu_1}{n-k}N)\psi_{N-n+k}\\
&-(-1)^{n-k+1}\binom{n-3}{k-1}(3+2s+u\nu_0\frac{2N}{N-n+k+1}+u\nu_1\frac{N}{n-k-1})\psi_{N-n+k+1}\\
&+(-1)^{n-k}\binom{n-3}{k-2}(1+s+u\nu_0\frac{N}{N-n+k+1})\psi_{N-n+k+1}\\
&+(-1)^{n-k+1}\binom{n-3}{k-1}(1+s+u\nu_0\frac{N}{N-n+k+2})\psi_{N-n+k+2}.
\end{aligned}
$$

The step in line $*$ may still be unclear. For completeness, we must check the coefficients of ψ_{N-n+k}. The coefficients corresponding to s fit obviously, corresponding to $Nu\nu_1$ and 1 they do so, too. After all, it is not obvious that the coefficients of $Nu\nu_0$ fit. Some calculation gives that the difference between

$$
-u\nu_0\sum_{\ell=n-k+1}^{n-1}(-1)^{n-\ell-1}\frac{(n-1)!}{\ell!}\frac{N}{(N-n+1)\cdots(N-\ell)}\sum_{j=n-k+1}^{\ell}(-1)^j\binom{\ell}{j}\psi_{N-j}
$$

and

$$
-u\nu_0\sum_{\ell=n-k}^{n-1}(-1)^{n-\ell-1}\frac{(n-1)!}{\ell!}\frac{N}{(N-n+1)\cdots(N-\ell)}\sum_{j=n-k}^{\ell}(-1)^j\binom{\ell}{j}\psi_{N-j}
$$

is

$$
-u\nu_0\sum_{\ell=1}^{k}(-1)^{n-k+\ell-1}\frac{(n-1)!}{(n-k)!(k-\ell)!}\frac{N}{(N-n+1)\cdots(N-n+\ell)}.
$$

This equals

$$
-u\nu_0(-1)^{n-k}\frac{N}{N-n+k}\binom{n-1}{k-1}\psi_{N-n+k}
$$

due to Lemma 3.27. □

Lemma 3.27. *Let $k \in \mathbb{N}$, $a \leq 0$, then*

$$\sum_{\ell=1}^{k}(-1)^{\ell-1}\frac{(k-1)!}{(k-\ell)!}\frac{1}{(a+1)\cdots(a+\ell)} = \frac{1}{(a+k)}.$$

Proof. The proof is a straight forward induction:

$$\begin{aligned}
&\sum_{\ell=1}^{k+1}(-1)^{\ell-1}\frac{k!}{(k+1-\ell)!(a+1)\cdots(a+\ell)} \\
&\qquad = \sum_{\ell=0}^{k}(-1)^{\ell}\frac{k!}{(k-\ell)!(a+1)\cdots(a+\ell+1)} \\
&\qquad = -\frac{k}{a+1}\sum_{\ell=1}^{k}(-1)^{\ell-1}\frac{(k-1)!}{(k-\ell)!}\frac{1}{(a+2)\cdots(a+1+\ell)} + \frac{1}{a+1} \\
&\qquad = -\frac{k}{a+1}\frac{1}{a+1+k} + \frac{1}{a+1} = \frac{1}{a+k+1}.
\end{aligned}$$

□

With Lemma 3.26 follows

Theorem 3.28. *For $n \geq 3$ and ψ given by the boundary value problem* (3.26) *we have*

$$\psi^{(n)}(1) = \frac{n}{n+\theta\nu_1}\Big[-\sigma\psi^{(n-1)}(1) - \theta\nu_0\sum_{k=1}^{n-1}(-1)^{k-1}\frac{(n-1)!}{(n-k)!}\psi^{(n-k)}(1)\Big]$$

or equivalently

$$\psi^{(n)}(1) = \frac{n}{n+\theta\nu_1}\Big[-(n-1+\sigma+\theta\nu_0+\theta\nu_1)\sigma\psi^{(n-1)}(1) - (n-1)\sigma\psi^{(n-2)}(1)\Big],$$

which holds for $n \geq 4$.

Proof. We evaluate the result in Lemma 3.26 at $k = n-1$:

$$\begin{aligned}
\sum_{k=0}^{n}(-1)^k\binom{n}{k}\psi_{N-k} = \frac{n}{n+Nu\nu_1}\Big[&-(n+Nu\nu_1)\psi_{N-1} - s\sum_{j=2}^{n-1}(-1)^j\binom{n-1}{j}\psi_{N-j} \\
&- u\nu_0\sum_{\ell=2}^{n-1}(-1)^{n-\ell-1}\frac{(n-1)!N(N-\ell-1)!}{\ell!(N-n+1)!}\sum_{j=2}^{n-1}(-1)^j\binom{\ell}{j}\psi_{N-j} \\
&+ (n-3)(1+s+\frac{Nu\nu_0}{N-1})\psi_{N-1} - (1+s+u\nu_0)\psi_N \\
&+ (3+2s+2\frac{Nu\nu_0}{N-1}+Nu\nu_1)\psi_{N-1}\Big] \\
=\frac{n}{n+Nu\nu_1}\Big[&- s\sum_{j=2}^{n-1}(-1)^j\binom{n-1}{j}\psi_{N-j} \\
&- u\nu_0\sum_{\ell=2}^{n-1}(-1)^{n-\ell-1}\frac{(n-1)!N(N-\ell-1)!}{\ell!(N-n+1)!}\sum_{j=2}^{n-1}(-1)^j\binom{\ell}{j}\psi_{N-j} \\
&+ \Big((n-1)(s+\frac{Nu\nu_0}{n-1})\Big)\psi_{N-1}\Big] \\
=\frac{n}{n+Nu\nu_1}\Big[&- s\sum_{j=1}^{n-1}(-1)^j\binom{n-1}{j}\psi_{N-j} \\
&- u\nu_0\sum_{\ell=1}^{n-1}(-1)^{n-\ell-1}\frac{(n-1)!N(N-\ell-1)!}{\ell!(N-n+1)!}\sum_{j=1}^{n-1}(-1)^j\binom{\ell}{j}\psi_{N-j}\Big].
\end{aligned}$$

From this, the first assertion in the theorem follows as $N\to\infty$.

The first assertion is equivalent with

$$\begin{aligned}
(n+\theta\nu_1)\psi^{(n)}(1) + n\sigma\psi^{(n-1)}(1) &= -\theta\nu_0\sum_{k=1}^{n-1}(-1)^{k-1}\frac{n!}{(n-k)!}\psi^{(n-k)}(1) \\
\Leftrightarrow -(n+\theta\nu_1)\psi^{(n)}(1) - n\sigma\psi^{(n-1)}(1) &= -\theta\nu_0\sum_{k=2}^{n}(-1)^{k-1}\frac{n!}{(n+1-k)!}\psi^{(n+1-k)}(1),
\end{aligned}$$

which we plug in the expression for the $(n+1)$th derivative again:

$$\begin{aligned}
\psi^{(n+1)}(1) &= \frac{n+1}{n+1+\theta\nu_1}\Big[-\sigma\psi^{(n)}(1) - \theta\nu_0\sum_{k=1}^{n}(-1)^{k-1}\frac{n!}{(n+1-k)!}\psi^{(n+1-k)}(1)\Big] \\
&= \frac{n+1}{n+1+\theta\nu_1}\Big[-\sigma\psi^{(n)}(1) - \theta\nu_0\psi^{(n)}(1) - (n+\theta\nu_1)\psi^{(n)}(1) - n\sigma\psi^{(n-1)}(1)\Big],
\end{aligned}$$

which is the second assertion. □

With Proposition 3.24,

$$a_{n-2} = \sum_{k=1}^{n-1} (-1)^{k-1} \frac{(n-1)!}{(n-k)!} \psi^{(n-k)}(1).$$

This term emerges in the first assertion in Theorem 3.28. So,

$$\psi^{(n)}(1) = \frac{n}{n+\theta\nu_1}\Big[\big(-\sigma-\theta\nu_0\big)\psi^{(n-1)}(1) + \theta\nu_0(n-1)!(-1)^{n-2}a_{n-2}\Big].$$

Together with Proposition 3.24 we can deduce for $n \geq 3$

$$\begin{aligned}
a_n =& (-1)^n \frac{\psi^{(n)}(1)}{n!} + a_{n-1} \\
=& \frac{(-1)^n}{n!}\frac{n}{n+\theta\nu_1}\big[\big(-\sigma-\theta\nu_0\big)\psi^{(n-1)}(1) + \theta\nu_0(n-1)!(-1)^{n-2}a_{n-2}\big] + a_{n-1} \\
=& \frac{1}{n+\theta\nu_0}\Big[(\sigma+\theta\nu_0)\frac{(-1)^{n-1}\psi^{(n-1)}(1)}{(n-1)!} + \theta\nu_0 a_{n-2}\Big] + a_{n-1} \\
=& \frac{1}{n+\theta\nu_0}\Big[(\sigma+\theta\nu_0)(a_{n-1}-a_{n-2}) + \theta\nu_0 a_{n-2}\Big] + a_{n-1} \\
=& \frac{1}{n+\theta\nu_1}\Big[(n+\sigma+\theta)a_{n-1} - \sigma a_{n-2}\Big],
\end{aligned}$$

which is the same recursion that one may deduce from the definition of the λ_j in Theorem 3.7.

We need to determine the seed value a_2:

$$\begin{aligned}
a_2 =& -\psi'(1) + \frac{1}{2}\psi''(1) \\
=& -\psi'(1) + \frac{2}{2+\theta\nu_1}\Big[-(\theta\nu_0+\sigma)\psi'(1) - \sigma\Big] \\
=& a_1 + \frac{2}{2+\theta\nu_1}\Big[(\theta\nu_0+\sigma)a_1 - \sigma\Big] \\
=& \frac{2}{2+\theta\nu_1}\Big[-(2+\theta+\sigma)a_1 - \sigma\Big].
\end{aligned}$$

We already determined a_1 in Lemma 3.18, so we may state

Theorem 3.29. *The coefficients a_j in ansatz (3.27) are given by*

$$a_1 = \frac{\sigma}{1+\nu_1}(1-\tilde{p}), \quad \textit{and} \quad a_2 = \frac{2}{2+\theta\nu_1}\Big[\frac{(2+\theta+\sigma)\sigma}{1-\theta\nu_1}(1-\tilde{p}) - \sigma\Big]$$

and for $j \geq 3$ by the recursion

$$a_j = \frac{1}{j+\theta\nu_1}\Big[(j+\sigma+\theta)a_{j-1} - \sigma a_{j-2}\Big].$$

□

Remark 3.30. In this section, we presented an alternative approach towards the conditional distribution h_k. Since it is derived directly from the elementary model, this approach may allow to find a heuristic interpretation of the coefficients a_j. In Sections 3.4 and 3.5, we present some results in this context.

3.3.6 Alternative solution of the boundary value problem

In the previous sections we conducted and solve a power series approach to the boundary value problem for h. Nevertheless, one may also directly solve the differential equation and impose the boundary conditions. This solution permits a reasonably faster way of determining a_1. Again, we start with the boundary value problem (3.26)

$$\psi''(x) + \Big(\frac{\theta\nu_0}{x} - \frac{\theta\nu_1}{1-x} + \sigma\Big)\psi'(x) - \Big(\frac{\theta\nu_0}{x^2} + \frac{\theta\nu_1}{(1-x)^2}\Big)\psi(x) = -\sigma$$

on $(0,1)$ with boundary conditions $\psi(0) = \psi(1) = 0$.

We define

$$\begin{aligned} a(x) :=& x(1-x), \\ b(x) :=& \theta\nu_0(1-x) - \theta\nu_1 x + \sigma x(1-x), \quad \text{and} \\ c(x) :=& \theta\nu_0\frac{1-x}{x} + \theta\nu_1\frac{x}{1-x}. \end{aligned}$$

Then the differential equation reads

$$\psi''(x) + \frac{b(x)}{a(x)}\psi'(x) - \frac{c(x)}{a(x)}\psi(x) = -\sigma.$$

We first consider the following first order linear homogeneous differential equation

$$\varphi'(x) + \frac{b(x)}{a(x)}\varphi(x) = 0.$$

A solution is

$$\varphi_1(x) = \exp\Big(-\int^x \frac{b(y)}{a(y)}dy\Big) = x^{-\theta\nu_0}(1-x)^{-\theta\nu_1}e^{-\sigma x}.$$

Since

$$\begin{aligned} \varphi_1''(x) &= -\frac{b(x)}{a(x)}\varphi_1'(x) - \frac{d}{dx}\frac{b(x)}{a(x)} \quad \text{and} \\ \frac{d}{dx}\left(\frac{b(x)}{a(x)}\right) &= \frac{c(x)}{a(x)}, \end{aligned} \tag{3.34}$$

φ_1 solves the homogeneous equation

$$\psi''(x) + \frac{b(x)}{a(x)}\psi'(x) - \frac{c(x)}{a(x)}\psi(x) = 0.$$

According to the method of variation of parameters we know that

$$z(x) = \varphi_1 \int^x \frac{h(x)}{\varphi_1(y)} dy$$

is the solution of the inhomogeneous equation

$$\varphi'(x) + \frac{b(x)}{a(x)}\varphi(x) = h(x).$$

Set $h(x) = -\sigma x + c$,

$$\varphi'(x) + \frac{b(x)}{a(x)}\varphi(x) = -\sigma x + c.$$

Differentiating of this equation gives

$$\psi''(x) + (\frac{\theta\nu_0}{x} - \frac{\theta\nu_1}{1-x} + \sigma)\psi'(x) - (\frac{\theta\nu_0}{x^2} + \frac{\theta\nu_1}{(1-x)^2})\psi(x) = -\sigma$$

which is our original differential equation. So,

$$z(x) = -\sigma\varphi_1(x) \int^x \frac{y+c}{\varphi_1(y)} dy$$

solves the differential equation.

We still have to consider boundary conditions. Since $\psi(0) = 0$, set the lower integration bound 0 and due to $\psi(1) = 0$, choose c such that

$$-c = \frac{\int_0^1 \frac{y}{\varphi_1(y)} dy}{\int_0^1 \frac{1}{\varphi_1(y)} dy} = \frac{\int_0^1 y^{\theta\nu_0+1}(1-y)^{\theta\nu_1} e^{\sigma y} dy}{\int_0^1 y^{\theta\nu_0}(1-y)^{\theta\nu_1} e^{\sigma y} dy} = \tilde{p}, \tag{3.35}$$

where $\tilde{p}$ is defined as in Lemma 3.18.

Remark 3.31. The constant $\tilde{p}$ is both the expectation of the frequency of fit alleles under the variance-biased stationary distribution with density $C\sigma y(1-y)\pi$ (where C is a normalizing constant) and also the probability that a sample of three individuals from a stationary population contains exactly two individuals of type 0 conditional on it containing at least one individual of each genotype [50].

Lemma 3.8 tells us that this solution is unique. So, we may state

Theorem 3.32. *The unique solution of the boundary value problem* (3.26) *is*

$$\psi(x) = \sigma \int_0^x (\tilde{p} - y)\Big(\frac{y}{x}\Big)^{\theta\nu_0}\Big(\frac{1-y}{1-x}\Big)^{\theta\nu_1} e^{\sigma(y-x)}dy. \tag{3.36}$$

□

With this solution we can calculate the coefficient a_1.

Corollary 3.33.

$$a_1 = \frac{\sigma}{1+\theta\nu_1}(1-\tilde{p}).$$

Proof.

$$\psi'(x) = \frac{d}{dx}\psi(x) = \sigma(\tilde{p} - x) + \psi(x)(-\sigma - \theta\nu_0 x^{-1} + \theta\nu_1(1-x)^{-1}).$$

Since $\psi'(x)$ can be continuously extended to $[0,1]$ (Lemma 3.8), we have

$$\psi'(1) = \lim_{x\nearrow 1}\psi'(x) = \sigma(\tilde{x} - 1) + \theta\nu_1 \lim_{x\nearrow 1}\frac{\psi(x)}{1-x} = \sigma(\tilde{p} - 1) - \theta\nu_1\psi'(1).$$

So,

$$a_1 = -v'(1) = -\psi'(1) = \frac{\sigma}{1+\theta\nu_1}(1-\tilde{p}).$$

□

Remark 3.34. Taylor derives the same solution [50, Eq. (23)]. He pursues a more complicated power series approach. The simplicity of our approach relies on the properties of the coefficient functions in (3.34).

3.4 No mutation

In Section 3.3.3 we calculated an explicit formula for $h(x)$ and the seed value a_1 directly from the solution of the recursion (3.14). In the case without mutation this derivation simplifies considerably. This hinges on the fact that the solution of the recursion is then much simpler since $A_\ell^{\ell+m} = 1$, $B_\ell^{\ell+m} = (1+s)^{m+1}$:

$$\psi_{N-k} = \sum_{\ell=1}^{k}(1+s)^{k-\ell}[\psi_{N-1} - (\ell-1)\frac{s}{N}] = \sum_{j=0}^{k-1}(1+s)^{j}[\psi_{N-1} - (k-j-1)\frac{s}{N}]. \tag{3.37}$$

Evaluating this at $\psi_0 = 0$ leads to

$$\begin{aligned} 0 = \psi_0 &= \sum_{j=0}^{N-1}(1+s)^j[\psi_{N-1} - (N-j-1)\frac{s}{N}] \\ &= N\psi_{N-1}\frac{1}{N}\sum_{j=0}^{N-1}(1+\frac{Ns}{N})^{N\frac{j}{N}} - Ns\frac{1}{N}\sum_{j=0}^{N-1}\frac{N-j-1}{N}(1+\frac{Ns}{N})^{N\frac{j}{N}}. \end{aligned}$$

Letting N tend to infinity, we receive an expression for $\psi'(1)$:

$$\psi'(1) = -\frac{\sigma\int_0^1(1-y)e^{\sigma y}dy}{\int_0^1 e^{\sigma y}dy}.$$

With

$$\tilde{p} := \frac{\int_0^1 ye^{\sigma y}dy}{\int_0^1 e^{\sigma y}dy},$$

which is the $\tilde{p}$ from equation (3.35), we again have

$$\psi'(1) = \sigma(\tilde{p}-1).$$

We can also calculate $\psi(x)$ from equation (3.37). Define a sequence $k_N \in \mathbb{N}$ with $k_N \leq N$ such that $\frac{k_N}{N} \to x$ as $N \to \infty$ for some $x \in [0,1]$. Then

$$\psi_{N-k_N} = N\psi_{N-1}\frac{1}{N}\sum_{j=0}^{k_N-1}(1+\frac{Ns}{N})^{N\frac{j}{N}} - Ns\frac{1}{N}\sum_{j=0}^{k_N-1}\frac{k_N-j-1}{N}(1+\frac{Ns}{N})^{N\frac{j}{N}}.$$

Letting $N \to \infty$ leads to

$$\begin{aligned} \psi(1-x) &= -\psi'(1)\int_0^x e^{\sigma y}dy - \sigma\int_0^x(x-y)e^{\sigma y}dy \\ &= \sigma\int_0^x(1-x-\tilde{p}+y)e^{\sigma y}dy \\ &= \sigma\int_0^{1-x}(\tilde{p}-y)e^{\sigma y}dy. \end{aligned}$$

With (3.18) we can give a simple formula for h_k in the case without mutation:

$$
\begin{aligned}
h_{N-k} &= \frac{N-k}{N} + \psi_{N-k} \\
&= \frac{N-k}{N} + \sum_{\ell=1}^{k} A_\ell^k B_\ell^{k-1} s \left[\frac{\sum_{\ell=0}^{N-1} \frac{N-1-\ell}{N} A_{N-\ell}^N B_{N-\ell}^{N-1}}{\sum_{\ell=0}^{N-1} A_{N-\ell}^N B_{N-\ell}^{N-1}} - \frac{(\ell-1)}{N} \right] \\
&= \frac{1}{N \sum_{j=0}^{N-1} (1+s)^j} \Bigg[(N-k) \sum_{i=0}^{N-1} (1+s)^i \\
&\qquad + \big((1+s)^k - 1\big) \sum_{i=0}^{N-1} (N-1-i)(1+s)^i \\
&\qquad - \big((1+s)^N - 1\big) \sum_{i=0}^{k-1} (k-1-i)(1+s)^i \Bigg] \\
&= \frac{1}{N \sum_{j=0}^{N-1} (1+s)^j} \Bigg[\sum_{i=0}^{k-1} \big(N-k-N+1+i+k-1-i\big)(1+s)^i \\
&\qquad + \sum_{i=k}^{N-1} (N-k-N+1+i)(1+s)^i \\
&\qquad + \sum_{i=k}^{N+k-1} (N-1+k-i)(1+s)^i \\
&\qquad - \sum_{i=N}^{N+k-1} (N-1+k-i)(1+s)^i \Bigg] \\
&= \frac{1}{N \sum_{j=0}^{N-1} (1+s)^j} \left[(N-k) \sum_{i=k}^{N-1} (1+s)^i + k \sum_{i=k}^{N-1} (1+s)^i \right] \\
&= \frac{\sum_{i=k}^{N-1} (1+s)^i}{\sum_{j=0}^{N-1} (1+s)^j},
\end{aligned}
\tag{3.38}
$$

where we used in the second step that

$$
1 + s \sum_{i=0}^{N-1} (1+s)^i = (1+s)^N.
$$

Of course, this formula is well known (for a nice derivation via a martingale argument, see [17]). It motivates a new particle model which we present in Section 3.5.

Path mirroring principle

We found an alternative approach to this equation via a path mirroring principle which we shortly present here.

Equation (3.38) gives us the absorption probability of the fit individuals, given that there are k fit individuals in the current population. When we look at the "dual" event, absorption of the unfit individual, given that there are k unfit individuals present, we have:

$$1 - h_{N-k} = \frac{\sum_{j=0}^{k-1} (1+s)^j}{\sum_{j=0}^{N-1} (1+s)^j} = (1+s)^{-(N-k)} h_k. \tag{3.39}$$

We can deduce this relation not only from (3.38) but also from a path mirroring principle. To this end, consider a direct path from k to N, this means the embedded Markov chain $(X_i)_{i\in\mathbb{N}}$ has the states $k, k+1, k+2, \ldots, N$. The probability for this path is

$$\frac{(1+s)^{N-k}}{(2+s)^{N-k}}.$$

Look at the path $N-k, N-k-1, \ldots, 0$. Its probability is

$$\frac{1}{(2+s)^{N-k}}.$$

Of course, there are many other possible paths from k to N. For each of these paths we want to define a *mirrored* path. The original path always starts in k and ends in N, therefore the mirrored path starts in $N-k$. Whenever the original path increases by one, the mirrored path decreases by one and vice versa. Thus, the mirrored path always ends in 0. For example, the mirrored path of the direct path $k, k+1, \ldots, N$ is $N-k, N-k-1, \ldots, 0$. All the other paths contain loops, some states are visited more than once. In a loop the number of increments is obviously the same as the number of decrements. So, the probability of the loop in the original path and in the mirrored path are the same. This means, the probabilities of the original and the mirrored path always only differ by the factor $(1+s)^{N-k}$ which we already calculated for the direct paths. Since the probability for all paths from k to N, $N-k$ to 0 respectively, is h_k, $1-h_{N-k}$ respectively, this reasoning already establishes equation (3.39).

We do not know yet how to devolve this reasoning onto the situation with selection and mutation.

3.5 New particle model

In the Moran model with selection but without mutation, the conditional probability h_k reads

$$h_k = \frac{\sum_{i=1}^{k}(1+s)^{N-i}}{\sum_{i=0}^{N-1}(1+s)^i}$$

as we know from (3.38). We want to formulate a new particle model with the same empirical distribution in which we may assign an absorption probability to the summands on the right-hand side.

Again, we look at a population with N individuals of type 0 or 1. We introduce some spatial structure by enumerating them starting with the fit individuals. Each individual reproduces itself by rate 1, which we may regard as the neutral reproduction as in the Moran model. Then, a randomly chosen individual dies. This event also induces a change in the spatial structure. Given the parental individual is number j and the number of the individual that dies is i:

- If $i = j$, this corresponds to an empty event, nothing happens;
- if $i < j$, the offspring individual is numbered with $j - 1$, the individuals' numbers from $i + 1$ to $j - 1$ are decreased by one;
- if $i > j$, the offspring individual is numbered with $j + 1$, the individuals' numbers from $j + 1$ to $i - 1$ are augmented by one.

Furthermore, each individual reproduces at rate s. Again, the individual which is replaced is chosen randomly. This occurs with the constraint, that only those individuals are replaced whose number is greater than the parental one's. If their number is smaller, the offspring individual does not survive.

Remark 3.35. Effectively, this new model introduces N types with different fitness values. The individual with number 1 is the fittest, itself and its offspring cannot be harmed by selective reproduction events which happen at higher numbered individuals. So, this new model takes an idea from the look-down process.

We can look at it in another way: we start at time 0, and neutral and selective events mean arrows in the graphical representation. Then we resolve the arrows in the way that selective arrows are only valid if they end at an individual which is left to or at such an offspring. This resolving of selective arrows is similar to the procedure in the construction of the ancestral selection graph.

Remark 3.36. We can additionally keep ancestry in mind. Therefore, for each time $t \geq 0$, we define a partition $\mathcal{A}_t := \{A_t^i,\ i = 1, \ldots, N\}$ of the set $\{1, \ldots, N\}$ at time 0, where A_t^i contains all offspring of individual i at time t. Then, the sets A_t^i can be written as $A_t^i = \{\ell | k_1 \leq \ell \leq k_2\}$ for some $k_1 \leq k_2 \in \{1, \ldots, N\}$. Thus, reproduction events preserve the ordering of the individuals corresponding to the ancestry.

Lemma 3.37. *In the new particle model, look at the individuals numbered with* $1, \ldots, j$ *at time* 0. *Let* Xj_t *be the number of their offspring at time* t. *Then,* Xj_t *has the same law as* Y_t, *the empirical distribution from the Moran model, given* $Y_0 = j$.

Proof. Both chains can only jump by one. Let $Xj_t = k$. Since the offspring are the first k individuals (Remark 3.36), all selective reproduction events are successful. Thus, the rate for the transition $k \to k+1$ is

$$(N-k)(1+s_N)\frac{k}{N}.$$

On the other hand, no selective reproduction event that is initiated by an individual with number $m > k$ can delete one of the first k individuals. Thus, transition $k \to k-1$ happens at rate

$$k\frac{N-k}{N}.$$

These are the same jumps and rates as for Y_t. □

The following theorem is a direct consequence of this lemma.

Theorem 3.38. Y_t, *the number of fit individuals, has the same the law in the new and the old model.* □

The next theorem addresses the question at the beginning of this section. It gives an interpretation of the conditional probability h_k.

Theorem 3.39. *The absorption probability of individual* i *is*

$$\frac{(1+s)^{N-i}}{\sum_{j=0}^{N-1}(1+s)^j}.$$

Proof. We proof this by induction. With Lemma 3.37, it is obvious that the fixation probability of individual 1 is

$$\frac{(1+s)^{N-1}}{\sum_{j=0}^{N-1}(1+s)^j}.$$

With Lemma 3.37, we can determine the fixation probability of the first $(i-1)$th and of the first i individuals. The difference is

$$\frac{(1+s)^{N-i}}{\sum_{j=0}^{N-1}(1+s)^j}.$$

Since the fixation probability of the first $i-1$ individuals is, according to the induction hypothesis, the sum of the fixation probabilities of each single individual $1,\dots,i-1$, it is clear that this is exactly the additional fixation probability coming from individual i.

□

Remark 3.40. The new particle model permits an interpretation for the coefficient a_1. For this, we look at

$$\begin{aligned}\psi_{N-1} =& h_{N-1} - (1-\frac{1}{N}) = -\frac{1}{\sum_{j=0}^{N-1}(1+s)^j} + \frac{1}{N}\\ =& \frac{-N+1+\sum_{j=1}^{N-1}(1+s)^{N-j}}{N\sum_{j=0}^{N-1}(1+s)^j}.\end{aligned}$$

We look at the summands in the numerator. Each of them is the fixation probability for a particle. With

$$(1+s)^{N-j} = 1 + s\sum_{k=0}^{N-j-1}(1+s)^k,$$

we split the summands in the numerator into a "neutral" and a "selective" part. Thus, we have

$$N\psi_{N-1} = \frac{-N+N+s\sum_{k=0}^{N-2}(N-1-k)(1+s)^k}{\sum_{j=0}^{N-1}(1+s)^j} \xrightarrow[N\to\infty]{} a_1.$$

So, a_1 is the part of the fixation probability that is contributed by that offspring of the first $N-1$ individuals that arose by selective events.

3.6 Conclusion

In this chapter, we presented the approaches of Fearnhead and Taylor towards the common ancestor distribution and their solutions. We could give another way to this solution directly derived from the Moran model with selection and mutation.

However, we restricted ourselves to a two-type model with constant selection coefficient. In contrast to this, Fearnhead solved this problem for a multi-type model (with the restriction on only two fitness classes and constant selection coefficient). It should be possible to apply our ansatz onto this extension successfully.

Taylor only considered a two-type model, too. However, he could prove existence and uniqueness of the solution for the boundary value problem (3.10) with density-dependent selection (which means that the selection coefficient $\sigma = \sigma(x)$ depends on the frequency of fit alleles x. However, he does not present an explicit solution in this case. It seems that the first-step argument to derive the bvp should work. However, it is arguable whether the ansatz we followed to solve the bvp would be successful. For example, the independence of the term $\frac{s}{N}$, that leads to (3.30) would not be valid anymore.

The passage to the diffusion limit is a common approach in population genetics. Since in this limit the particle presentation gets lost, one is concerned with the question whether one may reach the results directly from the particle model or regain a particle presentation from the diffusion limit. An example for this is the look-down process. In this chapter we provided a contribution to these efforts. Especially, we presented another particle model in the case without mutation. Unfortunately, we have not been able to give a comparable model in the case with selection and mutation yet.

Danksagung

Ich möchte mich noch bei jenen bedanken, die an der Anfertigung dieser Arbeit unverzichtbaren Anteil hatten.

An erster Stelle gilt mein Dank Ellen Baake für die Betreuung der Arbeit und die Vergabe des Themas. Sie stand mir jederzeit mit Rat, Hilfe und auch Motivation zur Seite. Vielen Dank dafür.

Barbara Gentz danke ich für ihr Interesse an meiner Arbeit und die Bereitschaft zur Begutachtung.

Ich danke allen Mitgliedern der AG Biomathematik für anregende fachliche und nicht-fachliche Gespräche. Insbesondere waren der fachliche Austausch mit Sandra Kluth und der fachliche und allgemeinere mit Ute von Wangenheim immer sehr gewinnbringend.

Ich danke Fabian Meier und Ute von Wangenheim für das Korrekturlesen und Auffinden des ein oder anderen sprachlichen Fehlers.

Für die Erstellung dieser Arbeit bedurfte es doch ab und an der Ermunterung. Ich danke allen guten Freunden und meiner Schwester Thurid, die mich in dieser Hinsicht immer wieder unterstützt haben.

Bibliography

[1] Abraham, V.M. (1980). Linearising quadratic transformations in genetic algebras. *Proc. London Math. Soc.* **40,** 346-363.

[2] Aigner, M. (1979). *Combinatorial Theory.* Springer, Berlin.

[3] Baake, E.; Baake, M. (2003). An exactly solved model for mutation, recombination and selection. *Canad. J. Math.*, **55**, 3-41 and **60** (2008), 264–265 (Erratum).

[4] Baake, E., Baake, M., Bovier, A., Klein, M. (2005). An asymptotic maximum principle for essentially nonlinear evolution models. *J. Math. Biol.*, **50**, 83-114.

[5] Baake, E., Bialowons, Robert (2008). Ancestral processes with selection: Branching and Moran models. *Banach Center Publ.*, **80**, 33-52.

[6] Baake, E., Herms, I. (2008). Single-crossover dynamics: finite versus infinite populations. *Bull. Math. Biol.* **70,** 603–624.

[7] Baake, E., Hustedt, T. (2011). Moment closure in a Moran model with recombination. *Mark. Proc. and rel. Fields* **17,** 429–446.

[8] Baake, M. (2005). Recombination semigroups on measure spaces. *Monatsh. Math.* **146,** 267–278 and **150** (2007), 83–84 (Addendum).

[9] Barton, N., Etheridge, A.M., Sturm, A.K. (2004). Coalescence in a random background. *Ann. Appl. Prob.* **14,** 754 – 785.

[10] Bobrowski, A., Wojdyla, T., Kimmel, M. (2010). Asymptotic behavior of a Moran model with mutations, drift and recombination among multiple loci. *J. Math. Biol.* **61,** 455–473.

[11] Brémaud, P. (1999). *Markov Chains, Gibbs Fields, Monte Carlo Simulation, and Queues.* Springer, New York.

[12] Bürger, R. (2000). *The Mathematical Theory of Selection, Recombination and Mutation.* Wiley, Chichester.

[13] Burke, C., Rosenblatt, M. (1958). A Markovian function of a Markov chain. *Ann. Math. Stat.* **29,** 1112–1122.

[14] Dieckmann, U., Law, R. (2000). Relaxation projections and the method of moments. In: The Geometry of Ecological Interactions: Simplifying Spatial Complexity, eds Dieckmann, U., Law, R., Metz, J.A.J., Cambridge University Press, 412–455.

[15] Donnelly, P., Kurtz, T. (1996). A countable Representation of the Fleming-Viot Measure-Valued Diffusion. *Ann.Prob.* **24,** 698–742.

[16] Donnelly, P., Kurtz, T. (1999). Genealogical processes for Fleming-Viot models with selection and recombination. *Ann. Appl. Prob.* **9,** 1091–1148.

[17] Durrett, R. (2008). *Probability Models for DNA Sequence Evolution*, 2nd edn. Springer, New York.

[18] Etheridge, A. (2000). *An Introduction to Superprocesses*. AMS, Providence, RI.

[19] Ethier, S.N., Kurtz, T.G. (1986). *Markov Processes - Characterization and Convergence*. Wiley, New York. Reprint 2005.

[20] Ewens, W. (2004). *Mathematical Population Genetics*, 2nd edn. Springer, Berlin.

[21] Fearnhead, P. (2002). The common ancestor at a nonneutral locus. *J. App. Prob.*, **39,** 38–54.

[22] Graham, A. (1981). *Kronecker products and matrix calculus with applications.* Chichester, Horwood.

[23] Gonshor, H. (1971). Contributions to genetic algebras. *Proc. Edinburgh Math. Soc. (2)* **17,** 289–298.

[24] Haccou, P., Jagers, P., Vatutin, V.A. (2005). *Branching Processes: Variation, Growth, and Extintion of Populations*. Cambridge University Press, Cambridge.

[25] Haldane, J.B.S. (1930). Theoretical genetics of auto-polyploids. *J. Genetics* **22,** 359-372.

[26] Hausken, K., Moxnes, J.F. (2011). Systematization of a set of closure techniques. *Theor. Pop. Biol.*, **80,** 175–184.

[27] Heuch, I. (1973). The linear algebra for linked loci with mutation. *Math. Biosciences* **16,** 263-271.

[28] Hein, J., Schierupp, M., Wiuf, C.(2005) *Gene Genealogies, Variation and Evolution. A Primer in Coalescent Theory.* Oxford University Press, Oxford.

[29] Holgate, P. (1967). Sequences of powers in genetic algebras. *J. London Math. Soc.* **42,** 489-496.

[30] Holgate, P. (1968). The genetic algebra of k linked loci. *Proc.London Math. Soc.* **18,** 315-327.

[31] Kaplan, N.L., Darden, T., Hudson, R.R. (1988). The coalescent process in models with selection. *Genetics*, **120,** 819–829.

[32] Kemeney, J.G., Snell, J.L. (1981). *Finite Markov Chains.* Springer, New York.

[33] Kluth, S. (2010). *Vorfahrenprozesse mit Selektion und Mutation.* Diplomarbeit, Universität Bielefeld.

[34] Kingman, J.F.C. (1982). The coalescent. *Stoch. Proc. Appl.* **13,** 235–248.

[35] Kingman, J.F.C. (1982). On the genealogy of large populations. *J. Appl. Prob.* **19A,** 27–43.

[36] Krone, S., Neuhauser, C. (1997). Ancestral processes with selection. *Theor. Pop. Biol.* **51,** 210–237.

[37] Kurtz, T. G. (1971). Limit Theorems for Sequences of Jump Markov Processes Approximating Ordinary Differential Processes. *J. Appl. Prob.* **8,** 344–356.

[38] Lee, C.H., Kim, K. and Kim, P. (2009). A moment closure method for stochastic reaction networks. *J. Chem. Phys.* **130,** 134107.

[39] Levermore, C.D. (1996). Moment closure hierarchies for kinetic theories. *J. Stat. Phys.* **83,** 1021–1065.

[40] Lyubich, Y. (1992). *Mathematical Structures in Population Genetics.* Springer, New York.

[41] McHale, D., Ringwood, A. (1983). Haldane linearisation of baric algebras. *J. London Math. Soc.* (2) **28,** 17–26.

[42] Moran, P. (1958). Random processes in genetics. *Proc. Camb. Phil. Soc.* **54,** 60-71.

[43] Neuhauser, C., Krone, S. (1997). The genealogy of samples in models with selection. *Genetics* **145,** 519–534.

[44] Norris, J.R. (1997). *Markov Chains*, Cambridge University Press, Cambridge.

[45] Pfaffelhuber, P., Haubold, B., Wakolbinger, A. (2006). Approximate genealogies under genetic hitchhiking. *Genetics* **174,** 1995–2008.

[46] Reed, M. (1997). Algebraic structure of genetic inheritance. *Bull. AMS* **34,** 107–130.

[47] Reiersol, O. (1962). Genetic algebras studied recursively and by means of differential operators. *Math. Scand.* **10,** 25-44.

[48] Schafer, R.D. (1949). Structure of genetic algebras. *American J. of Mathematics*, **71,** 121–135.

[49] Stephens, M., Donnelly, P. (2003). Ancestral inference in population genetics models with selection. *Austral. New Zeal. J. Stat.* **51,** 901–931.

[50] Taylor, J. (2007). The Common Ancestor Process for a Wright-Fisher Diffusion. *Electron. J. Probab.*, **12**, 808–847.

[51] von Wangenheim, U. (2010). *unpublished work together with Ute von Wangenheim.*

[52] von Wangenheim, U., Baake, E., Baake, M. (2010). Single–crossover recombination in discrete time. *J. Math. Biol.* **60,** 727–760; `arXiv:0906.1678v1`.

[53] von Wangenheim, U.(2011). *Single-crossover recombination and ancestral recombination trees.* Dissertation, Universität Bielefeld.

[54] Wörz-Busekros, A. (1980). *Algebras in Genetics,* Lecture Notes in Biomathematics. **36,** Springer, Berlin.