universität wien

MASTERARBEIT

Titel der Masterarbeit

Parallel adaptation of a beneficial mutation with deterministic spread in a spatially structured population

Verfasser

Manuela Geiß, BSc

angestrebter akademischer Grad

Master of Science (MSc)

Wien, November 2015

Studienkennzahl lt. Studienblatt:	A 066 821
Studienrichtung lt. Studienblatt:	Mathematik
Betreuer:	Univ.-Prof. Dr. Joachim Hermisson

Contents

Acknowledgement

First of all I want to thank my supervisor Joachim Hermisson for giving me the chance to develop my own model during this thesis and supporting me with great patience throughout the whole process. I also want to thank my parents for their support over the last years of my studies, especially for the emotional support.

Abstract

Models treating the adaptation of a population to a new environment are often based on a single newly arisen beneficial mutation which then goes to fixation. However, also parallel adaptation generated by multiple mutations of similar phenotypic effect is possible and several examples are known. There already exist various models assuming a continuous space with global selection pressure as well as models treating a spatially structured population with local selection. In both models limited dispersal leads to a pattern of distinct areas formed by the propagation of mutations of separated origins.
In the present work we treat a model with discrete demes where we consider a deterministic spread of a beneficial mutation from any given deme to its neigboring demes. This already existed in continuous space, where the deterministic spread of an adaptation wave has been analyzed, but not in a discrete deme model.
We mainly focus on the one-dimensional model but also the case of the two- and the three-dimensional model will be treated.

1 Introduction

There exist many examples of parallel evolution of phenotypes within species caused by multiple, independent mutations at the same gene that induce the same or functionally equivalent adaptive phenotypic effect. Some well-studied examples are the cyclodiene insecticide resistance within several insect species (Ffrench Constant *et al.* (2000)) or the loss of pigmentation in Drosophila santomea due to a mutational inactivation of a specific *cis*-regulatory element through at least three independent loss-of-function alleles (Jeong *et al.* (2008)).
Parallel adaptation can occur in a panmictic population adapting to a global selection pressure where limited dispersal leads to a slow propagation of selected alleles allowing other beneficial alleles to arise and spread (see e.g. Ralph and Coop (2010)). However, in many known examples selection is local and rates of gene flow are low which leads to a spatially structured population where parallel adaptation is even more favored and produces a patchwork of the propagation areas of distinct selected alleles.
In this thesis we treat the case of a spatially structured population. The main questions that we consider here arised from Greven *et al.* (2014) where they were concerned with the time it takes for a strongly beneficial allele to eventually go to fixation in a spatially structured population.
From the panmictic case it is already known that the fixation time for large selection coefficient α is approximately $2\frac{\log(\alpha)}{\alpha}$. Although the fixation process may be slowed down by the spatial setting, Greven *et al.* showed that the same time-scale $\frac{\log(\alpha)}{\alpha}$ also applies in a structured population with migration. They found that the asymptotic of the fixation time as $\alpha \to \infty$ differs depending on the migration rate μ and came up with three regimes for the migration rate relative to the selection speed, namely with μ of order α, μ of order α^p $(0 \leq p \leq 1)$ and $\mu = \frac{1}{\log(\alpha)}$.
Using these results for our model, we consider migration from one colony to its neighbors to take place after a deterministic time after the arising of a new selected allele. To our knowlegde there exists no work about a model with such a deterministic feature.
Assuming the infinite alleles model, we are then interested in the spatial pattern formed by the spreading mutations that reflect their separate origins. More precisely, our main concern will be the expected proportion of new mutation origins and the propagation areas of these mutations which we will denote as clusters.
Let us now give a brief outline of the thesis. In the first part we introduce our model and apply it to the one-dimensional stepping stone model with a finite number of demes arranged in a circle. As it turns out that general results in this setting are not feasible

due to boundary conditions, we then modify the one-dimensional setting into a linear arrangement with infinitely many demes in the main part of the thesis. This modification leads to results about the expected frequency of new mutations and the distribution of cluster sizes, as well as to results about the correlation of sizes of neighboring clusters and the distance of adjacent cluster origins. In a further step we extend our model in a simplified version to two and three dimensions where we come up with similar results about the expected proportion of new mutations and the distribution of cluster sizes. In the last chapter of this thesis we finally draw a conclusion between our one-dimensional model and the model *coalescent with killings* by comparing results of our spatial model to results given by the Ewens' sampling formula.

2 The finite one-dimensional model with a circular arrangement of demes

In this chapter we introduce our model in detail and then start by looking at the model in one dimension. Here, finitely many demes are arranged in a circle and we are first interested in the expected number of spontaneous mutations that can occur until all demes finally carry a beneficial allele. It turns out that this can be explicitely calculated for a small number of demes, but for more demes, due to boundary conditions, no general solution that would be useful to answer our question, can be derived. Some details of the calculations for a small number of demes are presented in order to show what kind of problems occur for an arbitrary number of demes.

2.1 The model

We consider a spatially structured population model on k demes (for some $k \in \mathbb{N}$). The population is panmictic in all demes, and all demes are of equal size which remains constant for all time. Moreover, we assume this model to be a stepping stone model, hence migration is only possible between adjacent demes. After an environmental change, the population needs to adapt to its new environment where a global selection pressure acts. We assume that there exists some beneficial allele which is responsible for the adaptation and arises from the wild-type allele by a single mutation (e.g. a point mutation). This beneficial mutation occurs spontaneously and at the same rate λ in each deme and will then spread to all adjacent demes after a deterministic time if these demes do not carry the beneficial allele by then. By talking about an arising mutation, we always mean that the mutation arises successfully in a deme and immediately gets to fixation within that deme. Thus, in particular, we do not consider genetic drift and, as Greven *et al.* (2014) already treated it in detail, we do not care about the time it takes a mutation to get to fixation.
The following definition describes the dynamic of the process in detail.

Definition 2.1. *Consider the process* $\mathcal{Y} = (\underline{Y}(t))_{t\geq 0}$ *with state space* $(\{\circ\} \cup [0,1])^k$. *Here,* $Y_i(t) = \circ$ *means that deme* i *is not yet infected by the beneficial mutant. Any new origin of the beneficial allele will give a new color* $u \in [0,1]$*, so, e.g.* $Y_i(t) = Y_j(t) = u \neq \circ$

means that by time t both demes i and j carry the beneficial allele and both carry the same mutation. The process starts in $\underline{Y}(0) = \underline{\circ}$ and has the following dynamics:
i) As soon as for some k the process Y_k reaches the value $u \neq \circ$, then after a fixed additional time of length 1 all the neighboring demes Y_j of k for which $Y_j = \circ$ by that time are set to u.
ii) Each deme i with $Y_i(t) = \circ$ has an instantaneous rate of $\lambda \in (0, 1]$ to change to some uniformly distributed $u \in [0, 1]$.

For simplicitiy we consider the demes to be arranged cyclically where each deme is connected to its two neighbors by migration. This is shown in Fig. 2.1.

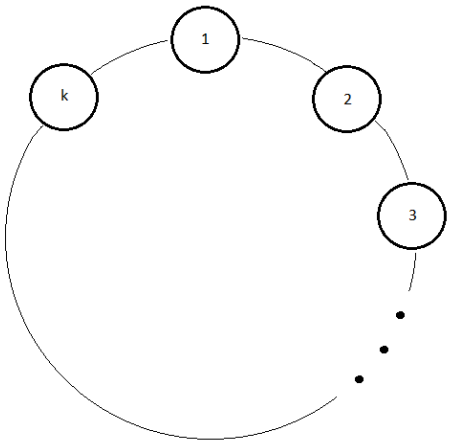

Figure 2.1: *Circular stepping stone model with k demes. The deme where the first mutation arises, is defined as deme 1.*

On the genetic level this corresponds to an infinite alleles model, so all mutations are distinguishable by their nucleotide sequence. But on the phenotypic level they show the same phenotype, or at least a phenotype of similar functional effect, so that they all have the same fitness. Note that in contrast to our model, Greven *et al.* (2014) do not assume distinguishable mutations. In fact, they do not assume a mutation rate at all but consider one single mutation that is already given at the beginning and then spreads throughout the structured population.
Moreover, note that no other mutation can infect a deme anymore once that deme carries a beneficial mutation.

2.2 Expected number of new mutations

As a first aim, we treat in this section the question about how many independent origins the beneficial allele has when it finally gets to fixation in the whole population.

Let T_i be the time of the first appearance of a spontaneous mutation in deme i and M the number of independent origins of the beneficial allele. The random variables T_i are all exponentially distributed with parameter λ.
Throughout the whole chapter we always label the deme where the first mutation appears as deme 1 and the occuring mutation as u_1. Moreover, we can set $T_1 = 0$ due to the memoryless property of the exponential distribution since for any strictly positive mutation rate λ there will always be a first mutation appearing somewhere within a finite waiting time.
Note that we did not define the model for $\lambda = 0$ in Definition 2.1 since, obviously, from a biological point of view there should be no mutation at all if the mutation rate equals zero. However, we will mention the case $\lambda = 0$ at some point in the next chapter as the limit case for very small mutation rates. In that context, we like to point out that all the following equations are likewise defined for $\lambda = 0$.
Before treating the general case with an arbitrary number of demes, we consider the cases $k = 2, 3, 4, 5$ and derive the expected number of independent spontaneous mutations in detail.

2.2.1 The case k=2

In the case of just two demes there are two possibilities. Either deme 2 gets infected by deme 1 after an additional time of length 1, or a spontaneous mutation u_2 occurs in deme 2 before the infection u_1 is transmitted from deme 1.
For the first case, a new mutation in deme 2 must not occur before time 1 and hence, the probability for this event is given by

$$P(T_2 > 1) = e^{-\lambda}. \tag{2.1}$$

For the second case, which is exactly the counterevent of the first, it thus yields

$$P(T_2 < 1) = 1 - e^{-\lambda}. \tag{2.2}$$

Therefore, the expectation for the number of independent spontaneous mutations is given by

$$\begin{aligned} E(M) &= 1 \cdot P(T_2 > 1) + 2 \cdot P(T_2 < 1) \\ &= 2 - e^{-\lambda}. \end{aligned} \tag{2.3}$$

2.2.2 The case k=3

For $k = 3$ we can distinguish three cases.

1) Only one mutation u_1 occurs and infects the other demes. The probability for this event is given by

$$P(T_2, T_3 > 1) = \left(e^{-\lambda}\right)^2. \tag{2.4}$$

Here we used independence of the events $T_2 > 1$ and $T_3 > 1$ and the assumption that the distribution ot mutation times is the same in all demes.

2) Two mutations occur which means that either in deme 2 or in deme 3 a new mutation can occur but not in both. By symmetry, both events have the same probability and hence, it suffices to look at the case where deme 3 is infected by one of the other demes and a new mutation can occur in deme 2. As by definition it always holds that $T_1 < T_2$, deme 3 gets infected by deme 1 before u_2 could be transmitted. Therefore, the probability for exactly two mutations is given by

$$\begin{aligned} 2 \cdot P(T_2 < 1, T_3 > \underbrace{\min(T_1 + 1, T_2 + 1)}_{=1}) &= 2 \cdot P\left(T_2 < 1\right) \cdot P(T_3 > 1) \\ &= 2 \cdot \left(1 - e^{-\lambda}\right) e^{-\lambda}, \end{aligned} \tag{2.5}$$

where the factor 2 occurs because of symmetry.

3) Three spontaneous mutations occur independently, so mutations in demes 2 and 3 have to occur before the mutation u_1 can be transmitted. This immediately implies that deme 2 and 3 cannot infect each other since both mutations have to arise before time 1. It follows for the probability of this event

$$P(T_2, T_3 < 1) = \left(1 - e^{-\lambda}\right)^2. \tag{2.6}$$

As a consequence we obtain for the expected number of independent spontaneous mutations

$$\begin{aligned} E(M) &= 1 \cdot \left(e^{-\lambda}\right)^2 + 2 \cdot 2 \cdot \left(1 - e^{-\lambda}\right) e^{-\lambda} + 3 \cdot \left(1 - e^{-\lambda}\right)^2 \\ &= 3 - 2 \cdot e^{-\lambda}. \end{aligned} \tag{2.7}$$

2.2.3 The case k=4

We treat the case $k = 4$ in a similar way and obtain four different cases.

1) One mutation: The beneficial allele of deme 1 can spread to all other demes before another mutation can arise in one of them. This means that mutations in demes 2 and 4 must not appear before time 1 whereas in deme 3 no mutation occurs before time 2. The corresponding probability is given by

$$P(T_2, T_4 > 1; T_3 > 2) = P(T_2, T_4 > 1) P(T_3 > 2), \tag{2.8}$$

where all the events are independent of each other.

2) Two mutations: Either the second mutation happens in deme 2 or 4, which by symmetry has the same probability, or deme 1 transmits its beneficial allele to deme 2

and 4 and the second mutation arises in deme 3. Hence, we obtain for the probability of exactly two mutations

$$\begin{aligned} &2 \cdot P(T_4 > 1, T_2 < 1; T_3 > T_2 + 1) + P(T_2, T_4 > 1; T_3 < 2) \\ &= 2 \cdot P(T_4 > 1) P(T_2 < 1; T_3 > T_2 + 1) + P(T_2, T_4 > 1) P(T_3 < 2), \end{aligned} \tag{2.9}$$

where the factor 2 appears by symmetry and some of the events are independent.

3) Three mutations: Only one of the three remaining demes gets infected by one of its neighbors while new mutations can successfully arise in the other demes before they get infected. This is given by

$$\begin{aligned} &2 \cdot P(T_4 > 1; T_2 < 1; T_3 < T_2 + 1) + P(T_2, T_4 < 1; T_3 > \min(T_2, T_4) + 1) \\ &= 2 \cdot P(T_4 > 1) P(T_2 < 1; T_3 < T_2 + 1) + P(T_2, T_4 < 1; T_3 > \min(T_2, T_4) + 1), \end{aligned} \tag{2.10}$$

where again the factor 2 occurs because of symmetry. Note that here we need a distinction of cases for the explicit calculation of $P(T_2, T_4 < 1; T_3 > \min(T_2, T_4) + 1)$. This calculation can be found in Appendix A.

4) Four mutations: No deme gets infected by an adjacent deme. As always $T_3 > T_1$, deme 2 and 4 could only get infected by deme 1 and hence, a new mutation has to arise before time 1 in both demes. The admissible time interval for the mutation of deme 3 depends on which of its neighbors has the first mutation, so which of its neighbors would be the first to infect deme 3. Thus, we obtain

$$P(T_2, T_4 < 1; T_3 < \min(T_2, T_4) + 1). \tag{2.11}$$

This leads again to the distinction of two different cases, namely $T_2 < T_4$ and $T_2 > T_4$. The corresponding calculation is shown in Appendix A.

Finally, the expectation of M is given by

$$E(M) = 4 - \frac{8}{3} e^{-\lambda} - \frac{1}{3} e^{-4\lambda}. \tag{2.12}$$

2.2.4 The case k=5

Before considering the general case, we still treat the case of five demes where we can have up to five separate origins of the beneficial allele.

1) One mutation: The only mutation is u_1 and it is transmitted to all other demes before a spontaneous mutation can arise in one of them. Thus, no mutation can occur in demes 2 and 5 before time 1 while in demes 3 and 4 there must not be a new mutation before time 2. This leads to

$$P(T_2 > 1)^2 \cdot P(T_3 > 2)^2. \tag{2.13}$$

Note that again we used the fact that the distribution of mutation time does not depend on the demes.

2) Two mutations: A second mutation occurs either at a distance of one from deme 1 or at a distance of two. The first case corresponds to the arising of a new mutation in deme 2 or 5, the latter to a mutation in 3 or 4. If a new mutation occurs in deme 2, so before time 1, then u_1 is still the first to reach demes 4 and 5 but deme 3 necessarily gets infected by deme 2. If it is deme 3 where the second mutation occurs, then the mutation u_1 spreads to the demes 2 and 5 before u_3, but which one of the two mutations reaches deme 4 first depends on the time when the mutation in deme 3 takes place. By symmetry, the equivalent holds for a mutation in deme 4 or 5 and therefore, it yields

$$\begin{aligned} &2P(T_2 < 1; T_3 > T_2 + 1)P(T_4 > 2)P(T_5 > 1) \\ &+ 2P(T_2 > 1)^2 P(T_3 < 2; T_4 > \min(2, T_3 + 1)) \end{aligned} \tag{2.14}$$

3) Three mutations: There are four possible patterns for the case of three mutations.
The second and the third mutation can arise in deme 2 and 5 which immediately implies that deme 3 gets infected by deme 2 and deme 4 by deme 5.
On the other hand, demes 2 and 5 can get infected by deme 1, hence the time interval for a mutation in deme 3 depends on T_4 and that of deme 4 on T_3.
Moreover, the two mutations can also occur in deme 2 and 3. In that case T_2 takes place before time 1, deme 5 gets infected by deme 1 and since u_1 cannot reach deme 3 before time 3, T_3 only depends on T_2. It then depends on T_3 by which mutation deme 4 gets infected, either by u_1 or u_3. Of course, the symmetrical case of new mutations in demes 4 and 5 has the same probability.
As a last case, we can have new mutations in demes 2 and 4. This directly implies that deme 1 transmits its beneficial allele to deme 5 and deme 3 is either infected by deme 2 or 4, depending on T_2 and T_4. Again, the equivalent holds for the symmetrical case of new mutations in demes 3 and 5.
This can be summarized as

$$\begin{aligned} &P(T_2 < 1; T_3 > T_2 + 1)P(T_5 < 1; T_4 > T_5 + 1) \\ &+ 2P(T_2 < 1; T_3 < T_2 + 1; T_4 > \min(2, T_3 + 1))P(T_5 > 1) \\ &+ 2P(T_2 < 1; T_4 < 2; T_3 > \min(T_2, T_4) + 1)P(T_5 > 1) \\ &+ P(T_2 > 1)^2 P(T_3 < \min(2, T_4 + 1); T_4 < \min(2, T_3 + 1)) \end{aligned} \tag{2.15}$$

4) Four mutations: In the case of four origins of the beneficial allele, there is only one deme which cannot generate its own mutation before being infected by one of its neighbors.
If this deme is at distance one from deme 1, let us say deme 5, then u_1 spreads to deme

5 and u_2 has to occur before time 1. The time intervals for the mutations in deme 3 and 4 depend on the mutation times of their adjacent demes where of course deme 5 is infected by deme 1 exactly at time 1.
If the deme without an own mutation is at distance two from deme 1, let us say deme 4, deme 3 cannot infect demes 2 and 5 before they are reached by u_1 but, depending on T_3 and T_5, it could reach deme 4 before this deme is reached by u_5. On the other hand, the time interval for T_3 only depends on T_2 because it holds that $T_2, T_5 < 1$ and therefore, u_2 reaches deme 3 before u_5.
By symmetry, it yields

$$\begin{aligned} &2P(T_2, T_5 < 1; T_3 < T_2 + 1; T_4 > \min(T_3, T_5) + 1) \\ &+ 2P(T_5 > 1)P(T_2 < 1; T_3 < \min(T_2, T_4) + 1; T_4 < \min(2, T_3 + 1)) \end{aligned} \tag{2.16}$$

5) Five mutations: Finally, the probability that a new mutation can arise in every deme is given by

$$P(T_2, T_5 < 1; T_3 < \min(T_2, T_4) + 1; T_4 < \min(T_3, T_5) + 1). \tag{2.17}$$

The time intervals in which the mutations in demes 3 and 4 have to occur, depend on the mutation times of the adjacent demes. Mutations in demes 2 and 5 have, of course, to arise before time 1.

Note that for the calculation of some of the probabilities from above a distinction of several cases is required. These calculations are shown in detail in Appendix A.

Finally, we obtain the expected number of independent spontaneous mutations which is given by

$$E(M) = 5 - \frac{10}{3}e^{-\lambda} - \frac{3}{2}e^{-3\lambda} + \frac{7}{3}e^{-4\lambda} - \frac{3}{2}e^{-5\lambda}. \tag{2.18}$$

2.2.5 The general case

The general case for an arbitrary k can be treated in a similar way but it quickly turns out that this does not lead to any feasible expression because it requires the distinction of many cases. We briefly outline the problem and then slightly modify our model in the next chapter in order to obtain tractable expressions.
For only one successful mutation, i.e. the beneficial mutation u_1 can be transmitted to all other demes before a new mutation could successfully arise in one of them, we obtain because of symmetry

$$\prod_{j=2}^{\lfloor \frac{k}{2} \rfloor} P\left(T_j > j - 1\right)^2 \cdot P\left(T_{\lfloor \frac{k}{2} \rfloor + 1} > \left\lfloor \frac{k}{2} \right\rfloor\right)^{\alpha} \tag{2.19}$$

with $\alpha = 1$ for k even and $\alpha = 2$ for k odd. Here, we used again that the distribution of the mutation times is independent of the demes. This expression can still be calculated easily.
We now treat the case of $i+1$ mutations (for $1 \leq i \leq k-1$) which leads to far more complex expressions.
In order to obtain the time interval in which a spontaneous mutation can arise in an arbitrary deme $l > 1$ before an infection is transmitted from one of the adjacent demes, we need to know the nearest mutation origins on each side of l. Therefore, we define

$$a_l := \max_{1 \leq j \leq k-1} (\text{Deme } j \text{ is the origin of a beneficial mutation and } j < l\ )$$

and

$$b_l := \min\left(k+1, \min_{2 \leq j \leq k} (\text{Deme } j \text{ is the origin of a beneficial mutation and } j > l\ )\right).$$

By these definitions, we immediately obtain the distance between deme l and the nearest origin demes on each side of l as $|l - a_l|$ and $|l - b_l|$ respectively.
Using these definitions yields the probability that exactly $i+1$ mutations (for $1 \leq i \leq k-1$) occur

$$\begin{aligned} \frac{1}{i!} \sum_{x_1=2}^{k} \sum_{\substack{x_2=2 \\ x_2 \neq x_1}}^{k} \cdots \sum_{\substack{x_i=2 \\ x_i \neq x_{i-1},\ldots,x_1}}^{k} P\Big(\bigcap_{l=1}^{i} T_{x_l} < \min(T_{a_{x_l}} + |x_l - a_{x_l}|, T_{b_{x_l}} + |x_l - b_{x_l}|, \\ \bigcap_{\substack{j=2 \\ j \neq x_1,\ldots,x_i}}^{k} T_j > \min(T_{a_j} + |j - a_j|, T_{b_j} + |j - b_j|\Big), \end{aligned} \tag{2.20}$$

where $i!$ appears because of the number of possible permutations. For an explicit calculation of this expression one needs to distinguish many cases.
Even for the case $i = 1$ this expression gets rather complex and requires a distinction of many cases as we briefly like to point out. By symmetry, (2.20) can be simplified for the case $i = 2$ and it yields

$$\begin{aligned} &2 \cdot \sum_{j=2}^{\lfloor \frac{k}{2} \rfloor} P\Big(T_j < j-1; \bigcap_{\substack{l=2 \\ l \neq j}}^{k} T_l > \min(l-1, T_j + |j-l|) \Big) \\ &+ \alpha \cdot P\left(T_{\lfloor \frac{k}{2} \rfloor + 1} < \left\lfloor \frac{k}{2} \right\rfloor ; \bigcap_{\substack{l=2 \\ l \neq \lfloor \frac{k}{2} \rfloor + 1}}^{k} T_l > \min\left(l-1, T_{\lfloor \frac{k}{2} \rfloor + 1} + \left| \left\lfloor \frac{k}{2} \right\rfloor + 1 - l \right| \right) \right) \end{aligned} \tag{2.21}$$

with α defined as above.
We have a closer look at the first expression of equation (2.21):

$$\begin{aligned}
&P\Big(T_j < j-1; \bigcap_{\substack{l=2 \\ l\neq j}}^{k} T_l > \min(l-1, T_j + |j-l|)\Big) \\
&= \int_0^{j-1} \lambda e^{-\lambda t_j} \int_{c_2}^{\infty} \dots \int_{c_{j-1}}^{\infty} \int_{c_{j+1}}^{\infty} \dots \int_{c_{\lfloor \frac{k}{2} \rfloor}}^{\infty} \lambda^{\lfloor \frac{k}{2} \rfloor - 2} e^{-\lambda(t_2+\dots+t_{j-1}+t_{j+1}+\dots+\lfloor \frac{k}{2} \rfloor)} \mathrm{d}t_{\lfloor \frac{k}{2} \rfloor} \dots \mathrm{d}t_{j+1} \mathrm{d}t_{j-1} \dots \mathrm{d}t_2 \mathrm{d}t_j \\
&= \int_0^{j-1} \lambda e^{-\lambda t_j} \cdot e^{-\lambda(c_2+\dots+c_{j-1}+c_{j+1}+\dots+c_{\lfloor \frac{k}{2} \rfloor})} \mathrm{d}t_j
\end{aligned} \tag{2.22}$$

with $c_l := \min(l-1, t_j + |j-l|)$.

At this point, due to the boundary conditions that follow from the circular arrangement of demes, we need to distinguish all different cases for the c_l and as one can easily imagine, this gets even more complex for larger i.
Hence, this is not feasible in a general way for larger k, so in a next step we will slightly modify our model regarding the arrangement of demes in order to get rid of these boundary conditions. This modified model will be the subject of the next chapter.

3 The one-dimensional model with an infinite chain of demes

In this chapter we extend the one-dimensional model of the previous chapter where we considered only a finite number of demes in a circular arrangement, to an infinite, linear chain of demes. Arranging demes in a linear way allows us to avoid the boundary conditions which we faced in the first chapter.
Similarly as before, we start by deriving the expected frequency of new mutations and compare this to our results of the finite model. Moreover, we compare this result of the infinite model to results of a continuous space model from the literature. We then have a closer look at the distribution of cluster sizes and the correlation of the sizes of adjacent clusters. At the end of this chapter about the one-dimensional infinite model we briefly focus on the distance between the nucleation points of two neighboring clusters and compare our model to a model with independent nucleation origins.

3.1 Expected frequency of new mutations

In this section we derive the probability for an arbitrary deme that a spontaneous mutation successfully arises and gets to fixation. But as the arising of a new mutation is identically distributed for all demes and we consider infinitely many demes, this is not only the probability for a single deme to have a spontaneous mutation but at the same time it is also the expected proportion of demes with a new mutation in the whole one-dimensional space.

Throughout the whole chapter demes are labeled by integers. We fix an arbitrary focal deme which we sometimes may also call "deme 0".
Let n denote the number of demes that a mutation u_0 which has arised from deme 0, can cover before being stopped by other mutations. With

$$n_r := \text{number of covered demes on the right side of deme 0}$$

and

$$n_l := \text{number of covered demes on the left side of deme 0,}$$

the total number of covered demes for deme 0 can be written as

$$n = n_r + n_l + 1. \tag{3.1}$$

We start by calculating the probability that a spontaneous mutation arises and gets to fixation in the focal deme. This probability will be denoted by $P(n > 0)$. Let $T \in \mathbb{R}_{\geq 0}$ be the time of the first successful mutation that occurs in deme 0 and let us keep T fixed for a moment.
The distance of two distinct demes x and y is defined by $d(x, y) = |x - y|$. Obviously, only demes j ($j \neq 0$) with $d(0, j) \leq T$ could infect deme 0 with a mutation u_j until time T and hence,

$$d_T := \lfloor T \rfloor$$

is the maximal distance that an arbitrary mutation can spread in one direction until time T.
Therefore, if $d_T > 0$, the probability that no infection from any other deme reaches the focal deme until a spontaneous mutation u_0 successfully arises at time T is given by

$$P(n > 0|T_0 = T) = P(T_{\pm 1} > T - 1, T_{\pm 2} > T - 2, ..., T_{\pm d_T} > T - d_T). \tag{3.2}$$

Since all these events are independent of each other and the distribution of the mutation times T_i does not depend on the demes, (3.2) simplifies to

$$\begin{aligned} P(n > 0|T_0 = T) &= \prod_{j=1}^{d_T} P\left(T_{\pm j} > T - j\right) \\ &= \prod_{j=1}^{d_T} P\left(T_0 > T - j\right)^2 \\ &= e^{-2\lambda \sum_{j=1}^{d_T} (T-j)} \\ &= e^{-2\lambda T d_T + \lambda d_T (d_T+1)}. \end{aligned} \tag{3.3}$$

For the special case of $d_T = 0$, no other deme is able to infect deme 0 before time T and hence, $P(n > 0|T_0 = T) = 1$.
Using this expression, integration over T yields the probability for an arbitrary deme to get infected by a spontaneous mutation instead of being infected by another deme. It is given by

$$\begin{aligned} P(n > 0) &= \int_0^\infty \lambda e^{-\lambda T} \cdot P(n > 0|T_i = T) \mathrm{d}T \\ &= \int_0^1 \lambda e^{-\lambda T} \mathrm{d}T + \int_1^\infty \lambda e^{-\lambda T} \cdot e^{-2\lambda T d_T + \lambda d_T (d_T+1)} \mathrm{d}T \\ &= 1 - e^{-\lambda} + \sum_{j=1}^{\infty} \int_j^{j+1} \lambda e^{-\lambda T} \cdot e^{-2\lambda T j + \lambda j (j+1)} \mathrm{d}T. \end{aligned} \tag{3.4}$$

We obtain for the integral over this step function with $j \geq 1$

$$\begin{aligned}\int_{j}^{j+1} \lambda e^{-\lambda T} \cdot e^{-2\lambda j T + \lambda j(j+1)} \mathrm{d}T &= \frac{1}{2j+1} e^{\lambda j(j+1)} \left(e^{-\lambda(2j+1)j} - e^{-\lambda(2j+1)(j+1)} \right) \\ &= \frac{1}{2j+1} \left(e^{-\lambda j^2} - e^{-\lambda(j+1)^2} \right).\end{aligned} \tag{3.5}$$

Hence, (3.4) becomes

$$P(n > 0) = \sum_{j=0}^{\infty} \frac{1}{2j+1} \left(e^{-\lambda j^2} - e^{-\lambda(j+1)^2} \right). \tag{3.6}$$

We can compare this result for the expected frequency of new mutations with the previous results that we obtained for two, three, four and five demes in Chapter 1. This is shown in Fig. 3.1 for different values of λ.

Remark 3.1. *Throughout the whole thesis we choose the approximation of all infinite sums, including step integrals, in a way that the graphical visualisation does hardly differ from the true values.*

As a first general observation we see that for both models the expected frequency of new mutations increases with increasing mutation rate which we have of course expected. Moreover, for all values of λ the expected proportion of new mutations in the finite case approaches the expected frequency in the infinite case asymptotically. This convergence becomes faster with increasing mutation rate and for large values of λ the cases $k = 4, 5$ and even $k = 3$ are already very good approximations for the infinite model.
In order to explain this, let us first have a closer look at the special case $\lambda = 0$. The result shown in the figure reflects the different assumptions in the finite and the infinite model. In the finite model we assume that a mutation occurs in the origin deme 1 at time $T_1 = 0$, so we always start in our model with a mutation regardless of the value of the mutation rate. For all $\lambda > 0$, this can be justified by the Memoryless property because there will always be a first mutation appearing somewhere within a finite waiting time, even if this time can be very large for small λ. The case $\lambda = 0$ can be considered as the limit case for very small mutation rates. Hence, there is exactly one mutation for $\lambda = 0$. In contrast, in the infinite model we do not start with a mutation in any deme and new mutations can only arise if the mutation rate is strictly positive, so there will be no mutation at all for $\lambda = 0$. Moreover, the probability for a new mutation in any finite subset of demes always tends to zero for very low λ whereas in the finite model there must be at least one origin.
These different assumptions are also the main reason why for smaller λ the difference between finite and infinite model is larger, especially for only a small number of demes. For larger mutation rates, which implies that more new mutations can occur, the very

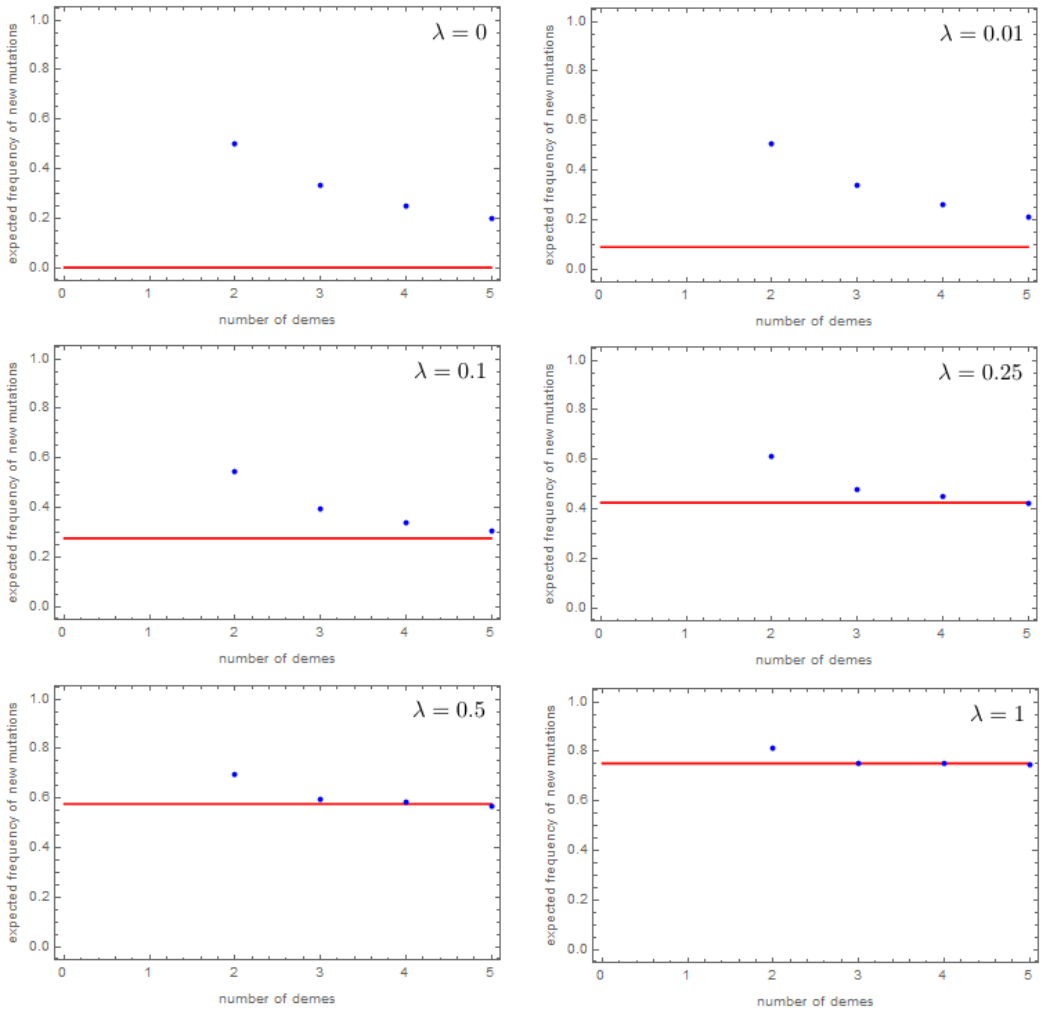

Figure 3.1: *The expected relative frequencies of demes with a new mutation for the finite (blue dots) and the infinite deme model (red line) are shown for different values of* λ.

first mutation in the finite model becomes less relevant and has less influence. This argument also contributes to the observation that the approximation is asymptotic since the influence of this main difference in model assumptions decreases with increasing number of demes. Moreover, it explains why we see an approximation from above, so why for a small number of demes the expected frequency of new mutations is larger in the finite deme model compared to the infinite deme model.
Interestingly, for larger mutation rates the expected frequency of new mutations is slightly smaller in the finite model for $k = 5$ than in the infinite model. This can be explained by the fact that we assume a circle of demes in the finite model whereas in the infinite model we assume a linear chain of demes and, of course, the mean time it takes a randomly positioned mutation to cover a finite number of demes arranged in

a circle is shorter than for the same number of demes arranged as a chain. Thus, the mean time that is left for new mutations to arise is smaller for a circular arrangement and therefore, less mutation events are expected to take place. This effect becomes smaller with an increasing number of demes. So, if we derived results for more demes of a finite circle, we would expect that the mutation frequency first drops but later increases again to the red line.

3.2 Comparison with a continuous space model

In this section we compare the expected frequency of new mutations from our discrete space model, which we obtained in (3.6), with the expected value given by a continuous space model. We use for our discussion some results from Evans (1945).
Evans considers raindrops which randomly fall on a two-dimensional area and produce expanding circles. In a first step he assumes an unbounded growth of these circles, in particular this means that circles do not disturb each others propagation.
Then, for a random distribution of the raindrops, a constant radial velocity v for the growth of the circles and a rain at uniform rate, the number of circles which pass over an arbitrary point P within a certain time from the start of the rain, is Poisson distributed. Consequently, the probability that a certain point is not passed by any circle within that time is given by

$$p_0 = e^{-E}, \tag{3.7}$$

where E is the mean number of passing circles.
In order to compute E, Evans states that only raindrops fallen within a distance vt from point P can reach this point within time t. Hence, it follows

$$E = \int_0^{vt} \Omega(t - \frac{r}{v}) 2\pi r \mathrm{d}r = \frac{\pi \Omega v^2 t^3}{3}, \tag{3.8}$$

where r is the distance from point P and Ω the two-dimensional nucleation rate. This is a time-dependent Poisson process.
Expanding his model by introducing the additional assumption that raindrops can only fall on areas which are not yet covered by other circles, Evans argues that in general the number of circles passing P is no longer Poisson distributed but for the special case of P remaining uncovered until a certain time t, the expression $p_0 = e^{-E}$ is still valid because only the first arriving circle matters and this circle is not affected by the constraint that new drops are only allowed to fall on uncovered area.

We can now use these results to derive the probability that at some point a raindrop falls on an arbitrary point P which has to remain uncovered until that event. As we would like to compare this to the results from our one-dimensional mutation model, we

slightly modify Evans' results of his two-dimensional model, more precisely we consider the one-dimensional area $2r$ instead of $2\pi r$ and we use our one-dimensional mutation rate λ. The spreading velocity v can be set to 1, corresponding to the deterministic time it takes a mutation to spread from one deme to the next.
With these modifications, Evans model is exactly the corresponding continuous space model to our one-dimensional discrete space model that we have treated so far and hence, the probability that a raindrop falls on a still uncovered area is equivalent to the probability that a mutation successfully arises in a still uninfected deme.
By introducing the notation of our model, this probability then becomes

$$\begin{aligned} P_{cont}(n > 0) &= \int_0^\infty \lambda e^{-\lambda t} p_0 \mathrm{d}t \\ &= \int_0^\infty \lambda e^{-\lambda t} e^{-\frac{\lambda t^3}{3}} \mathrm{d}t. \end{aligned} \tag{3.9}$$

In Fig. 3.2 we compare this result with the result of (3.6) of our discrete space model.

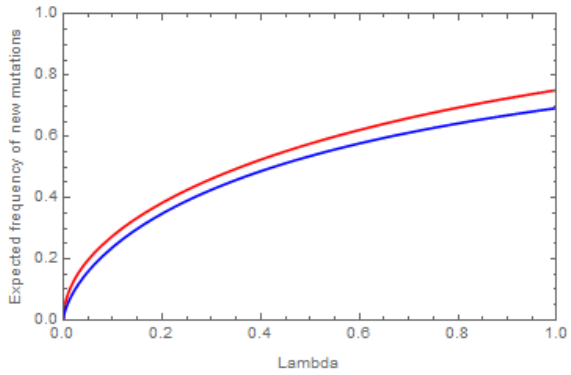

Figure 3.2: *The expected frequency of new mutations in the discrete space and the continuous space model is shown. The red line represents the discrete case, the blue one the continuous space model.*

Obviously, the expected proportion of demes with a new mutation is larger in the discrete space model than in the continuous space model and the larger the mutation rate, the larger is the absolute difference. This perfectly reflects the fact that in the discrete space model a mutation does not at all affect the adjacent demes until time 1 while in the continuous space model mutations spread in a wave-like manner which means that they continuously cover the area around them. So, a mutation travels the same distance within time 1 in both models but in one case it is a continuous spreading while in the other case the area remains uncovered until 1 and then gets completely covered from one moment to the next. Hence, in the discrete case there is more uncovered area at any time before 1 and therefore, more mutations are expected to take place compared to the continuous model.

3.3 Distribution of cluster sizes

Next, we are interested in the spreading of mutations, more precisely in the number of demes that are infected by one single mutation, and in the distribution of these cluster sizes.
Again, we start by considering an arbitrary deme and label it as deme 0. We already defined n_r and n_l as the number of demes located to the right and to the left side of the focal deme that are infected by mutation u_0.
The distribution of cluster sizes in general is given by

$$\frac{P(n=k)}{P(n>0)} = \frac{\int_0^\infty \lambda e^{-\lambda T} \cdot P(n=k|T_0=T)\mathrm{d}T}{P(n>0)} \qquad \text{for } k \geq 1 \text{ and } \lambda \neq 0. \tag{3.10}$$

In order to ensure that the denominator $P(n > 0)$ does not equal zero, we have to exclude the case $\lambda = 0$. But as we already mentioned before, this is just the trivial limit case where the mutation rate equals zero and therefore, no mutations occur at all.

We now derive the probability for any cluster to have size k for an arbitrary $k \geq 1$. The probability for a mutation to arise in deme 0 and to form a cluster of size $k \geq 1$ given the condition $T_0 = T$ can be written as

$$P(n=k|T_0=T) = \sum_{j=0}^{k-1} P(n_r=j, n_l=k-1-j|T_0=T). \tag{3.11}$$

We start by deriving the joint probability $P(n_r = k_1, n_l = k_2|T_0 = T)$ for arbitrary integers $k_1, k_2 \geq 0$.
First of all note that a cluster of size k with $n_r = k_1$ and $n_l = k_2$ is described by the intersection of the events given by the following definition.

Definition 3.2. *For an arbitrary deme 0 and integers $k_1, k_2 \geq 0$ we define:*

$N_r^1(k_1) :=$ *u_0 can reach all demes on the right side of deme 0 until deme k_1 before a spontaneous mutation arises in one of these demes,*
$N_l^1(k_2) :=$ *u_0 can reach all demes on the left side of deme 0 until deme $-k_2$ before a spontaneous mutation arises in one of these demes,*
$B_r^1(k_1) :=$ *a mutation coming from the right side of deme k_1 reaches deme k_1+1 before time T_0+k_1+1,*
$B_l^1(k_2) :=$ *a mutation coming from the left side of deme $-k_2$ reaches deme $-(k_2+1)$ before time T_0+k_2+1,*

$C_r^1(k_1) :=$ u_0 *can reach deme* k_1 *before another mutation coming from the right side of deme* k_1,
$C_l^1(k_2) :=$ u_0 *can reach deme* $-k_2$ *before another mutation coming from the left side of deme* $-k_2$.

Note that one of these events alone cannot guarantee that deme 0 is not yet infected by another mutation at time T_0 but the events $N_r^1(k_1), N_l^1(k_2), C_r^1(k_1)$ and $C_l^1(k_2)$ together do ensure it.
As we will need them in the following, we now calculate the probabilities of these events, each of them under the condition $T_0 = T$.
Let us start with the events $N_r^1(k_1)$ and $N_l^1(k_2)$. First, we set $P(N_r^1(k_1)|T_0 = T) = 1$ and $P(N_l^1(k_2)|T_0 = T) = 1$ for $k_1 = 0$ and $k_2 = 0$ respectively. Note that this definition is perfectly consistent with the fact that a right cluster size of zero is sufficiently charaterized by $B_r^1(k_1)$ and $C_r^1(k_1)$ and the equivalent holds for k_2 as well.
For $k_1 > 0$ and $k_2 > 0$ we obtain

$$\begin{aligned} P\left(N_r^1(k_1)|T_0 = T\right) &= P\left(T_1 > T+1, ..., T_{k_1} > T+k_1\right) \\ &= \prod_{j=1}^{k_1} P\left(T_j > T+j\right) \\ &= \prod_{j=1}^{k_1} P\left(T_0 > T+j\right) \\ &= \prod_{j=1}^{k_1} e^{-\lambda(T+j)} \\ &= e^{-\lambda T k_1} e^{-\lambda \frac{k_1(k_1+1)}{2}}, \end{aligned} \tag{3.12}$$

where we again used independence of the mutation events above and that the distribution of mutation times is independent of the demes.
Similarly, we get

$$\begin{aligned} P\left(N_l^1(k_2)|T_0 = T\right) &= P\left(T_{-1} > T+1, ..., T_{-k_2} > T+k_2\right) \\ &= \prod_{j=1}^{k_2} P\left(T_{-j} > T+j\right) \\ &= \prod_{j=1}^{k_2} P\left(T_0 > T+j\right) \\ &= P\left(N_r^1(k_2)|T_0 = T\right). \end{aligned} \tag{3.13}$$

For the other events we can consider arbitrary $k_1, k_2 \geq 0$ and by using the same argu-

ments as above, we obtain for $B_r^1(k_1)$and $B_l^1(k_2)$

$$\begin{aligned}
P\left(B_r^1(k_1)|T_0=T\right) &= P\left(T+k_1+1>\min\left(T_{k_1+1},...,T_{d_T+2(k_1+1)}+d_T+k_1+1\right)\right)\\
&= 1-P\left(T+k_1+1<\min\left(T_{k_1+1},...,T_{d_T+2(k_1+1)}+d_T+k_1+1\right)\right)\\
&= 1-\prod_{j=1}^{d_T+k_1+2} P\left(T_{k_1+j}>T+k_1+2-j\right)\\
&= 1-\prod_{j=1}^{d_T+k_1+2} P\left(T_0>T+k_1+2-j\right)\\
&= 1-\prod_{j=1}^{d_T+k_1+2} e^{-\lambda(T+k_1+2-j)}\\
&= 1-e^{-\lambda T(d_T+k_1+2)}e^{-\lambda(k_1+2)(d_T+k_1+2)}e^{\lambda\frac{(d_T+k_1+2)(d_T+k_1+3)}{2}}
\end{aligned} \tag{3.14}$$

and by symmetry, it obviously holds

$$P\left(B_l^1(k_2)|T_0=T\right)=P\left(B_r^1(k_2)|T_0=T\right). \tag{3.15}$$

In the same way we calculate the two remaining events $C_r^1(k_1)$ and $C_l^1(k_2)$ and obtain

$$\begin{aligned}
P\left(C_r^1(k_1)|T_0=T\right) &= P\left(T+k_1<\min\left(T_{k_1+1}+1,...,T_{d_T+2k_1}+d_T+k_1\right)\right)\\
&= \prod_{j=1}^{d_T+k_1} P\left(T_{k_1+j}>T+k_1-j\right)\\
&= \prod_{j=1}^{d_T+k_1} P\left(T_0>T+k_1-j\right)\\
&= \prod_{j=1}^{d_T+k_1} e^{-\lambda(T+k_1-j)}\\
&= e^{-\lambda T(d_T+k_1)}e^{-\lambda k_1(d_T+k_1)}e^{\lambda\frac{(d_T+k_1)(d_T+k_1+1)}{2}}
\end{aligned} \tag{3.16}$$

and

$$P\left(C_l^1(k_2)|T_0=T\right)=P\left(C_r^1(k_2)|T_0=T\right). \tag{3.17}$$

The joint probability for the right cluster size to be k_1 and the left cluster size to be k_2 respectively conditioned on $T_0=T$ is then given by

$$\begin{aligned}
&P(n_r=k_1,n_l=k_2|T_0=T)\\
&=P(N_r^1(k_1),N_l^1(k_2),B_r^1(k_1),B_l^1(k_2),C_r^1(k_1),C_l^1(k_2)|T_0=T)\\
&=P(N_r^1(k_1)|T_0=T)\cdot P(N_l^1(k_2)|T_0=T)\cdot P(B_r^1(k_1),C_r^1(k_1)|T_0=T)\\
&\quad\cdot P(B_l^1(k_2),C_l^1(k_2)|T_0=T),
\end{aligned} \tag{3.18}$$

where in the last step we used that some of these events are independent of each other. Obviously, the events $B_r^1(k_1)$ and $C_r^1(k_1)$ are not independent of each other but the probability $P(B_r^1(k_1), C_r^1(k_1)|T_0 = T)$ can be further simplified into

$$\begin{aligned}
&P(B_r^1(k_1), C_r^1(k_1)|T_0 = T) \\
&= P(C_r^1(k_1)|T_0 = T) - P(C_r^1(k_1), B_r^{1,C}(k_1)|T_0 = T) \\
&= P(C_r^1(k_1)|T_0 = T) - P(C_r^1(k_1)|T_0 = T, B_r^{1,C}(k_1))P(B_r^{1,C}(k_1)|T_0 = T) \\
&= P(C_r^1(k_1)|T_0 = T) + P(B_r^1(k_1)|T_0 = T) - 1 \\
&= \prod_{j=1}^{d_T+k_1} e^{-\lambda(T+k_1-j)} - \prod_{j=1}^{d_T+k_1+2} e^{-\lambda(T+k_1+2-j)} \\
&= \prod_{j=1}^{d_T+k_1} e^{-\lambda(T+k_1-j)} \left(1 - e^{-2\lambda(d_T+k_1+2)} e^{-\lambda(T+1-d_T)} e^{-\lambda(T-d_T)}\right) \\
&= e^{-\lambda T(d_T+k_1)} e^{-\lambda k_1(d_T+k_1)} e^{\lambda \frac{(d_T+k_1)(d_T+k_1+1)}{2}} \left(1 - e^{-2\lambda T} e^{-\lambda(2k_1+5)}\right),
\end{aligned} \tag{3.19}$$

where $B_r^{1,C}(k_1)$ denotes the complementary event of $B_r^1(k_1)$. In the second equation we used the fact that $P(C_r^1(k_1)|T_0 = T, B_r^{1,C}(k_1)) = 1$ because if deme $k_1 + 1$ has not been reached by a mutation coming from the right side until time $T_0 + k_1$, then u_0 can reach deme $k_1 + 1$, and therefore also deme k_1, before another mutation coming from the right side of deme k_1.
By symmetry, it follows for the left side of the focal deme

$$\begin{aligned}
&P(B_l^1(k_2), C_l^1(k_2)|T_0 = T) \\
&= P(C_l^1(k_2)|T_0 = T) + P(B_l^1(k_2)|T_0 = T) - 1 \\
&= e^{-\lambda T(d_T+k_2)} e^{-\lambda k_2(d_T+k_2)} e^{\lambda \frac{(d_T+k_2)(d_T+k_2+1)}{2}} \left(1 - e^{-2\lambda T} e^{-\lambda(2k_2+5)}\right).
\end{aligned} \tag{3.20}$$

We are now able to derive $P(n_r = k_1, n_l = k_2)$ explicitely for arbitrary integers $k_1, k_2 \geq 0$ and obtain

$$\begin{aligned}
&P(n_r = k_1, n_l = k_2) \\
&= \int_0^\infty \lambda e^{-\lambda T} P(n_r = k_1, n_l = k_2|T_0 = T) \mathrm{d}T \\
&= \int_0^\infty \lambda e^{-\lambda T} P(N_r^1(k_1)|T_0 = T) \Big[P(C_r^1(k_1)|T_0 = T) + P(B_r^1(k_1)|T_0 = T) - 1\Big] \\
&\quad \cdot P(N_l^1(k_2)|T_0 = T) \Big[P(C_l^1(k_2)|T_0 = T) + P(B_l^1(k_2)|T_0 = T) - 1\Big] \mathrm{d}T.
\end{aligned} \tag{3.21}$$

Although the calculation is straightforward, it becomes rather complex and there is no need to go into details at this point. All further details can be found in Appendix B.

The distribution of the cluster sizes is shown in Fig. 3.3 for different mutation rates. Obviously, the larger the mutation rate, the earlier the propagation of beneficial alleles is stopped by other mutations, i.e. the smaller the chance of forming large clusters. Therefore, for large λ, the distribution puts almost all weight on small clusters, for $\lambda = 1$ even more than 70% of the clusters are only of size one, i.e. beneficial alleles do not even succeed in spreading to adjacent demes.
On the other hand, the smaller λ, the higher is the chance that also large clusters can be generated and thus, the distribution becomes broader.

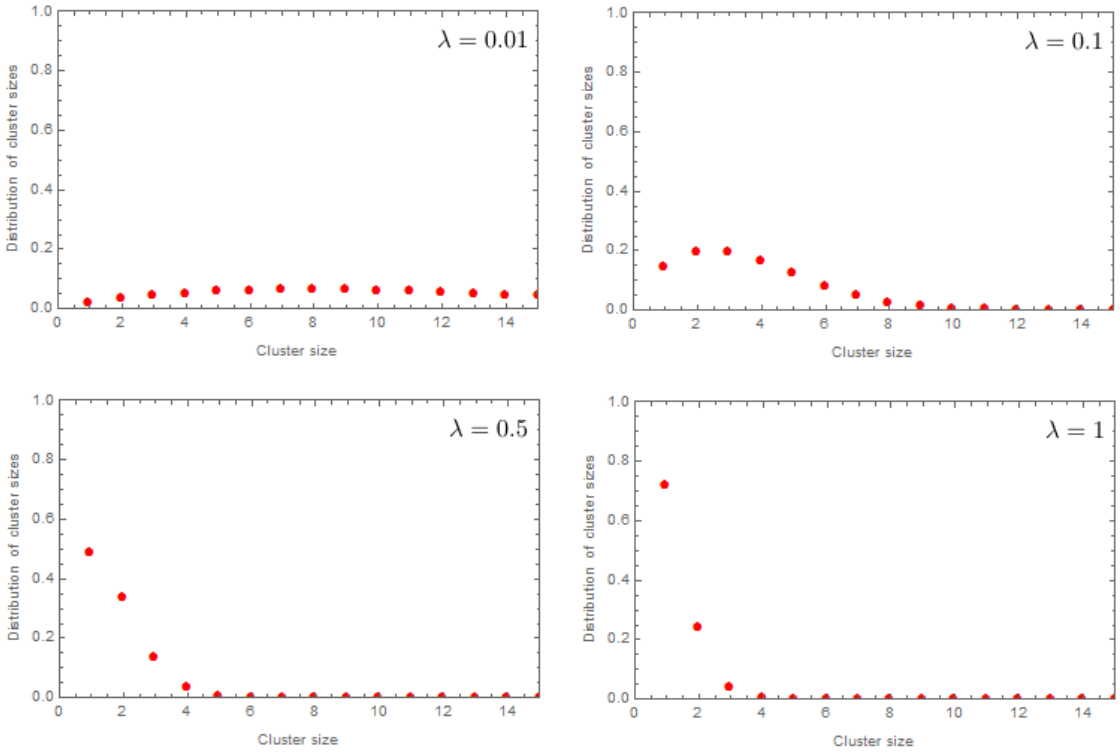

Figure 3.3: *The distribution of cluster sizes is shown for different values of* λ.

The expected cluster size and the corresponding variance, shown in Fig. 3.4, reflect the pattern of the above distribution. Both are small for large λ but become arbitrarily large for arbitrary small mutations rates since mutations happen so slowly that also larger clusters can be formed with an increasing chance. In particular, the variance is high for low λ which goes very well together with the broad distribution of cluster sizes for small mutation rates.

Remark 3.3. *These results are likewise reflected by the number of steps which are needed for good approximations of the step integrals. There are far more steps needed for low* λ *since mutations can spread more and also for increasing time, demes are still likely to be uninfected by then and to become the origin of a new beneficial allele. For a high*

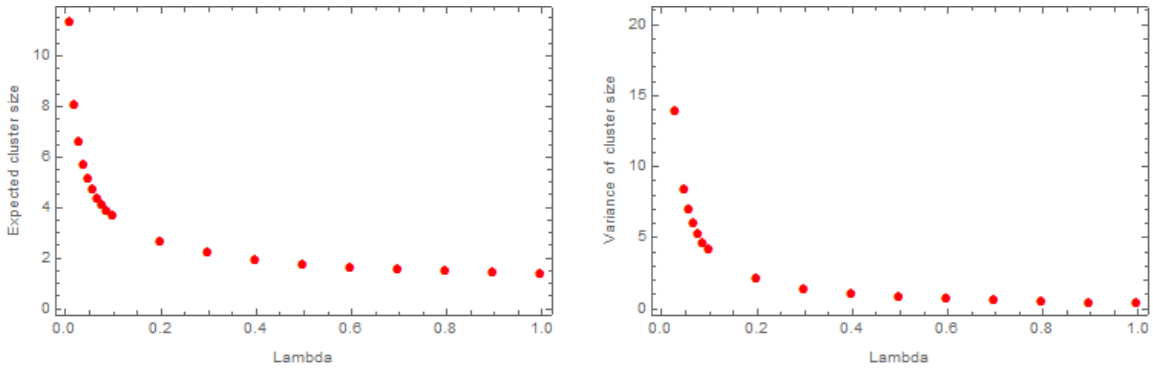

Figure 3.4: *The expectation value and the variance of the cluster size are shown.*

mutation rate, the spreading of mutations is stopped earlier and demes are unlikely to stay uninfected for a long time.

3.4 Distribution of right and left cluster sizes

From the results above we can easily derive the distribution of the right and left cluster size of an arbitrary cluster X which we define as

$$X_r := n_r^x \text{ given } n^x > 0 \tag{3.22}$$

and

$$X_l := n_l^x \text{ given } n^x > 0, \tag{3.23}$$

where n^x, n_r^x and n_l^x are defined as n, n_r and n_l for cluster X.
Given the condition $T_0 = T$, the random variables X_r and X_l are independent and identically distributed. Without this condition, right and left cluster sizes are still identically distributed but not independent since cluster sizes – and as a consequence also right and left cluster sizes – strongly depend on the time when the origin mutation arises – the earlier a mutation arises, the higher is the chance to produce large clusters.

Let us derive the probability $P(n_r = k)$. For an arbitrary cluster with $n_r = k$ and n_l of arbitrary size, it is sufficient to consider $C^1_{l,0}$ as the only condition for the left side of the cluster in order to make sure that no mutation coming from the left side can infect the origin deme of our cluster before time T. Hence, it yields for an arbitrary integer $k \geq 0$

$$\begin{aligned} P(n_r = k) &= \int_0^\infty \lambda e^{-\lambda T} P(n_r = k | T_0 = T) \mathrm{d}T \\ &= \int_0^\infty \lambda e^{-\lambda T} P(C_l^1(0), B_r^1(k), C_r^1(k), N_r(k)^1 | T_0 = T) \mathrm{d}T \end{aligned}$$

$$= \int_0^\infty \lambda e^{-\lambda T} P(C_l^1(0)|T_0 = T) P(N_r^1(k)|T_0 = T)$$
$$\cdot \Big[P(C_r^1(k)|T_0 = T) + P(B_r^1(k)|T_0 = T) - 1 \Big] \mathrm{d}T, \tag{3.24}$$

where we used (3.19) in the third equation. Again, we do not show all details here and leave the rest of the calculation for Appendix B.
The distribution of the random variable X_r is given by $\frac{P(n_r=k)}{P(n>0)}$ which is shown in Fig. 3.5.
Of course, the distribution of the right cluster size strongly ressembles that of the size of the whole cluster. But, as the left cluster size is not taken into account, the distribution of right cluster sizes puts even more weight on small values compared to the distribution of the size of the whole cluster.

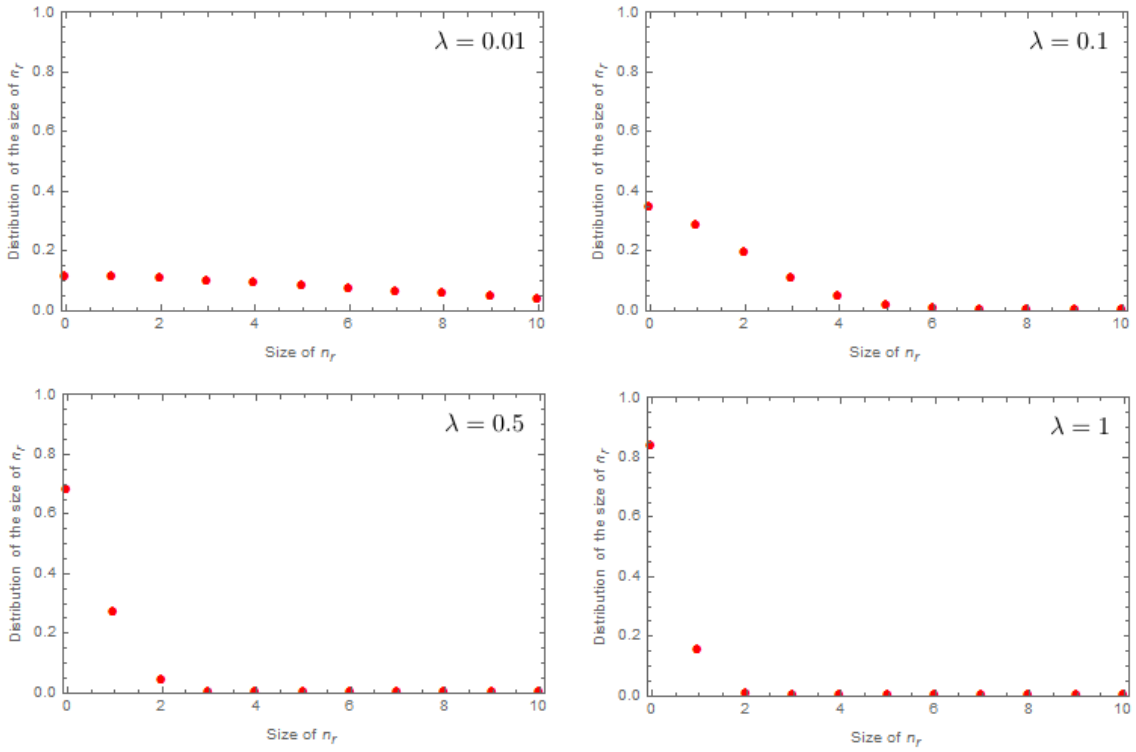

Figure 3.5: *The distribution of X_r is shown for different values of λ.*

3.5 Correlation of right and left cluster sizes

We are now interested in the correlation of the random variables X_r and X_l which means the correlation between right and left cluster sizes. As we already mentioned,

they are only independent of each other if conditioned on a fixed mutation time $T_0 = T$. Consequently, they are also uncorrelated given the condition $T_0 = T$ but this turns out to be wrong for general T.

We already have all the ingredients to calculate the covariance of X_r and X_l. Using that X_r and X_l are identically distributed and that therefore, it holds $\mathrm{E}(X_r) = \mathrm{E}(X_l)$, we obtain

$$\begin{aligned}\mathrm{Cov}(X_r, X_l) &= \mathrm{E}(X_r X_l) - \mathrm{E}(X_r)\mathrm{E}(X_l) \\ &= \mathrm{E}(X_r X_l) - \mathrm{E}(X_r)^2 \\ &= \sum_{k_1=1}^{\infty}\sum_{k_2=1}^{\infty} k_1 k_2 \cdot \frac{P(n_r = k_1, n_l = k_2)}{P(n > 0)} - \left(\sum_{k_1=1}^{\infty} k_1 \cdot \frac{P(n_r = k_1)}{P(n > 0)}\right)^2 .\end{aligned} \tag{3.25}$$

Using again that X_r and X_l are identically distributed, the correlation can be simplified into

$$\begin{aligned}\rho_{X_r,X_l} &= \frac{\mathrm{Cov}(X_r, X_l)}{\sqrt{\mathrm{Var}(X_r)}\sqrt{\mathrm{Var}(X_l)}} \\ &= \frac{\mathrm{Cov}(X_r, X_l)}{\mathrm{Var}(X_r)},\end{aligned} \tag{3.26}$$

where the variance of X_r is given by

$$\begin{aligned}\mathrm{Var}(X_r) &= E(X_r^2) - E(X_r)^2 \\ &= \sum_{k_1=1}^{\infty} k_1^2 P(X_r = k_1) - \left(\sum_{k_1=1}^{\infty} k_1 P(X_r = k_1)\right)^2 .\end{aligned} \tag{3.27}$$

The correlation of X_r and X_l is illustrated in Fig. 3.6. This correlation is necessarily positive since, as we already indicated before, the cluster size, and hence also the right and left cluster sizes, strongly depend on the time at which the origin mutation takes place – the earlier, the higher the probability to generate a large cluster which in general means that both, right and left cluster size, become large. Moreover, the correlation becomes of course stronger with decreasing λ as low mutation rates allow larger cluster. For $\lambda \to 1$, which leads to clusters of size one with high probability, right and left cluster sizes are zero with high probability anyway, and therefore, correlation is low.
Interestingly, the correlation is approximately linear, in other words the ratio between mutation rate and correlation remains the same for all λ.

3.6 Correlation of cluster sizes

We are now interested in the correlation of adjacent cluster sizes, let us say in the correlation of cluster X and its right neighbor Y with nucleation points 0 and d_c, where

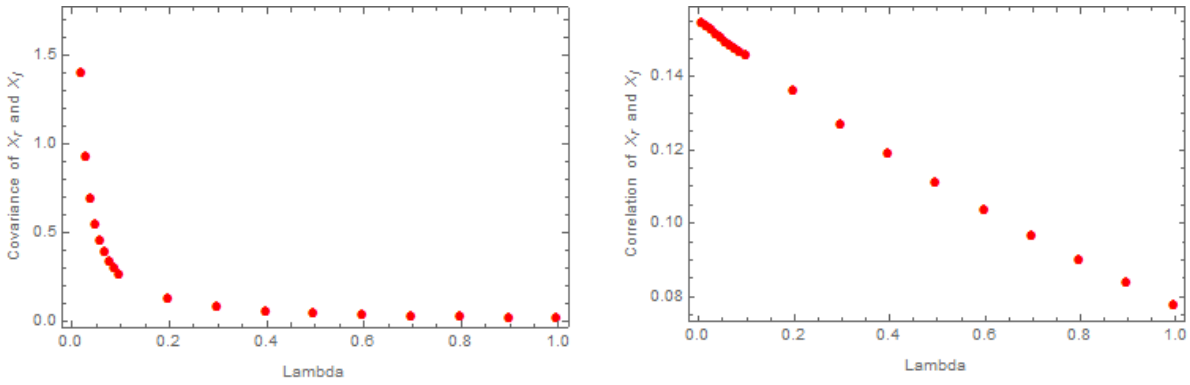

Figure 3.6: *This figure shows the covariance and the correlation of X_r and X_l.*

d_c is the distance between the two nucleation points. The corresponding mutation times are denoted by T_x and T_y, and we call right and left cluster sides X_r, Y_r and X_l, Y_l respectively.
Then, the covariance is given by

$$\begin{aligned} \mathrm{Cov}(X,Y) &= \mathrm{E}(XY) - \mathrm{E}(X)\mathrm{E}(Y) \\ &= \mathrm{E}(XY) - \mathrm{E}(X)^2, \end{aligned} \tag{3.28}$$

where the last equation follows by symmetry since X and Y are identically distributed. By similar arguments as in the previous section, we obtain for the correlation of X and Y

$$\begin{aligned} \rho_{X,Y} &= \frac{\mathrm{Cov}(X,Y)}{\sqrt{\mathrm{Var}(X)}\sqrt{\mathrm{Var}(Y)}} \\ &= \frac{\mathrm{Cov}(X,Y)}{\mathrm{Var}(X)}, \end{aligned} \tag{3.29}$$

with

$$\begin{aligned} \mathrm{Var}(X) &= E(X^2) - E(X)^2 \\ &= \sum_{k=1}^{\infty} k_1^2 P(X=k) - \left(\sum_{k=1}^{\infty} k P(X=k) \right)^2 . \end{aligned} \tag{3.30}$$

Before going into details, we treat the simpler problem of the correlation between X_r and Y_l. We will later use the results that we get from this problem for our origin problem. So, we will have a closer look at

$$\begin{aligned} \mathrm{Cov}(X_r,Y_l) &= \mathrm{E}(X_rY_l) - \mathrm{E}(X_r)\mathrm{E}(Y_l) \\ &= \mathrm{E}(X_rY_l) - \mathrm{E}(X_r)^2, \end{aligned} \tag{3.31}$$

where again the last equation follows because of the identical distribution of X amd Y. We fix time T and obtain

$$\begin{aligned}
\mathrm{E}(X_r) &= \sum_{k_1=1}^{\infty} k_1 \cdot \frac{P(n_r^x = k_1)}{P(n^x > 0)} \\
&= \frac{1}{P(n^x > 0)} \sum_{k_1=1}^{\infty} k_1 \cdot \int_0^{\infty} \lambda e^{-\lambda T} P(C_l^{1,x}(0), N_r^{1,x}(k_1), B_r^{1,x}(k_1), C_r^{1,x}(k_1) | T_x = T) \mathrm{d}T \\
&= \frac{1}{P(n^x > 0)} \sum_{k_1=1}^{\infty} k_1 \cdot \int_0^{\infty} \lambda e^{-\lambda T} P(C_l^{1,x}(0) | T_x = T) P(N_r^{1,x}(k_1) | T_x = T) \\
&\cdot P(B_r^{1,x}(k_1), C_r^{1,x}(k_1) | T_x = T) \mathrm{d}T,
\end{aligned} \tag{3.32}$$

where the events $B_r^{1,x}(k_1), B_l^{1,x}(k_2), C_r^{1,x}(k_1), C_l^{1,x}(k_2), N_r^{1,x}(k_1)$ and $N_l^{1,x}(k_2)$ are defined as before, so as $B_r^1(k_1)$, $B_l^1(k_2)$, $C_r^1(k_1)$, $C_l^1(k_2)$, $N_r^1(k_1)$ and $N_l^1(k_2)$ for cluster X and hence, (3.19) also applies here.
Next, we need to evaluate the expectation value $\mathrm{E}(X_r Y_l)$ which is given by

$$\begin{aligned}
\mathrm{E}(X_r Y_l) &= \sum_{k_1=1}^{\infty} \sum_{k_2=1}^{\infty} k_1 k_2 \cdot P(X_r = k_1, Y_l = k_2) \\
&= \sum_{k_1=1}^{\infty} \sum_{k_2=1}^{\infty} k_1 k_2 \cdot \frac{P(n_r^x = k_1, n_l^y = k_2)}{P(n^x > 0)}.
\end{aligned} \tag{3.33}$$

At this point we have to take into consideration that we want the cluster Y to be the right neighbor of X, that means no other mutation occurs between the nucleation points of these two clusters. In order to obtain clusters with $X_r = k_1 \geq 0$ and $Y_l = k_2 \geq 0$, so $d_c = k_1 + k_2 + 1$, the origin mutation of cluster Y has to arise in a certain time interval depending on T_x which is given by

$$\tau := \tilde{\tau} \backslash (\tilde{\tau} \cap \mathbb{R}_{<0}), \tag{3.34}$$

where

$$\tilde{\tau} := \{T_x + k_1 - k_2 - 1 < T_y < T_x + k_1 - k_2 + 1\}. \tag{3.35}$$

We briefly show that the events $N_r^{1,x}(k_1)$ and $N_l^{1,y}(k_2)$ together with the condition $\tilde{\tau}$ are sufficient to guarantee that no other mutations occur between the two nucleation points that could occupy one of them. Let us start with $k_1, k_2 > 0$. For demes $1 \leq j \leq k_1$, it immediately follows by $N_r^{1,x}(k_1)$ that deme 0 infects deme j before a new mutation can arise in this deme and therefore, deme j cannot infect any other deme, in particular not 0 and d_c. The same holds for demes $k_1 + 1 \leq j \leq d_c - 1$ because of $N_l^{1,y}(k_2)$. Because of the condition $\tilde{\tau}$, X and Y cannot infect each other either.
For $k_1 = 0$ and $k_2 > 0$, no mutation arises in demes j with $1 \leq j \leq d_c - 1$ because

of $N_l^{1,y}(k_2)$ and again, X and Y cannot infect each other because of $\tilde{\tau}$. The analogous argument holds for $k_1 > 0$ and $k_2 = 0$. For the special case $k_1 = k_2 = 0$, where the nucleation points of X and Y are located in adjacent demes, $\tilde{\tau}$ guarantees that the two clusters do not disturb each other.
For further illustration see Fig. 3.7.

Figure 3.7: *This figure shows the two adjacent clusters X and Y with $X_r = 7$ and $Y_l = 4$.*

Hence, the joint probability $P(n_r^x = k_1, n_l^y = k_2)$ for $k_1, k_2 \geq 0$ can be calculated as follows

$$
\begin{aligned}
&P(n_r^x = k_1, n_l^y = k_2) \\
&= \int_0^\infty \int_\tau \lambda e^{-\lambda T} \lambda e^{-\lambda \tilde{T}} P(C_l^{1,x}(0), C_r^{1,y}(0), N_r^{1,x}(k_1), N_l^{1,y}(k_2) | T_x = T, T_y = \tilde{T}, \tau) \mathrm{d}\tilde{T} \mathrm{d}T \\
&= \int_0^\infty \int_\tau \lambda e^{-\lambda T} \lambda e^{-\lambda \tilde{T}} P(C_l^{1,x}(0) | T_x = T, T_y = \tilde{T}, \tau) P(C_r^{1,y}(0) | T_x = T, T_y = \tilde{T}, \tau) \\
&\qquad P(N_r^{1,x}(k_1) | T_x = T, T_y = \tilde{T}, \tau) P(N_l^{1,y}(k_2) | T_x = T, T_y = \tilde{T}, \tau) \mathrm{d}\tilde{T} \mathrm{d}T \\
&= \int_0^\infty \lambda e^{-\lambda T} P(C_l^{1,x}(0) | T_x = T) P(N_r^{1,x}(k_1) | T_x = T) \int_\tau \lambda e^{-\lambda \tilde{T}} P(C_r^{1,y}(0) | T_y = \tilde{T}) \\
&\qquad P(N_l^{1,y}(k_2) | T_y = \tilde{T}) \mathrm{d}\tilde{T} \mathrm{d}T \\
&= \sum_{j=0}^\infty \int_j^{j+1} \lambda e^{-\lambda T} P(C_l^{1,x}(0) | T_x = T) P(N_r^{1,x}(k_1) | T_x = T) \\
&\quad \left(\int_{\max(0,T+k_1-k_2-1)}^{\max(0,j+k_1-k_2)} \lambda e^{-\lambda \tilde{T}} P(C_r^{1,y}(0) | T_y = \tilde{T}) P(N_l^{1,y}(k_2) | T_y = \tilde{T}) \mathrm{d}\tilde{T} \right. \\
&\quad + \int_{\max(0,j+k_1-k_2)}^{\max(0,j+k_1-k_2+1)} \lambda e^{-\lambda \tilde{T}} P(C_r^{1,y}(0) | T_y = \tilde{T}) P(N_l^{1,y}(k_2) | T_y = \tilde{T}) \mathrm{d}\tilde{T} \\
&\quad \left. + \int_{\max(0,j+k_1-k_2+1)}^{\max(0,T+k_1-k_2+1)} \lambda e^{-\lambda \tilde{T}} P(C_r^{1,y}(0) | T_y = \tilde{T}) P(N_l^{1,y}(k_2) | T_y = \tilde{T}) \mathrm{d}\tilde{T} \right) \mathrm{d}T, \qquad (3.36)
\end{aligned}
$$

where we used independence of some of the events for the second equality of (3.36). Moreover,

$$P(C_l^{1,x}(0)|T_x = T, T_y = \tilde{T}, \tau) = P(C_l^{1,x}(0)|T_x = T)$$

and

$$P(C_r^{1,y}(0)|T_x = T, T_y = \tilde{T}, \tau) = P(C_r^{1,y}(0)|T_y = \tilde{T})$$

hold since we defined the event τ such that clusters X and Y do not disturb each others nucleation. Similarly, it holds

$$P(N_r^{1,x}(k_1)|T_x = T, T_y = \tilde{T}, \tau) = P(N_r^{1,x}(k_1)|T_x = T)$$

and

$$P(N_l^{1,y}(k_2)|T_x = T, T_y = \tilde{T}, \tau) = P(N_l^{1,y}(k_2)|T_y = \tilde{T}).$$

Note that $P(C_r^{1,y}(0)|T_y = \tilde{T})$ and $P(N_l^{1,y}(k_2)|T_y = \tilde{T})$ do depend on $d_{\tilde{T}} := \left\lfloor \tilde{T} \right\rfloor$ just as $P(C_l^{1,x}(0)|T_x = T)$ and $P(N_r^{1,x}(k_1)|T_x = T)$ do depend on d_T.
Moreover, it holds that $P(X_r = k_1, Y_l = k_2) = 0$ if $\tau = \varnothing$ for this pair (k_1, k_2).

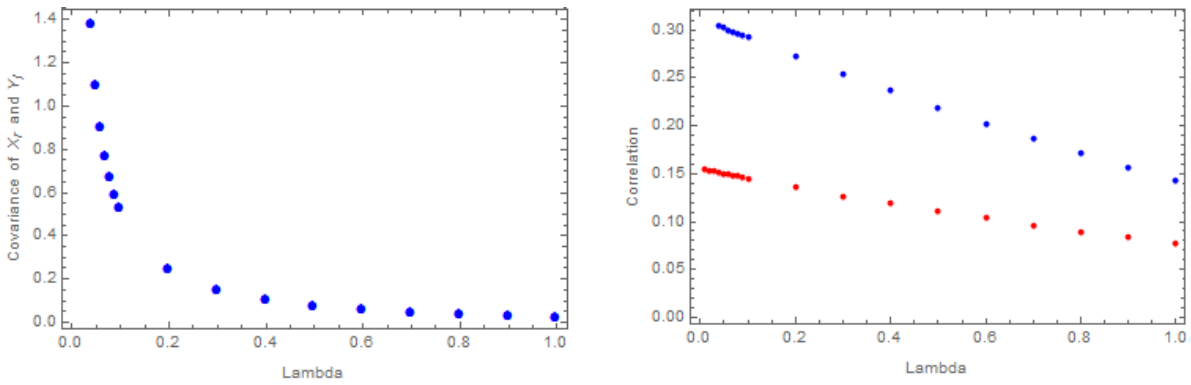

Figure 3.8: *The graphic on the left side shows the covariance of X_r and Y_l, on the right the correlation of X_r and Y_l (blue dots) is compared to that of X_r and X_l (red dots).*

Fig. 3.8 shows the correlation of X_r and Y_l in comparison to the correlation of X_r and X_l. In order to understand these results, we have to discuss two effects which influence the correlation of X_r and Y_l in opposite directions. First, one can fix the mutation times T_x and T_y and keep the distance between the corresponding nucleation points variable. This leads to a distribution of the distances which obviously satisfies that greater distances imply larger values of both, X_r and Y_l. In other words, there is a positive correlation. But on the other hand, there also exists a negative correlation when fixing the distance and keeping the mutation times variable instead. This time one obtains a distribution for the difference of the mutation times for every fixed distance. Now, since we assume an arbitrary but fixed distance, a difference of the mutation times that leads to a large X_r, corresponds to a small value of Y_l.

According to Fig. 3.8, it seems to be the first effect, i.e. the effect of the positive correlation, which dominates. Obviously, the dominance is even strong enough that the correlation of X_r and Y_l is larger than that of X_r and X_l for all mutation rates. Possible reasons for this last point will be discussed below.

We will now come back to our original question, namely the correlation of the cluster sizes X and Y. With the results from above, this becomes a rather simple calculation. The covariance of X and Y is given by

$$\begin{aligned}
&\mathrm{Cov}(X,Y)\\
&= \mathrm{E}(XY) - \mathrm{E}(X)\mathrm{E}(Y)\\
&= \mathrm{E}(XY) - \mathrm{E}(X)^2\\
&= \sum_{k_1=1}^{\infty}\sum_{k_2=1}^{\infty} k_1 k_2 P(X=k_1, Y=k_2) - \left(\sum_{k=1}^{\infty} k P(X=k)\right)^2\\
&= \sum_{k_1=1}^{\infty}\sum_{k_2=1}^{\infty} k_1 k_2 \left(\sum_{i=0}^{k_1-1}\sum_{j=0}^{k_2-1} \frac{P(n_l^x=i, n_r^x=k_1-1-i, n_r^y=j, n_l^y=k_2-1-j)}{P(n^x>0)}\right)\\
&\quad - \left(\sum_{k=1}^{\infty} k \frac{P(n^x=k)}{P(n^x>0)}\right)^2,
\end{aligned} \tag{3.37}$$

where $P(n^x=k)$ is already known from one of the last sections and the joint probability $P(n_l^x=i, n_r^x=k_1-1-i, n_r^y=j, n_l^y=k_2-1-j)$ can be easily derived as

$$\begin{aligned}
&P(n_l^x=i, n_r^x=k_1-1-i, n_r^y=j, n_l^y=k_2-1-j)\\
&= \int_0^{\infty} \lambda e^{-\lambda T} P(B_l^{1,x}(i), C_l^{1,x}(i) | T_x = T) P(N_l^{1,x}(i) | T_x = T) P(N_r^{1,x}(k_1-1-i) | T_x = T)\\
&\int_{\tau} \lambda e^{-\lambda \tilde{T}} P(B_r^{1,y}(j), C_r^{1,y}(j) | T_y = \tilde{T}) P(N_r^{1,y}(j) | T_y = \tilde{T}) P(N_l^{1,y}(k_2-1-j) | T_y = \tilde{T}) \mathrm{d}\tilde{T}\mathrm{d}T,
\end{aligned} \tag{3.38}$$

where $\tilde{\tau}$ is now given by

$$\tilde{\tau} := \{T_x + k_1 - k_2 - i + j - 1 < T_y < T_x + k_1 - k_2 - i + j + 1\}. \tag{3.39}$$

The correlation of X and Y is shown in Fig. 3.9 in comparison with the correlations of X_r/X_l and X_r/Y_l. Again, we basically have the same two opposed effects as for X_r and Y_l. But obviously, this time the positive correlation is weakened and the negative one is strengthened in comparison to before although the positive correlation still dominates the negative such that the total correlation becomes positive. Let us again have a closer look at these two effects in order to explain what we see.

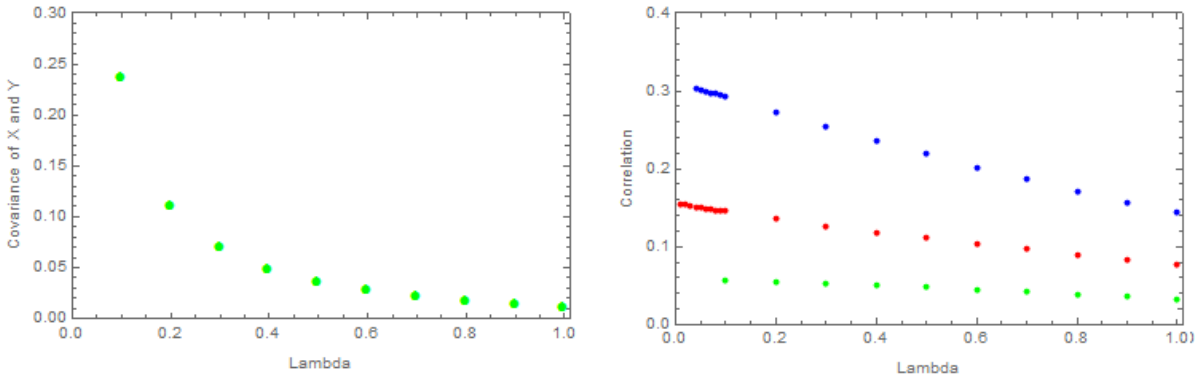

Figure 3.9: *The graphic on the left side shows the covariance of X and Y, on the right the correlation of X and Y (green dots) is compared to that of X_r/X_l (red dots) and X_r/Y_l (blue dots) respectively.*

We start with the negative correlation. We already know that for a fixed distance d_c large values of X_r imply small values of Y_l and vice versa. But a large X_r and small Y_l simply means that the mutation time T_x is relatively small compared to T_y and as a consequence, the whole cluster X has a high chance to become great in comparison to cluster Y. Obviously, this correlation is stronger than the one of X_r and Y_l since now X_l and Y_r are likewise affected and, as we already know, they are positively correlated with X_r and Y_l respectively.
On the other hand, the positive correlation is weakened because the size of the whole clusters X and Y does not only depend on the distance of their own nucleation points but also the left neighbor of X and the right neighbor of Y do play a critical role.
Let us extent this last argument a bit further in a way that it may also explain why the correlation of X_r and X_l is located in the middle of the two others. As we already noticed, the correlation of X_r and Y_l does only depend on two nucleation points. In contrast, two further nucleation points, namely those of the two adjacent clusters of X and Y, do influence the size of clusters X and Y, hence their correlation will be lower. Finally, there is the size of X_r and X_l which is determined by three nucleation points - the one of X and those of the two adjacent clusters.
In order to confirm this idea, let us assume $X_l := X_r$ and $Y_r := Y_l$, i.e. the size of cluster X and Y is no longer determined by four nucleation points but only by the two nucleation points of X and Y. If our idea holds true, then we will expect that the correlation of X and Y equals the one of X_r an Y_l. Indeed, we then obtain

$$\begin{aligned} \mathrm{Cov}(X,Y) &= \mathrm{Cov}(2X_r+1, 2Y_l+1) \\ &= \mathrm{Cov}(2X_r, 2Y_l) \\ &= 4 \cdot \mathrm{Cov}(X_r, Y_l) \end{aligned} \tag{3.40}$$

and

$$\begin{aligned} \mathrm{Var}(X, Y) &= \mathrm{Var}(2X_r + 1, 2Y_l + 1) \\ &= \mathrm{Var}(2X_r, 2Y_l) \\ &= 4 \cdot \mathrm{Var}(X_r, Y_l). \end{aligned} \tag{3.41}$$

Hence, it holds

$$\rho_{X,Y} = \rho_{X_r,Y_l}. \tag{3.42}$$

Therefore, we conclude that the correlation of cluster sizes is directly linked to the number of influencing nucleation points – the more nucleation points, the lower the correlation.

3.7 Distance between nucleation points of adjacent clusters

We like to close this chapter by some results about the distance between the nucleation points of two adjacent clusters. We will also compare this to results from a more general model where we assume nucleation points to be independent.

Let d_c be the distance between the nucleation points of two neighboring clusters. By the definition of X_r and Y_l, it yields $d_c = X_r + Y_l + 1$ and therefore, as X_r and Y_l are identically distributed, the expectation value of d_c becomes

$$\begin{aligned} \mathrm{E}(d_c) &= \mathrm{E}(X_r + Y_l + 1) \\ &= 2 \cdot \mathrm{E}(X_r) + 1 \\ &= \mathrm{E}(X). \end{aligned} \tag{3.43}$$

Similarly, it follows for the variance of d_c

$$\begin{aligned} \mathrm{Var}(d_c) &= \mathrm{Var}(X_r + Y_l) \\ &= 2 \cdot \mathrm{Var}(X_r) + 2 \cdot \mathrm{Cov}(X_r, Y_l). \end{aligned} \tag{3.44}$$

Using the results from the last section, we easily obtain the distribution of the distance d_c as

$$P(d_c = k) = \sum_{i=0}^{k-1} \frac{P(X_r = i, Y_l = k - 1 - i)}{P(n^x > 0)} \tag{3.45}$$

for an arbitrary $k \geq 1$.

Obviously, the nucleation points of clusters X and Y are not independent of each other in our model. If the mutation origins were independent of each other, this would lead to a geometric distribution with $P(n^x > 0)$ as the probability of success, which means that a new mutation can successfully arise in an adjacent deme of cluster X before the

infection of cluster X can spread and infect that neighbor. A distance of size k between neighboring nucleation points would therefore correspond to $k-1$ failures until the first success. We call the distance between neighboring nucleation points in this case D_c. The distribution of the distance between two adjacent nucleation points in this model is given by

$$P(D_c = k) = (1 - P(n^x > 0))^{k-1} P(n^x > 0) \tag{3.46}$$

according to the geometric distribution. Then it yields for the corresponding expectation value

$$\mathrm{E}(D_c) = \frac{1}{P(n^x > 0)} \tag{3.47}$$

and the variance is given by

$$\mathrm{Var}(D_c) = \frac{1}{P(n^x > 0)^2} - \frac{1}{P(n^x > 0)}. \tag{3.48}$$

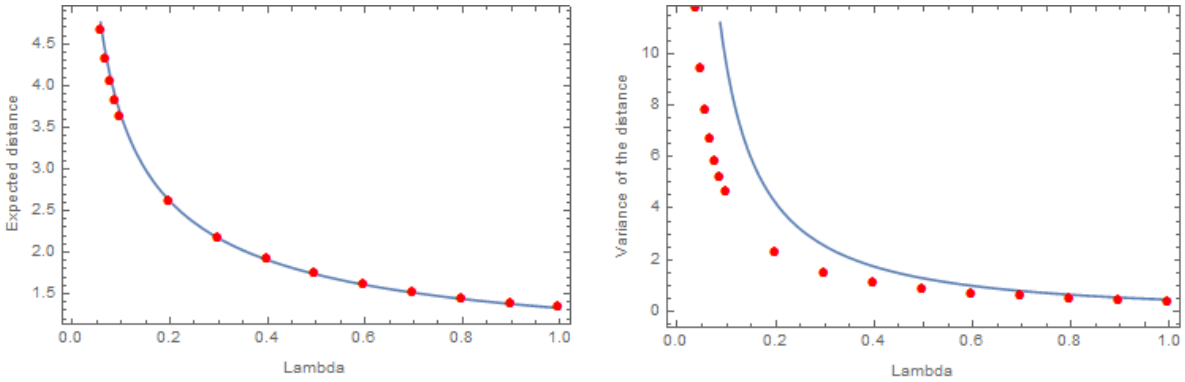

Figure 3.10: *The expectation value and the variance of the distance between adjacent nucleation points are shown. The red dots represent our model, the blue line shows the model with independent nucleation points.*

We compare the expected distance as well as the variance of both models in Fig. 3.10. The solid line represents the second model, i.e. the model with independent nucleation origins, whereas the red dots belong to our model where the origins are not independent of each other.

Obviously, the expected distance between adjacent nucleation origins is the same in both models. This holds since the expectation value of the geometric distribution equals $\frac{1}{P(n^x>0)}$ which, as $P(n^x > 0)$ is the density of clusters, is necessarily the average distance of two adjacent nucleation points in our model.

Of course, the expected distance tends to one for large λ because mutations happen so quickly that beneficial alleles have only a very small chance to spread outward their origin deme. On the other hand, for arbitrarily small λ, mutations happen so slowly

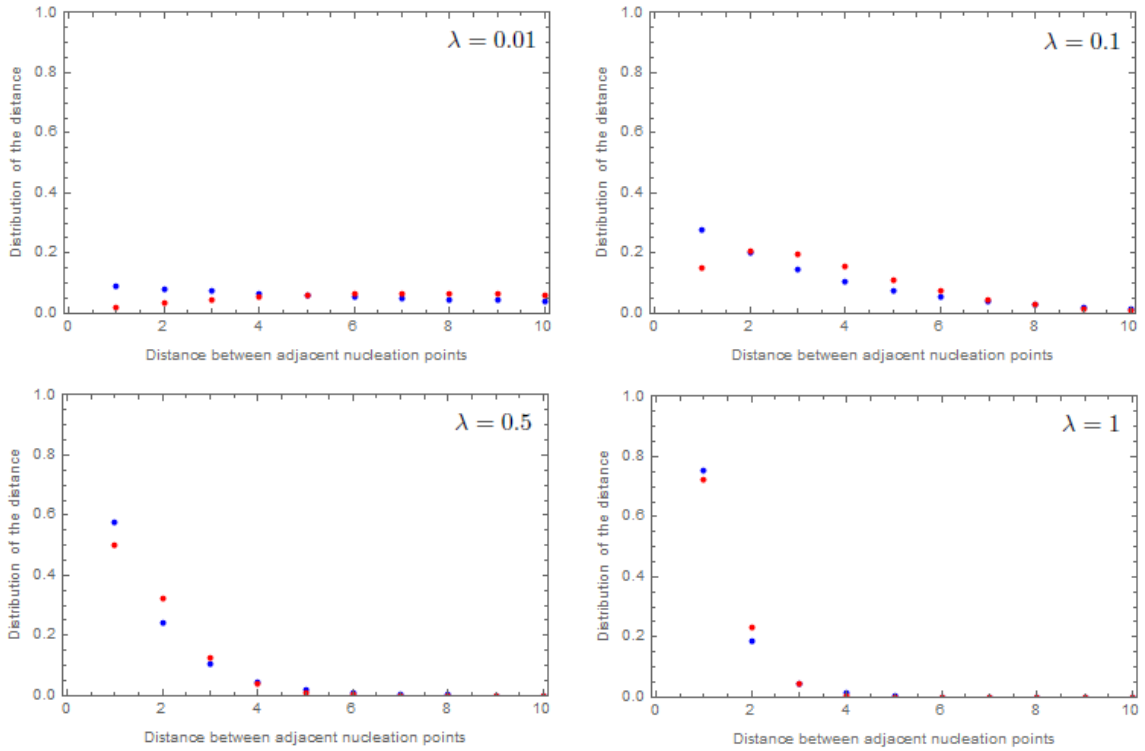

Figure 3.11: *The distribution of distances between adjacent nucleation points is shown for our model (red dots) and the model with independent nucleation points (blue dots).*

that beneficial alleles do have the possibility to form arbitrarily large clusters and hence, the expectation value becomes arbitrarily large for very low mutation rates.
In contrast, the variances of both models differ significantly. Similarly to the expectation value, both variances are small for large mutation rates and tend to infinity as λ goes to zero since the smaller λ, the broader the distribution of cluster sizes, i.e. the higher the chance that not only small but also large clusters do exist. Compared to our model, the variance is larger for the model with independent nucleation points for every mutation rate λ. This can be explained with the help of Fig. 3.11 where the distributions of the distance between adjacent nucleation points of both models are shown for different mutation rates. One can easily see that the distribution of the model with independent nucleation points puts more weight on small distances than the distribution of our model - in comparison with the corresponding expectation value in Fig. 3.10 this means further away from the expectation value. Obviously, the distribution is strictly monotonic decreasing for the model with independent nucleation origins as (3.46) strictly decreases with increasing k whereas, similar to the distribution of cluster sizes, there exists an intermediate maximum for our model.
In order to understand the differences that we observe in the distribution of both models, let us have a closer look at the differences in model assumptions and in particular

at the probability of success. For the model with independent nucleation origins, the probability of success always stays the same, independently of the time that has already past. In contrast, in our model the probability that a new mutation successfully arises in any deme in a certain time interval $[0, T]$ increases with increasing T, i.e. with increasing distance from a given focal deme. But by definition of $P(n^x > 0)$, this probability will always be strictly smaller than $P(n^x > 0)$ for any finite time interval. Hence, if we ignored that there are still other demes that could potentially infect a deme in a local neighborhood of a given focal deme, we would expect much greater distances between the nucleation points of this focal deme and its adjacent demes in our model in comparison to the model with independent nucleation origins. Of course, we can not simply neglect an infection by other demes since this strongly influences the probability that a deme already carries another beneficial allele at the time it would be reached by the infection of the given focal deme, but this does not completely neutralize the effect of the different probabilities of success as one can see in Fig. 3.11.

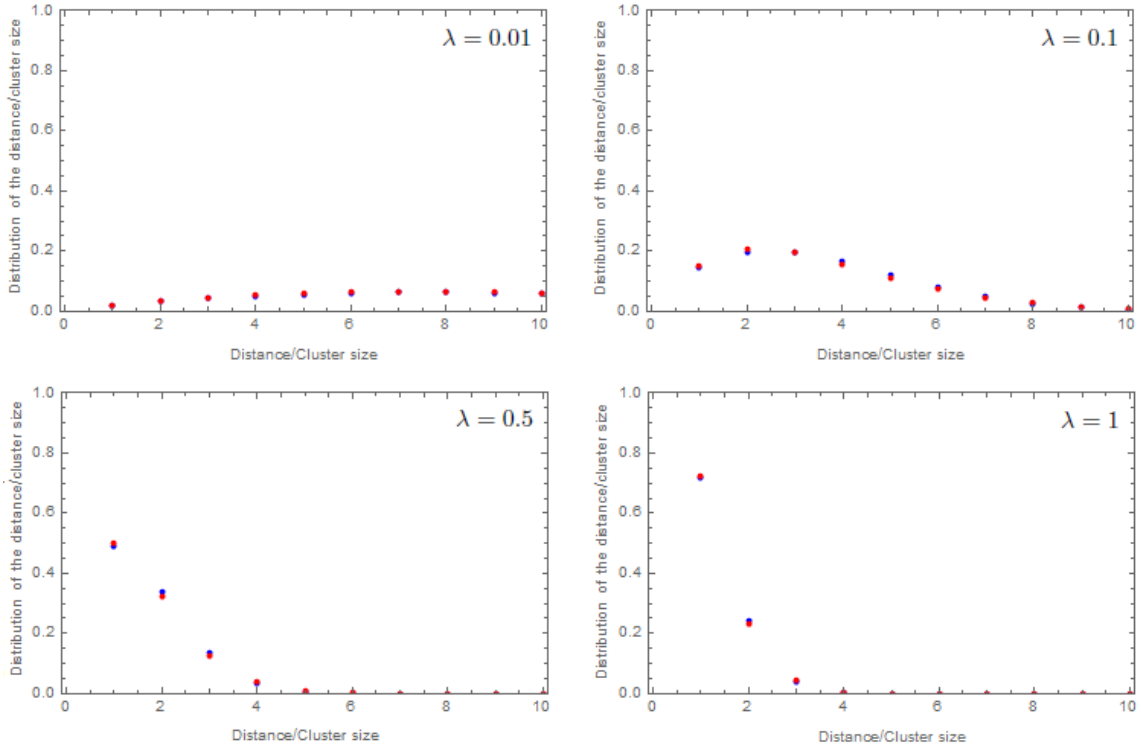

Figure 3.12: *The distribution of the distance between adjacent nucleation points is shown (red dots). In order to show the similarity between the distribution, it is compared to the distribution of cluster sizes (blue dots).*

By Fig. 3.12 we briefly like to point out that the distribution of cluster size and distance between adjacent nucleation points look very similar because they are, of course, related to each other but still, the derivations differ significantly.

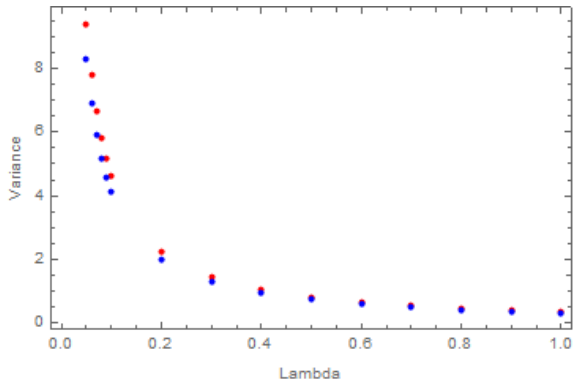

Figure 3.13: *The variance of distances between adjacent nucleation points (red dots) is compared to the variance of cluster sizes (blue dots).*

In Fig. 3.13 the corresponding variances are compared to each other. Although the differences in the distributions are small, the variance of the distance between new mutations is slightly larger than the variance of cluster sizes. This nicely fits the fact that the covariance of X_r and Y_l is larger than the one of X_r and X_l.

With these results about the distance between nucleation points of adjacent clusters, we like to close this chapter about the one-dimensional model. We will once again come back to this model in the last chapter of this thesis where results of this spatial model are compared to results of the model *coalescent with killings*.

4 The two-dimensional model with an infinite grid of quadratical demes

In this chapter we extend our model to an infinite two-dimensional grid of quadratical demes. In the first part we briefly present the model by pointing out the differences to the one-dimensional case and introduce the important parameter of v-step neighbors. As before, we then derive the expected proportion of new mutations as well as the distribution of cluster sizes and compare this to the results of the infinite one-dimensional model. Note that for the derivation of the distribution of cluster sizes we use a simplified version of the two-dimensional model, namely we measure cluster sizes along a horizontal line.

4.1 The model

We extend our model to an infinite two-dimensional grid of quadratical demes as it is shown in Fig. 4.1. Unless stated otherwise, all assumptions and notations of the one-dimensional model will be kept and, in particular, the underlying process is again given by Definition 2.1. In two dimensions each deme has four neighbors and any deme with a beneficial allele can transmit its infection to all of its neighbors that do not carry a mutation, after a deterministic time 1.
We label demes by pairs of numbers (x, y) with $x, y \in \mathbb{Z}$.
The time a mutation needs to spread from a deme (x_1, y_1) to another deme (x_2, y_2) is equivalent to the distance between the two demes which is given by

$$d((x_1, y_1), (x_2, y_2)) := |x_1 - x_2| + |y_1 - y_2|. \tag{4.1}$$

Obviously, this even defines a metric.

A very important parameter in this context is the number of demes that could be reached by a mutation of an arbitrary deme i within a certain time $v \in \mathbb{N}$, so all demes within a distance of at most v from the focal deme i. We call all demes which are reached exactly at time v, the *v-step neighbors* of deme i. The first v-step neighbors (until $v = 4$) of an arbitrary deme are shown in Fig. 4.1 and, as one can easily see by this figure, the

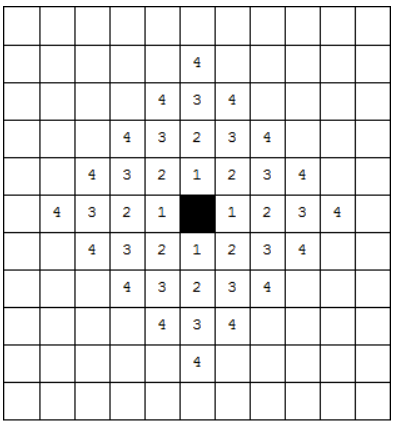

Fig. 4.1: *Demes are arranged chessboard-like in the two-dimensional plane. One deme and its corresponding v-step neighbors until $v = 4$ are highlighted.*

number of v-step neighbors is simply given by the function

$$f_2(v) = 4 * v. \tag{4.2}$$

Note that in the one-dimensional model the number of v-step neighbors is exactly 2 for each v.

We will need this function immediately in order to derive the probability for an arbitrary deme to be the origin of a new beneficial allele.

4.2 Expected frequency of new mutations

In this section we derive the expected proportion of new mutations on the whole grid which is the same as the probability that a new mutation successfully arises in an arbitrary deme. We focus on deme $(0, 0)$ as our focal deme which we will sometimes simply call "deme 0".
Exactly as in the previous chapter, we first fix time $T \in \mathbb{R}_{\geq 0}$ and focus on the probability that at this time the focal deme is not infected by a mutation of another deme and a new mutation u_0 successfully arises instead. In order to guarantee that no other mutation reaches the origin deme before time T, we have to consider all v-step neighbors within the distance $d_T := \lfloor T \rfloor$. Hence, as the following events are pairwise independent and the distribution of the mutation times does not depend on the demes, it follows

$$\begin{aligned} P\left(n > 0 | T_0 = T\right) &= \prod_{v=1}^{d_T} \prod_{v_1=0}^{v} P\left(T_{(\pm v_1, \pm(v-v_1))} > T - v\right) \\ &= \prod_{v=1}^{d_T} P\left(T_0 > T - v\right)^{f_2(v)} \end{aligned}$$

$$\begin{aligned} &= e^{-\lambda T \sum_{v=1}^{d_T} f_2(v)} \cdot e^{\lambda \sum_{v=1}^{d_T} v f_2(v)} \\ &= e^{-2\lambda T d_T(d_T+1)} \cdot e^{\frac{2}{3}\lambda d_T(d_T+1)(2d_T+1)} \end{aligned} \tag{4.3}$$

for $T \geq 1$. For $T < 1$, it obviously yields $P\left(n > 0 | T_0 = T\right) = 1$.
Again, a simple integration over a step function yields the probability for an arbitrary deme to be the origin of a newly arisen mutation

$$\begin{aligned} P(n>0) &= \int_0^\infty \lambda e^{-\lambda T} \cdot P\left(n > 0 | T_0 = T\right) \mathrm{d}T \\ &= \int_0^1 \lambda e^{-\lambda T} \mathrm{d}T + \sum_{j=1}^{\infty} \int_j^{j+1} \lambda e^{-\lambda T} e^{-2\lambda T j(j+1)} \cdot e^{\frac{2}{3}\lambda j(j+1)(2j+1)} \mathrm{d}T \\ &= \sum_{j=0}^{\infty} \frac{1}{2j^2+2j+1} e^{\frac{2}{3}\lambda j(j+1)(2j+1)} \left(e^{-\lambda(2j^2+2j+1)j} - e^{-\lambda(2j^2+2j+1)(j+1)} \right) \\ &= \sum_{j=0}^{\infty} \frac{1}{2j^2+2j+1} \left(e^{-\frac{1}{3}\lambda j(2j^2+1)} - e^{-\frac{1}{3}\lambda(j+1)(2j^2+4j+3)} \right). \end{aligned} \tag{4.4}$$

We compare the expected proportion of new mutations for the one- and the two-dimensional model in Fig. 4.2. The two curves closely resemble each other due to the very similar derivation that was used. But the number of v-step neighbors, i.e. the number of demes that could potentially infect the origin deme in each time step, is of course significantly larger in the two-dimensional model and, consequently, new mutations have a higher chance to arise in a given focal deme in the one-dimensional model than in the two-dimensional grid of demes. Therefore, it is not surprising that the red line lies above the blue one.

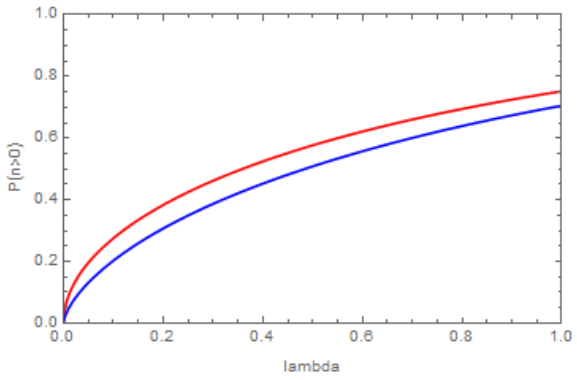

Fig. 4.2: *The expected frequency of new mutations in the one- and the two-dimensional model is shown. The red line represents the one-dimensional, the blue one the two-dimensional model.*

4.3 Distribution of cluster sizes along a horizontal line

As in the previous chapter, we are now interested in the clusters that are generated by the spreading mutations and the distribution of their sizes. But as this problem becomes very complex in the two-dimensional case, we focus on cluster sizes along a horizontal line. Let us consider the horizontal line $(x, 0)$ for $x \in \mathbb{Z}$.
This does not mean that mutations spread along this horizontal line exclusively. Demes still can transmit beneficial alleles in all four directions but the size of a cluster is measured only by demes along the horizontal line $(x, 0)$. For a further illustration see Fig. 4.3.

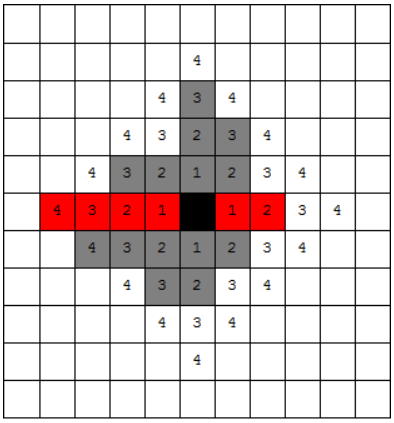

Fig. 4.3: *The beneficial allele of the focal deme (black) has spread and generated a cluster (grey and red). The size of the resulting cluster is measured along a horizontal line (red demes).*

Moreover, n_r and n_l describe again the spreading distance on the right and left side of the origin deme. For instance, the cluster shown in Fig. 4.3 is of size 7 with n_r=2 and n_l=4
We consider clusters of size k and their distribution which is given by

$$\begin{aligned} \frac{P(n=k)}{P(n>0)} &= \frac{\sum_{j=0}^{k-1} P(n_r = j, n_l = k-1-j)}{P(n>0)} \\ &= \frac{\sum_{j=0}^{k-1} \int_0^\infty \lambda e^{-\lambda T} \cdot P(n_r = j, n_l = k-1-j | T_0 = T) \mathrm{d}T}{P(n>0)} \end{aligned} \tag{4.5}$$

for $k \geq 1$ and $\lambda \neq 0$.
A cluster of size k with $n_r = k_1 \geq 0$ and $n_l = k_2 \geq 0$ is described by the intersection of the following events which are simply an extension of the corresponding events of the one-dimensional model, using the idea of a separation between demes on the left and the

right side of the focal deme. But since in the two-dimensional plane there exist demes which are neither on the left nor on the right side of the origin deme, namely all demes $(0, y)$ with $y \neq 0$, an additional condition N_0^2 is required in order to ensure that these demes cannot infect the origin deme before time T_0 either.
Similar to the previous chapter, we give the following definition.

Definition 4.1. *Let $v := v_1 + v_2$. Then, we define the following events:*

$$
\begin{aligned}
B_r^2(k_1) &:= \textit{a mutation coming from a deme } (x,y) \textit{ with } x > k_1 \textit{ reaches deme } (k_1+1,0) \\
&\quad \textit{before time } T_0 + k_1 + 1 \\
&= \{T_0 + k_1 + 1 > \min\left(T_{(k_1+v_1,\pm v_2)} + v - 1 | v_1 \geq 1, v_2 \geq 0, 1 \leq v \leq d_T + k_1 + 2\right)\}, \\
B_l^2(k_2) &:= \textit{a mutation coming from a deme } (x,y) \textit{ with } x < -k_2 \textit{ reaches deme} \\
&\quad (-k_2 - 1, 0) \textit{ before time } T_0 + k_2 + 1 \\
&= \{T_0 + k_2 + 1 > \min\left(T_{(-k_2-v_1,\pm v_2)} + v - 1 | v_1 \geq 1, v_2 \geq 0, 1 \leq v \leq d_T + k_2 + 2\right)\}, \\
C_r^2(k_1) &:= \textit{no mutation coming from a deme } (x,y) \textit{ with } x > k_1 \textit{ reaches deme } (k_1,0) \\
&\quad \textit{before time } T_0 + k_1 \\
&= \{T_0 + k_1 < \min\left(T_{(k_1+v_1,\pm v_2)} + v | v_1 \geq 1, v_2 \geq 0, 1 \leq v \leq d_T + k_1\right)\}, \\
C_l^2(k_2) &:= \textit{no mutation coming from a deme } (x,y) \textit{ with } x < -k_2 \textit{ reaches deme} \\
&\quad (-k_2, 0) \textit{ before time } T_0 + k_2 \\
&= \{T_0 + k_2 < \min\left(T_{(-k_2-v_1,\pm v_2)} + v | v_1 \geq 1, v_2 \geq 0, 1 \leq v \leq d_T + k_2\right)\}, \\
N_r^2(k_1) &:= u_0 \textit{ can reach all demes } (x,0) \textit{ with } 1 \leq x \leq k_1 \textit{ before these demes can be} \\
&\quad \textit{infected by another mutation coming from any deme of the form } (x,y) \\
&= \{T_0 + 1 < \min\left(T_{(1,0)}, T_{(1,\pm 1)} + 1, ..., T_{(1,\pm(d_T+1))} + d_T + 1\right), \\
&\quad\ T_0 + 2 < \min\left(T_{(2,0)}, T_{(2,\pm 1)} + 1, ..., T_{(2,\pm(d_T+2))} + d_T + 2\right), \\
&\quad\ ..., \\
&\quad\ T_0 + k_1 < \min\left(T_{(k_1,0)}, T_{(k_1,\pm 1)} + 1, .., T_{(k_1,\pm(d_T+k_1))} + d_T + k_1\right)\}, \\
N_l^2(k_2) &:= u_0 \textit{ can reach all demes } (-x,0) \textit{ with } 1 \leq x \leq k_2 \textit{ before these demes can be} \\
&\quad \textit{infected by another mutation coming from any deme of the form } (x,y) \\
&= \{T_0 + 1 < \min\left(T_{(-1,0)}, T_{(-1,\pm 1)} + 1, ..., T_{(-1,\pm(d_T+1))} + d_T + 1\right), \\
&\quad\ T_0 + 2 < \min\left(T_{(-2,0)}, T_{(-2,\pm 1)} + 1, ..., T_{(-2,\pm(d_T+2))} + d_T + 2\right), \\
&\quad\ ..., \\
&\quad\ T_0 + k_2 < \min\left(T_{(-k_2,0)}, T_{(-k_2,\pm 1)} + 1, .., T_{(-k_2,\pm(d_T+k_2))} + d_T + k_2\right)\}, \\
N_0^2 &:= \textit{no mutation coming from a deme } (0,y) \textit{ with } y \neq 0 \textit{ reaches the focal deme} \\
&\quad \textit{before time } T_0 \\
&= \{T_0 < \min\left(T_{(0,\pm v_2)} + v_2 | 1 \leq v_2 \leq d_T\right)\}.
\end{aligned}
$$

The probabilities of each of these events conditioned on $T_0 = T$ can be calculated completely analogous to the previous chapter. The calculations are shown in detail in Appendix C. Note that these probabilities strongly depend on the number of v-step neighbors, hence the function $f_2(v)$ plays an important role for these calculations.

The probability for an arbitrary cluster to have right and left cluster size k_1 and k_2 respectively is given by the joint probability of all the events above. Thus, it yields

$$\begin{aligned}
&P(n_r = k_1, n_l = k_2) \\
&= \int_0^\infty \lambda e^{-\lambda T} P(n_r = k_1, n_l = k_2 | T_0 = T) \mathrm{d}T \\
&= \int_0^\infty \lambda e^{-\lambda T} P(N_0^2, N_r^2(k_1), N_l^2(k_2), C_r^2(k_1), B_r^2(k_1), C_l^2(k_2), B_l^2(k_2) | T_0 = T) \mathrm{d}T \\
&= \int_0^\infty \lambda e^{-\lambda T} P(N_0^2 | T_0 = T) P(N_r^2(k_1) | T_0 = T) P(N_l^2(k_2) | T_0 = T) \cdot \Big[P(C_r^2(k_1) | T_0 = T) \\
&\quad + P(B_r^2(k_1) | T_0 = T) - 1 \Big] \cdot \Big[P(C_l^2(k_2) | T_0 = T) + P(B_l^2(k_2) | T_0 = T) - 1 \Big] \mathrm{d}T.
\end{aligned} \tag{4.6}$$

Note that for $P(B_r^2(k_1), C_r^2(k_1) | T_0 = T)$ it still holds

$$P(B_r^2(k_1), C_r^2(k_1) | T_0 = T) = P(C_r^2(k_1) | T_0 = T) + P(B_r^2(k_1) | T_0 = T) - 1 \tag{4.7}$$

as in equation (3.19) of the previous chapter. An analogous result holds, of course, for $P(B_l^2(k_2), C_l^2(k_2) | T_0 = T)$.
We spare ourselves the detailed calculation of $P(n_r = k_1, n_l = k_2)$ because it is very complex and does not yield further useful information. In principle, it follows the same steps as the corresponding calculation of the one-dimensional model which can be found in Appendix B.

Fig. 4.4 shows the distribution of cluster sizes of the one- and the two-dimensional model for different values of the mutation rate λ. Again, the underlying idea that we used for the derivation is very similar in both models. It is thus not surprising that the distributions closely resemble each other: Both models show a flat distribution for low λ allowing also clusters of larger size whereas for large λ most of the weight lies on smaller cluster sizes. But as the number of v-step neighbors is larger in the two-dimensional model, even more weight is put on small clusters compared to the one-dimensional case and this holds for all λ.

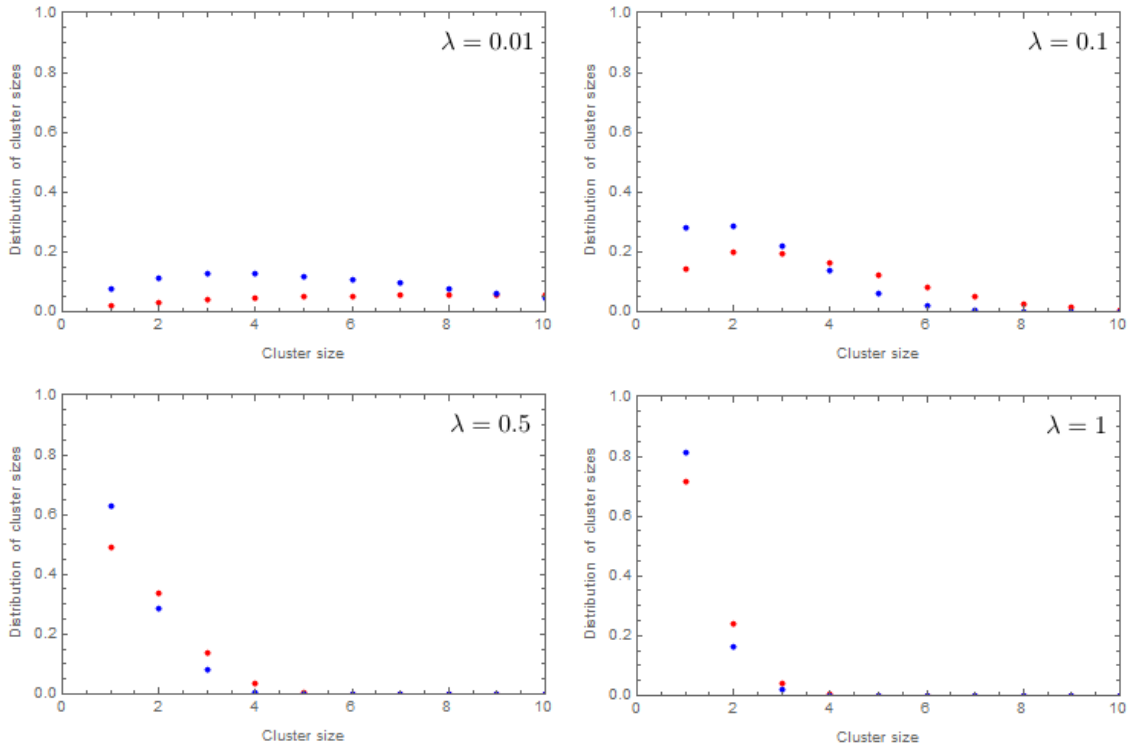

Figure 4.4: *The distribution of cluster sizes of the one-dimensional and the two-dimensional model are compared for different values of λ. Red dots represent the one-dimensional model, blue dots show the results for the two-dimensional model.*

Remark 4.2. *Instead of measuring the size of a cluster along the horizontal line, we could likewise measure it along the vertical line. Let X_h be the cluster size along the horizontal and X_v along the vertical line. These random variables are obviously independent of each and identically distributed, hence it yields*

$$\mathrm{E}(X_h + X_v) = 2 \cdot \mathrm{E}(X_h) - 1, \tag{4.8}$$

where, of course,

$$\mathrm{E}(X_h) = \sum_{k=0}^{\infty} k \cdot \frac{P(n = k)}{P(n > 0)}. \tag{4.9}$$

From this point the extension of the model to three dimensions is only a small step. We will do this in the next chapter where we derive the corresponding results in three dimensions.

5 The model in three dimensions with cubic demes

As a last step we now extend our model to three dimensions. This chapter follows the same structure than the previous. It starts with a short presentation of the model where, in particular, the definition of the number of v-step neighbors in three dimensions differs from the corresponding definition of the two-dimensional model. Then, we derive the expected proportion of new mutations as well as the distribution of cluster sizes and compare the results of all three models. As before, we consider a simplified version of our model for the derivation of the distribution of cluster sizes by measuring cluster sizes along a horizontal line.

5.1 The model

We consider an infinitely large, three-dimensional space filled with infinitely many cubic demes of the same size situated next to each other, so every deme has six direct neighbors in total. Demes are labeled by (x, y, z) with $x, y, z \in \mathbb{Z}$. Then, the distance between two demes (x_1, y_1, z_1) and (x_2, y_2, z_2) is defined by

$$d((x_1, y_1, z_1), (x_2, y_2, z_2)) := |x_1 - x_2| + |y_1 - y_2| + |z_1 - z_2|. \tag{5.1}$$

As long as it is not stated otherwise, all assumptions and notations of our model remain unchanged. The main difference to the assumptions of the model in one or two dimensions is the larger number of direct neighbors of each deme. As a consequence, the number of v-step neighbors also increases in comparison to the models of lower dimensions.
We briefly derive the number of v-step neighbors in three dimensions using the function f_2 which we obtained in two dimensions.
This problem can be reduced to a two-dimensional problem by counting the number of v-step neighbors of an arbitrary deme in each two-dimensional plane. This can easily be done by using our well-known function f_2.
We focus on deme $(0, 0, 0)$ as the focal deme and define plane $z \in \mathbb{Z}$ as the set of all demes whose last coordinate equals z. Then, in plane 0, the number of v-step neighbors equals $f_2(v)$. In plane 1, the number of v-step neighbors is given by $f_2(v-1)$ since it takes the mutation one additional time step to reach this plane. By symmetry, the same holds for plane -1. Similarly, the number of v-step neighbors in plane k and $-k$

(for $1 < k < v$) equals $f_2(v-k)$. Finally, in plane v and $-v$, the only demes which can be reached at time v by a mutation from the origin deme, are demes $(0, 0, v)$ and $(0, 0, -v)$ respectively. Obviously, planes z with $|z| > v$ cannot contain any v-step neighbors of our focal deme.
Putting all this together, it yields for the number of v-step neighbors of an arbitrary deme in three dimensions

$$\begin{aligned} f_3(v) &= \begin{cases} 6 & v = 1 \\ f_2(v) + 2 \cdot \sum_{j=1}^{v-1} f_2(j) + 2 & v > 1 \end{cases} \\ &= \begin{cases} 6 & v = 1 \\ 4v^2 + 2 & v > 1. \end{cases} \end{aligned} \tag{5.2}$$

Remark 5.1. *The function describing the number of v-step neighbors can easily be extended to any dimension $d \geq 1$ by the same underlying arguments that we used for the extension of f_2 to f_3. For an arbitrary $d > 1$, it is given by*

$$f_d(v) = \begin{cases} 2d & v = 1 \\ f_{d-1}(v) + 2 \cdot \sum_{j=1}^{v-1} f_{d-1}(j) + 2 & v > 1. \end{cases} \tag{5.3}$$

For $d = 1$, the number of v-step neighbors is simply the constant function $f_1(v) = 1$ for all v.

With the function f_3, we can now easily derive the expected number of independent origins of beneficial mutations as well as the distribution of the size of their propagation area by following the same steps and arguments that we already used in the previous chapters.

5.2 Expected frequency of new mutations

In this section we derive the expected proportion of new mutations for the three-dimensional model by using the same idea as in one and two dimensions, namely that this is the same as the probability for an arbitrary deme to be the origin of a newly arisen beneficial allele. We already know that at each time the probability that an arbitrary deme does not carry a beneficial mutation by then, is only influenced by all demes whose infection could potentially reach this deme until then.
Let be $T \in \mathbb{R}_{\geq 0}$ a fixed time. Then, the expected frequency of new mutations in our population is given by

$$P(n > 0) = \int_0^\infty \lambda e^{-\lambda T} P(n > 0 | T_0 = T) \mathrm{d}T, \tag{5.4}$$

where for the probability that deme 0 is not yet infected at time T under the condition $T_0 = T$, it holds

$$\begin{aligned} P(n > 0|T_0 = T) &= \prod_{v=1}^{d_T} \prod_{v_1=0}^{v} \prod_{v_2=0}^{v-v_1} P\left(T_{(\pm v_1, \pm v_2, \pm(v-v_1-v_2))} > T - v\right) \\ &= \prod_{v=1}^{d_T} P\left(T_0 > T - v\right)^{f_3(v)} \\ &= e^{-\lambda T \sum_{v=1}^{d_T} f_3(v)} \cdot e^{\lambda \sum_{v=1}^{d_T} v f_3(v)} \end{aligned} \tag{5.5}$$

for $T \geq 1$. For $T < 1$, it obviously yields $P\left(n > 0|T_0 = T\right) = 1$.
The result is shown in Fig. 5.1 where we also compare it to the proportion of new mutations that is expected in one and two dimensions.

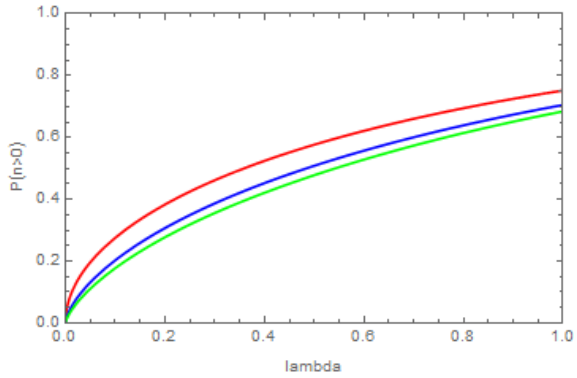

Figure 5.1: *The expected frequencies of new mutations in the one-, two- and three-dimensional model are shown. The red, blue and green line represent the one-, two- and the three-dimensional model respectively.*

We can basically draw the same conclusions from this figure than we did in the last chapter in the context of the comparison of only one and two dimensions. All curves look very similar but the higher the dimension, the higher the number of v-step neighbors at each time and therefore, the expected frequency of new beneficial alleles decreases with higher dimension.

Remark 5.2. *By using the function f_d, we could derive the expected frequency of new mutations for any dimension $d \geq 1$ as the conditional probability $P(n > 0|T_0 = T)$ in general is given by*

$$P(n > 0|T_0 = T) = \prod_{v=1}^{d_T} P(T_0 > T - v)^{f_d(v)} \tag{5.6}$$

for $T \geq 1$. For $T < 1$, it obviously yields $P(n > 0|T_0 = T) = 1$.
Obviously, we would expect the same pattern as can be seen in Fig. 5.1, i.e. the higher the dimension, the lower the expected proportion of new mutations.

The subject of the last part of this chapter will be the distribution of cluster sizes. With the results of this section, it is very likely that we will see a similar pattern as in the two-dimensional case, i.e. a tendency towards small clusters, even smaller than in two dimensions.

5.3 Distribution of cluster sizes along a horizontal line

Similarly to the two-dimensional model, we simplify the question about general cluster sizes by measuring the size of a cluster along a horizontal line. We focus on deme $(0, 0, 0)$ and measure the size of a cluster generated by this deme along the horizontal line $(x, 0, 0)$ for $x \in \mathbb{Z}$.
Given that a beneficial allele successfully arises in this deme, we can derive the probability that this cluster is of a certain size in a similar way as in the previous chapters. Since this probability is the same for all demes, the distribution of cluster sizes in the whole space is again given by our well-known formula

$$\begin{aligned} \frac{P(n=k)}{P(n>0)} &= \frac{\sum_{j=0}^{k-1} P(n_r = j, n_l = k-1-j)}{P(n>0)} \\ &= \frac{\sum_{j=0}^{k-1} \int_0^\infty \lambda e^{-\lambda T} \cdot P(n_r = j, n_l = k-1-j|T_0 = T)\mathrm{d}T}{P(n>0)} \end{aligned} \tag{5.7}$$

for $k \geq 1$ and $\lambda \neq 0$.
For the derivation of $P(n_r = j, n_l = k-1-j)$ we use the same idea as in one and two dimensions. Only some slight modifications concerning the additional third coordinate are needed in order to adjust the events of Definition 4.1 from the two-dimensional to the three-dimensional model.

Definition 5.3. *Let $v := v_1 + v_2 + v_3$. Then, we define the following events:*

$$\begin{aligned} B_r^3(k_1) &:= \textit{a mutation coming from a deme } (x,y,z) \textit{ with } x > k_1 \textit{ reaches deme} \\ &\quad (k_1+1,0,0) \textit{ before time } T_0 + k_1 + 1 \\ &= \{T_0 + k_1 + 1 > \min(T_{(k_1+v_1, \pm v_2, \pm v_3)} + v - 1 | v_1 \geq 1, v_2, v_3 \geq 0, \\ &\quad 1 \leq v \leq d_T + k_1 + 2)\}, \end{aligned}$$

$$
\begin{aligned}
B_l^3(k_2) &:= \textit{a mutation coming from a deme } (x,y,z) \textit{ with } x < -k_2 \textit{ reaches deme} \\
&\quad (-k_2+1,0,0) \textit{ before time } T_0 + k_2 + 1 \\
&= \{T_0 + k_2 + 1 > \min(T_{(-(k_2+v_1),\pm v_2,\pm v_3)} + v - 1 | v_1 \geq 1, v_2, v_3 \geq 0, \\
&\quad 1 \leq v \leq d_T + k_2 + 2)\}, \\
C_r^3(k_1) &:= \textit{no mutation coming from a deme } (x,y,z) \textit{ with } x > k_1 \textit{ reaches deme } (k_1,0,0) \\
&\quad \textit{before time } T_0 + k_1 \\
&= \{T_0 + k_1 < \min(T_{(k_1+v_1,\pm v_2,\pm v_3)} + v | v_1 \geq 1, v_2, v_3 \geq 0, 1 \leq v \leq d_T + k_1)\}, \\
C_l^3(k_2) &:= \textit{no mutation coming from a deme } (x,y,z) \textit{ with } x < -k_2 \textit{ reaches deme} \\
&\quad (-k_2,0,0) \textit{ before time } T_0 + k_2 \\
&= \{T_0 + k_2 < \min(T_{(-(k_2+v_1),\pm v_2,\pm v_3)} + v | v_1 \geq 1, v_2, v_3 \geq 0, 1 \leq v \leq d_T + k_2)\}, \\
N_r^3(k_1) &:= u_0 \textit{ can reach all demes } (x,0,0) \textit{ with } 1 \leq x \leq k_1 \textit{ before these demes can be} \\
&\quad \textit{infected by another mutation coming from any deme of the form } (x,y,z) \\
&= \{T_0 + 1 < \min(T_{(1,\pm v_2,\pm v_3)} + v_2 + v_3 | v_2, v_3 \geq 0, 0 \leq v_2 + v_3 \leq d_T + 1), \\
&\quad T_0 + 2 < \min(T_{(2,\pm v_2,\pm v_3)} + v_2 + v_3 | v_2, v_3 \geq 0, 0 \leq v_2 + v_3 \leq d_T + 2)k, \\
&\quad ..., \\
&\quad T_0 + k_1 < \min(T_{(k_1,\pm v_2,\pm v_3)} + v_2 + v_3 | v_2, v_3 \geq 0, 0 \leq v_2 + v_3 \leq d_T + k_1)\}, \\
N_l^3(k_2) &:= u_0 \textit{ can reach all demes } (-x,0,0) \textit{ with } 1 \leq x \leq k_2 \textit{ before these demes can be} \\
&\quad \textit{infected by another mutation coming from any deme of the form } (x,y,z) \\
&= \{T_0 + 1 < \min(T_{(-1,\pm v_2,\pm v_3)} + v_2 + v_3 | v_2, v_3 \geq 0, 0 \leq v_2 + v_3 \leq d_T + 1), \\
&\quad T_0 + 2 < \min(T_{(-2,\pm v_2,\pm v_3)} + v_2 + v_3 | v_2, v_3 \geq 0, 0 \leq v_2 + v_3 \leq d_T + 2), \\
&\quad ..., \\
&\quad T_0 + k_2 < \min(T_{(-k_2,\pm v_2,\pm v_3)} + v_2 + v_3 | v_2, v_3 \geq 0, 0 \leq v_2 + v_3 \leq d_T + k_2)\}, \\
N_0^3 &:= \textit{no mutation coming from a deme } (0,y,z) \textit{ with } y,z \neq 0 \textit{ reaches the focal} \\
&\quad \textit{deme before time } T_0 \\
&= \{T_0 < \min(T_{(0,\pm v_2,\pm v_3)} + \underbrace{v_2 + v_3}_{=v} | v_2, v_3 \geq 0, 1 \leq v \leq d_T)\}.
\end{aligned}
$$

Again, we need the probability of each of these events conditioned on $T_0 = T$. The corresponding calculations follow the same pattern as in one and two dimensions and can be found in Appendix D.

Moreover, also equation (3.19) is still valid here:

$$P\left(B_r^3(k_1), C_r^3(k_1) | T_0 = T\right) = P\left(C_r^3(k_1) | T_0 = T\right) + P\left(B_r^3(k_1) | T_0 = T\right) - 1 \qquad (5.8)$$

and the same applies to $P\left(B_l^3(k_2), C_l^3(k_2) | T_0 = T\right)$.

Now, we have all the ingredients for the calculation of the joint probability $P(n_r = k_1, n_l = k_2)$ which becomes

$$\begin{aligned}
&P(n_r = k_1, n_l = k_2) \\
&= \int_0^\infty \lambda e^{-\lambda T} P(n_r = k_1, n_l = k_2 | T_0 = T) \mathrm{d}T \\
&= \int_0^\infty \lambda e^{-\lambda T} P(N_0^3, N_r^3(k_1), N_l^3(k_2), C_r^3(k_1), B_r^3(k_1), C_l^3(k_2), B_l^3(k_2) | T_0 = T) \mathrm{d}T \\
&= \int_0^\infty \lambda e^{-\lambda T} P(N_0^3 | T_0 = T) P(N_r^3(k_1) | T_0 = T) P(N_l^3(k_2) | T_0 = T) \cdot \Big[P(C_r^3(k_1) | T_0 = T) \\
&\quad + P(B_r^3(k_1) | T_0 = T) - 1 \Big] \cdot \Big[P(C_l^3(k_2) | T_0 = T) + P(B_l^3(k_2) | T_0 = T) - 1 \Big] \mathrm{d}T.
\end{aligned} \tag{5.9}$$

Putting this into (5.7) yields the distribution of cluster sizes which is illustrated in Fig. 5.2 for different values of the mutation rate λ. Moreover, the results of one and two dimensions are also shown.

Here, we exactly see what we have already expected at the end of section 5.2, namely that the distribution is again similar to those of the one- and the two-dimensional case but the main weight lies even more on small cluster sizes. For large λ, especially for $\lambda = 1$ but even for $\lambda = 0.5$, almost all weight is put on clusters of size one, i.e., it is nearly impossible for a beneficial allele to spread at all after it has arised.

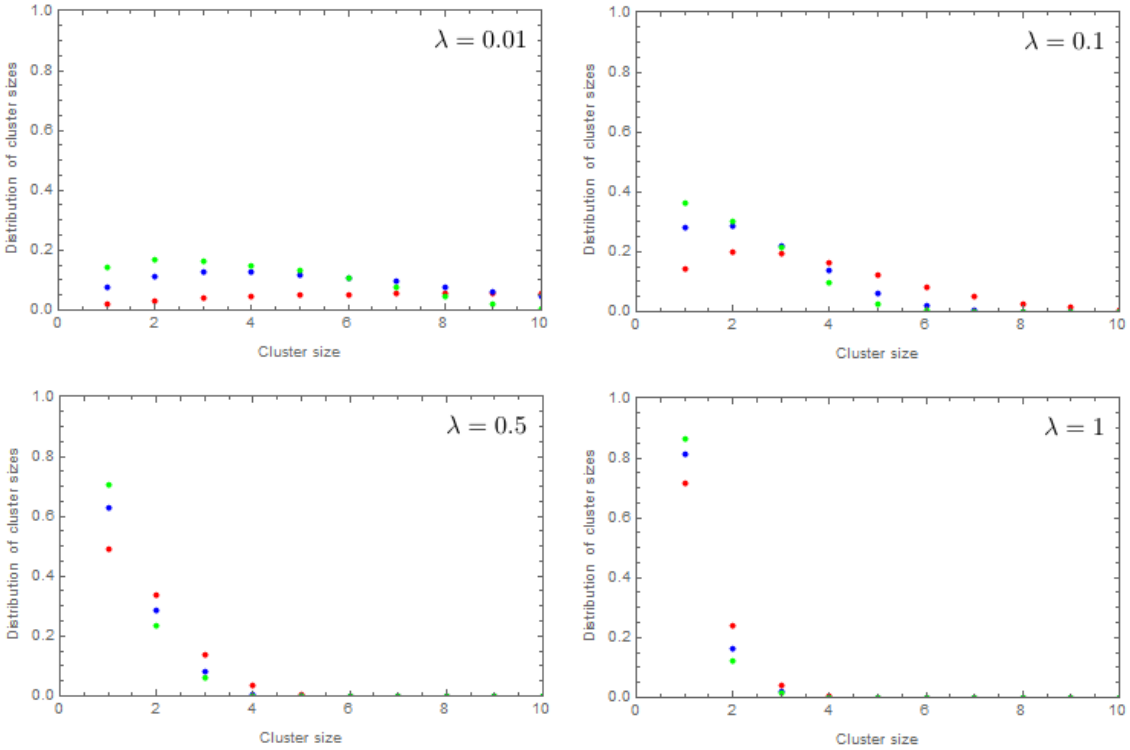

Figure 5.2: *The distribution of cluster sizes of the one-, two- and three-dimensional model are compared for different values of λ. Red dots represent the one-dimensional model, blue and green dots show the results for the two- and the three-dimensional model.*

6 Comparison with the model *coalescent with killings*

In this chapter we compare some of our results of the one-dimensional model to a model from coalescence theory, namely the model known as *coalescent with killings*. In the first two sections we derive the summary statistics that we then use in the last section for the comparison of the two models. We focus on the number of beneficial alleles and the distribution of cluster sizes. The last one is described by the well-known *Ewens' sampling formula* for the coalescent with killings.

6.1 The model *coalescent with killings*

The amount of variation in a panmictic population governed by the infinite alleles model can be characterized by drawing a random sample and calculating summary statistics as the number of different alleles or the allelic partition in the sample.
Let K_m be the number of different beneficial alleles which can be observed in a sample of size m and a_j the number of alleles that appear exactly j times in that sample. Then, the vector $\mathbf{a} := (a_1, ..., a_m)$ is called the *allelic partition* of our data.
As an easy consequence, the following identities will be satisfied

$$\sum_{j=1}^{m} a_j = k \tag{6.1}$$

and

$$\sum_{j=1}^{m} j \cdot a_j = m. \tag{6.2}$$

One approach to analyze the sampling distributions of these two statistics is the so-called *coalescent with killings*. This model is based on the assumption that for the infinite alleles model only the most recent mutations can be observed. Therefore, when constructing the genealogical tree of a sample, we only need to trace lineages back until we encounter the most recent mutation or the root of the tree. We then know the genetical state of this ancestor as well as of all its descendants, thus there is no need to construct the genealogy of this lineage further back in time, so we can *kill* this branch.
Based on these observations, our model contains two kind of events for a tree in state m, i.e. a tree with m ancestral lines. First, there exist coalescence events. This means that

two lineages are chosen uniformly at random and merged at rate $\binom{m}{2}$ which reduces the number of lineages by one. On the other hand, a mutation event occurs at rate $m \cdot \frac{\theta}{2}$ and since the corresponding branch is then removed from the coalescent, this reduces the number of lineages to $m-1$ as well and moreover, the number of alleles decreases by one. The parameter θ is called the *population mutation parameter*. It is defined as $\theta := 2Nu$ for a haploid population where N denotes the size ot the total population and u is the mutation rate per generation per individual in a Wright-Fisher model. Note that the waiting time until the next event, which is either a coalescence or a mutation, is exponentially distributed with parameter $\binom{m}{2} + m \cdot \frac{\theta}{2}$.
By using these observations, we find the following relation of K_m and K_{m-1}

$$\begin{aligned} P(K_m = k) &= \frac{m \cdot \frac{\theta}{2}}{\binom{m}{2} + m \cdot \frac{\theta}{2}} P(K_{m-1} = k-1) + \frac{\binom{m}{2}}{\binom{m}{2} + m \cdot \frac{\theta}{2}} P(K_{m-1} = k) \\ &= \frac{\theta}{m-1+\theta} P(K_{m-1} = k-1) + \frac{m-1}{m-1+\theta} P(K_{m-1} = k). \end{aligned} \tag{6.3}$$

Given the initial condition $P(K_1 = 1)$, this recursion can be solved by

$$P(K_m = k) = \frac{\theta^k}{\theta_{(m)}} \cdot S(m,k) \tag{6.4}$$

where

$$\theta_{(m)} = \theta(\theta+1)...(\theta+m-1). \tag{6.5}$$

The expression $S(m,k)$ denotes the *Stirling numbers of the first kind* which satisfy the recurrence

$$S(m,k) = S(m-1,k-1) + (m-1) \cdot S(m-1,k). \tag{6.6}$$

Moreover, we are interested in the allelic partition of our data which is given by the *Ewens' sampling formula.*

Theorem 6.1. *Let a_i be the number of alleles that appear exactly i times in a sample of size m. Then, under the standard neutral model, the distribution of the vector $\boldsymbol{a} := (a_1, ..., a_m)$ is given by the* ***Ewens' sampling formula****:*

$$P_m(\boldsymbol{a}) = \frac{m!}{\theta_{(m)}} \prod_{j=1}^{m} \frac{(\frac{\theta}{j})^{a_j}}{a_j!}. \tag{6.7}$$

For a proof of this theorem see e.g. Durrett (2008), or Griffiths and Lessard (2005) where a proof based on combinatorical arguments is given.
Obviously, for $m = 2$ and $m = 3$ the Ewens' sampling formula is equivalent to (6.4).

In the following section we compare results that we obtain from the model described above to results that can be received from our model.

6.2 Derivation of $P(K_m = k)$ in our model

In Chapter 3 we considered a chain of infinitely many demes and derived the distribution of cluster sizes. Out of these demes, we now choose finitely many demes in a local neighborhood, i.e. M demes next to each other, and have a look at how many different clusters will be contained in a sample of m individuals that we draw with replacement from these M demes. We restrict ourselves to the cases $m = 2$ and $m = 3$ as the case $m > 3$ becomes much more complex and would exceed the scope of this thesis.
As all individuals of a deme are genetically equal and we sample with replacement, it makes no difference if we draw a single individual or the whole deme instead. In the follwing we will always consider a sample of demes that we draw with replacement.
The underlying idea is to characterize the probability that the sample contains a certain number of demes drawn from the same cluster, based on the distance between the sampled demes.

6.2.1 Sample of size two

We denote the two sampled demes by m_1 and m_2 such that, if the sample is not drawn from the same deme anyway, m_1 is located to the left of m_2. The distance between them is defined as $d_s \in [0, M-1]$.
Using the notation of the previous section, the probability that K_2 takes the value one, hence m_1 and m_2 belong to the same cluster, is given by

$$P(K_2 = 1) = \sum_{k=0}^{M-1} P(K_m = 1 | d_s = k) P(d_s = k). \tag{6.8}$$

First, we derive the distribution of d_s. There exist $\frac{M(M-1)}{2}$ possibilities to draw m_1 and m_2 from two distinct demes and M possibilities to draw them from the same. Moreover, for a fixed $k \in [0, M-1]$, there are $M - k$ possibilities to place the left deme m_1 in a way that m_2 is still contained in our selected set of M demes. Hence, we obtain for the distribution of d_s

$$\begin{aligned} P(d_s = k) &= \frac{M-k}{\frac{M(M-1)}{2} + M} \\ &= \frac{2(M-k)}{M(M+1)}. \end{aligned} \tag{6.9}$$

Next, we have a closer look at the derivation of the conditioned probability $P(K_2 = 1 | d_s = k)$. In order to guarantee that, given a distance of k between them, the two sampled demes are part of the same cluster, m_1 has to be contained in a cluster of size at least $k + 1$ and it needs to be at least k demes away from the right border of this cluster. The probability that m_1 is sampled from a deme of size j is $j \cdot P(n = j)$. With

probability $\frac{j-k}{j}$ its distance to the right border of the cluster is at least k. Therefore, it yields

$$\begin{aligned} P(K_2 = 1) &= \sum_{k=0}^{M-1} P(K_2 = 1 | d_s = k) P(d_s = k) \\ &= \sum_{k=0}^{M-1} \sum_{j>k} (j-k) \cdot P(n = j) \cdot \frac{2(M-k)}{M(M+1)}. \end{aligned} \tag{6.10}$$

As a simple consequence, the probability to draw m_1 and m_2 from two different clusters is given by

$$P(K_2 = 2) = 1 - P(K_2 = 1). \tag{6.11}$$

6.2.2 Sample of size three

In the case of $m = 3$, there are three possibilities. Either we sample all demes from the same cluster, or two of the demes come from the same cluster whereas the third belongs to a different cluster, or all three sampled demes are located in different clusters.
We start by deriving the probability to sample three demes from the same cluster in the same way as for $m = 2$.
Similarly as before, we label the sampled demes as m_1, m_2 and m_3 such that m_1 is situated to the left, m_2 in the middle and m_3 to right of the other two demes. The distances between m_1 and m_2 as well as between m_2 and m_3 are denoted by $d_{s_1} \in [0, M-1]$ and $d_{s_2} \in [0, M-1-d_{s_1}]$ respectively. Defining d_s as the distance between left and right deme, it immediately yields $d_s = d_{s_1} + d_{s_2}$.
Then, we obtain for the probability that K_3 takes the value one

$$P(K_3 = 1) = \sum_{k_1=0}^{M-1} \sum_{k_2=0}^{M-1-k_1} P(K_3 = 1 | d_{s_1} = k_1, d_{s_2} = k_2) P(d_{s_1} = k_1, d_{s_2} = k_2). \tag{6.12}$$

The distribution of $P(d_{s_1} = k_1, d_{s_2} = k_2)$ can easily be derived. We set $k := k_1 + k_2$ and observe that for given values of k_1 and k_2, there are $M-k$ possibilities to place m_1 such that m_2 and m_3 are still contained in our set of M demes. If only k is given, there still exist $M-k$ possibilities to place m_1 in an appropriate way and moreover, there are $k+1$ possible positions for m_2 between the two other demes of the sample. Hence, it yields

$$P(d_{s_1} = k_1, d_{s_2} = k_2) = \frac{M-k}{\sum_{k=0}^{M-1} (M-k)(k+1)}. \tag{6.13}$$

Using exactly the same arguments for the derivation of $P(K_3 = 1|d_{s_1} = k_1, d_{s_2} = k_2)$ as in the previous section, we finally obtain

$$\begin{aligned} P(K_3 = 1) &= \sum_{k_1=0}^{M-1} \sum_{k_2=0}^{M-1-k_1} P(K_3 = 1|d_{s_1} = k_1, d_{s_2} = k_2) P(d_{s_1} = k_1, d_{s_2} = k_2) \\ &= \sum_{k_1=0}^{M-1} \sum_{k_2=0}^{M-1-k_1} \sum_{j>k} (j-k) \cdot P(n = j) \cdot \frac{M-k}{\sum_{k=0}^{M-1}(M-k)(k+1)}. \end{aligned} \tag{6.14}$$

For the case $K_3 = 2$, there are now two possibilities. Either m_1 and m_2 belong to the same cluster and m_3 does not, or m_2 and m_3 come from another cluster than m_1. Let us call these two clusters A and B. Then, we obtain

$$\begin{aligned} &P(K_3 = 2) \\ &= \sum_{k_1=0}^{M-1} \sum_{k_2=0}^{M-1-k_1} \Big[P(m_1, m_2 \in A, m_3 \in B|d_{s_1} = k_1, d_{s_2} = k_2) \\ &\quad + P(m_1 \in A, m_2, m_3 \in B|d_{s_1} = k_1, d_{s_2} = k_2) \Big] P(d_{s_1} = k_1, d_{s_2} = k_2). \end{aligned} \tag{6.15}$$

Let us focus on the first case, i.e. m_1, $m_2 \in A$ and $m_3 \in B$. The remaining case then immediately follows by symmetry.
In order to guarantee that m_1 and m_2 are part of the same cluster, it is again required that cluster A is at least of size $k_1 + 1$ and that the distance between m_1 and the right border of this cluster is at least k_1. Concerning the condition that m_3 belongs to another cluster, we have to distinguish two cases for the size j of cluster A. If $j \leq k$, then it is impossible that all demes are sampled from the same cluster and the probability $P(m_1, m_2 \in A, m_3 \in B|d_{s_1} = k_1, d_{s_2} = k_2)$ reduces to $P(m_1, m_2 \in A|d_{s_1} = k_1, d_{s_2} = k_2)$. But if $j > k$, we need an additional constraint for the position of m_1 within cluster A. Its maximal distance from the right border is now given by $k_1 + k_2 - 1$, otherwise m_3 would also be part of cluster A. Taking also the minimal distance between m_1 and the right border into account, which is given by k_1, it yields $\frac{k_2}{j}$ for the probability of this event and hence, we obtain

$$\begin{aligned} &P(m_1, m_2 \in A, m_3 \in B|d_{s_1} = k_1, d_{s_2} = k_2) \\ &= \sum_{j \geq k_1} jP(n = j) \cdot \left(\mathbb{1}_{\{j \leq k\}} \cdot \frac{j - k_1}{j} + \mathbb{1}_{\{j>k\}} \cdot \frac{k_2}{j} \right) \end{aligned} \tag{6.16}$$

By symmetry, equation (6.15) becomes

$$\begin{aligned} &P(K_3 = 2) \\ &= \sum_{k_1=0}^{M-1} \sum_{k_2=0}^{M-1-k_1} \Bigg[\sum_{j>k_1} P(n=j) \cdot \left(\mathbb{1}_{\{j \leq k\}} \cdot (j-k_1) + \mathbb{1}_{\{j>k\}} \cdot k_2 \right) \\ &\quad + \sum_{j>k_2} P(n=j) \cdot \left(\mathbb{1}_{\{j \leq k\}} \cdot (j-k_2) + \mathbb{1}_{\{j>k\}} \cdot k_1 \right) \Bigg] \frac{M-k}{\sum_{k=0}^{M-1} (M-k)(k+1)}. \end{aligned} \tag{6.17}$$

Finally, it yields for the probability that all sampled demes belong to different clusters

$$P(K_3 = 3) = 1 - P(K_3 = 1) - P(K_3 = 2). \tag{6.18}$$

Remark 6.2. *Note that for $m = 2$ and $m = 3$ this ansatz requires no information about the correlation of cluster sizes. But it is not an appropriate ansatz for larger sample size as we would not only need information about the correlation of adjacent cluster sizes but also about the correlation of clusters which are not next to each other.*

6.3 Comparison of the two models

In this section we finally compare the two models to each other by developping a concept which could be further used for hypothesis testing in order to decide which model best describes a given population with unknown structure.

In the model *coalescent with killings*, we consider a panmictic population with N individuals whereas in our model we have a spatially structured population in a local neighborhood of M demes of equal size, so each deme consists of $\frac{N}{M}$ individuals. Here, we assume haploid populations, hence the population mutation parameter is given by $\theta = 2Nu$ where u is again the mutation rate per generation per individual.
The first step of this comparison consists of an appropriate choice of the mutation parameter θ. We define θ such that for a sample of size two, the models are not distinguishable, i.e. they have the same probabilities for $K_2 = 1$ as well as for $K_2 = 2$. Let $P^*(K_2 = 1)$ and $P^*(K_2 = 2)$ be the solution of (6.10) and (6.11) respectively. Then, as $P^*(K_2 = 1) + P^*(K_2 = 2) = 1$, it immediately follows for θ by (6.4)

$$\theta = \frac{P^*(K_2 = 2)}{P^*(K_2 = 1)}. \tag{6.19}$$

This θ is then fixed for the coalescent with killings. Note that it depends on the corresponding mutation rate λ of our spatial model.
Now, the underlying idea is that one can assume one of the models as the null model which describes the structure of the given population, and although we cannot decide

on the underlying model by samples of size two, we could do so by randomly sampling three individuals and using some statistical methods as hypothesis testing. Although we do not want to go further into details concerning hypothesis testing here, we briefly like to discuss the results of the two models for a sample of size three. These results are shown in Fig. 6.1.

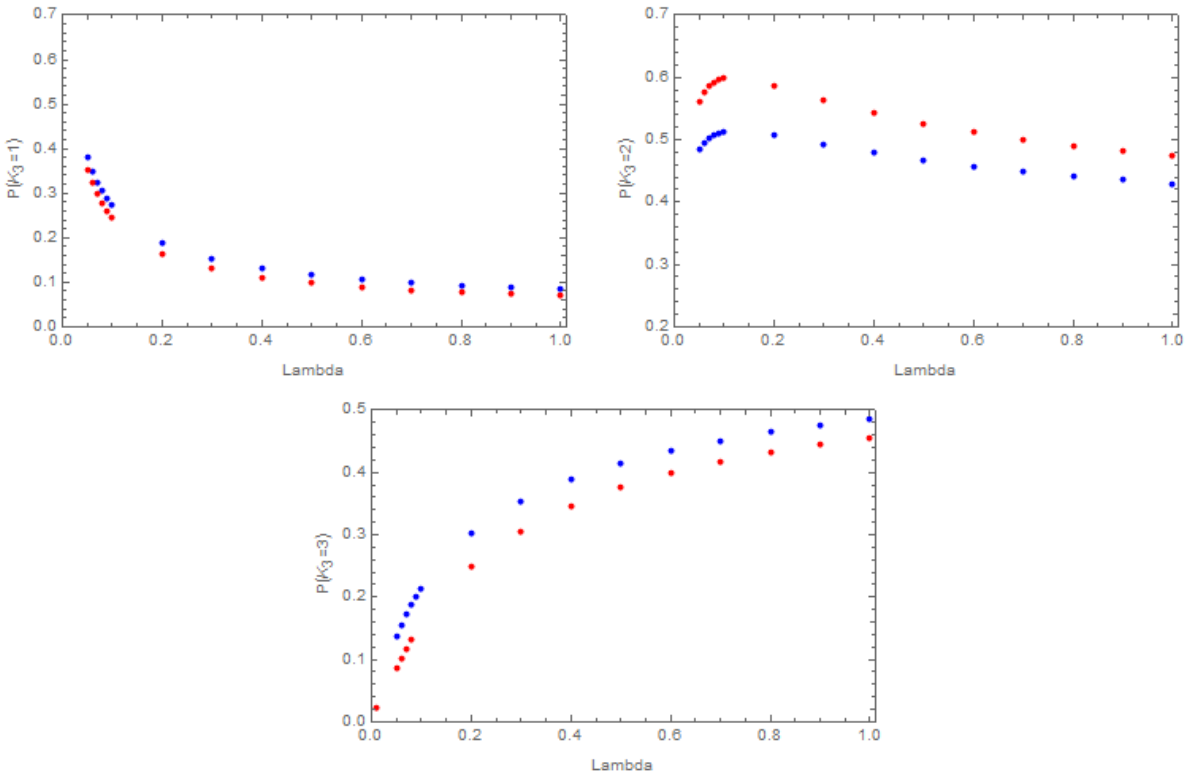

Figure 6.1: *The grafics at the top on the left and on the right show the probability for K_3 to take the values one and two respectively, below the results for $P(K_3 = 3)$ are shown. The red dots represent the spatial model, the results of the coalescent with killings are illustrated by blue dots. The values $N = 100000$ and $M = 10$ were used as parameters.*

Apparently, the differences between the two models are rather small for $M = 10$ according to Fig. 6.1. For both models the probability to draw all individuals from the same cluster tends to one as λ tends to zero since the number of clusters tends to infinity. This probability is slightly smaller for the spatial model, thus, as $P(K_2 = 1)$ is the same for both models, it is more likely for the coalescent with killings that one draws a third individual from a certain cluster after already having drawn two other individuals from this cluster. On the other hand, the probability that all sampled individuals belong to distinct clusters is likewise larger for the coalescent with killings. Consequently, it is more likely to draw three individuals from two different clusters in our spatial model than for the coalescent with killings.

In order to explain what we see, let us have a look at the following approach. We define p_i as the proportion of all individuals of cluster i in the total population. Note that the total number if clusters, which we denote as i_{max}, as well as the proportion p_i vary among replicates.
As the number of individuals of any cluster is large, we can assume that we draw random samples with replacement. Then, the probability across replicates that we sample two individuals of the same cluster becomes

$$P(K_2 = 1) = \langle \sum_{i=1}^{i_{max}} p_i^2 \rangle, \tag{6.20}$$

where $\langle ... \rangle$ denotes the average across replicates.
Similarly, one obtains

$$P(K_3 = 1) = \langle \sum_{i=1}^{i_{max}} p_i^3 \rangle \tag{6.21}$$

and

$$P(K_3 = 2) = 3 \cdot \langle \sum_{\substack{i,j=1,\\ j \neq i}}^{i_{max}} p_i^2 p_j \rangle, \tag{6.22}$$

where the factor 3 appears because either the first, the second or the third sampled individual belongs to cluster j.
It always holds

$$\begin{aligned} \sum_{\substack{i,j=1,\\ j \neq i}}^{i_{max}} p_i^2 p_j &= \sum_{i,j=1}^{i_{max}} p_i^2 p_j - \sum_{i}^{i_{max}} p_i^3 \\ &= \sum_{i=1}^{i_{max}} p_i^2 - \sum_{i=1}^{i_{max}} p_i^3 \end{aligned} \tag{6.23}$$

for any replicate and thus also for the average over replicates. Hence, it yields

$$P(K_3 = 2) = 3 \cdot \left(P(K_2 = 1) - P(K_3 = 1) \right). \tag{6.24}$$

All these equations hold for both, for the spatial model as well as for the coalescent with killings.
Now, we set $P(K_2 = 1)$ as well as $P(K_2 = 2)$ equal for both models. Then, we find that

$$\Delta_1 := P(K_3 = 1) - P^*(K_3 = 1) > 0 \tag{6.25}$$

according to Fig. 6.1. Again, P^* denotes the corresponding probabilities for the spatial model.

With Δ_1, we are now able to explain the differences that we see for the cases $K_3 = 2$ and $K_3 = 3$. By (6.24), it immediately follows

$$\begin{aligned}\Delta_2 &:= P(K_3 = 2) - P^*(K_3 = 2) \\ &= 3\left(P(K_2 = 1) - P(K_3 = 1)\right) - 3\left(P^*(K_2 = 1) - P^*(K_3 = 1)\right) \\ &= 3\left(P^*(K_3 = 1) - P(K_3 = 1)\right) \\ &= -3 \cdot \Delta_1 \\ &< 0\end{aligned} \tag{6.26}$$

and

$$\begin{aligned}\Delta_3 &:= P(K_3 = 3) - P^*(K_3 = 3) \\ &= (1 - P(K_3 = 1) - P(K_3 = 2)) - (1 - P^*(K_3 = 1) - P^*(K_3 = 2)) \\ &= 2\left(P(K_3 = 1) - P^*(K_3 = 1)\right) \\ &= 2 \cdot \Delta_1 \\ &> 0.\end{aligned} \tag{6.27}$$

Thus, if we know the difference Δ_1 between the two models, we are able to explain the sign as well as the absolute values of the differences Δ_2 and Δ_3 by the relations given above.

Appendices

A Calculations related to the finite, circular one-dimensional model

Some of the probabilites that we obtained in Chapter 2, namely for the cases $k = 4$ and $k = 5$, lead to more complex calculations or even require a distinction of cases. In this appendix we show the explicit calculations in detail.

A.1 The case $k = 4$

In the case of four demes we obtained in the context of three and two mutations the probabilities $P(T_2 < 1, T_3 < T_2 + 1)$ and $P(T_2 < 1, T_3 > T_2 + 1)$ respectively. We calculate the first expression explicitly.

$$\begin{aligned} P(T_2 < 1, T_3 < T_2 + 1) &= \int_0^1 \int_0^{t_2+1} \lambda e^{-\lambda t_2} \lambda e^{-\lambda t_3} \mathrm{d}t_3 \mathrm{d}t_2 \\ &= \int_0^1 \lambda e^{-\lambda t_2} \left(1 - e^{-\lambda(t_2+1)}\right) \mathrm{d}t_2 \\ &= \int_0^1 \lambda e^{-\lambda t_2} dt_2 - e^{-\lambda} \int_0^1 \lambda e^{-2\lambda t_2} \mathrm{d}t_2 \\ &= 1 - \frac{3}{2} e^{-\lambda} + \frac{1}{2} e^{-3\lambda}. \end{aligned} \tag{A.1}$$

In a similar way we get

$$P(T_2 < 1, T_3 > T_2 + 1) = \frac{1}{2} e^{-\lambda} \left(1 - e^{-2\lambda}\right). \tag{A.2}$$

In the context of three and four mutations we were confronted with a distinction of two cases. We start by calculating the probability $P(T_2, T_4 < 1; T_3 < \min(T_2, T_4) + 1)$ where we need to distinguish the cases $T_2 < T_4$ and $T_2 > T_4$.

$$\begin{aligned} P(T_2, T_4 < 1; T_3 < \min(T_2, T_4) + 1) &= \int_0^1 \int_0^1 \int_0^{\min(t_2,t_4)+1} \lambda e^{-\lambda t_2} \lambda e^{-\lambda t_4} \lambda e^{-\lambda t_3} \mathrm{d}t_3 \mathrm{d}t_4 \mathrm{d}t_2 \\ &= \int_0^1 \int_0^1 \lambda e^{-\lambda t_2} \lambda e^{-\lambda t_4} \left(1 - e^{-\lambda(\min(t_2,t_4)+1)}\right) \mathrm{d}t_4 \mathrm{d}t_2. \end{aligned} \tag{A.3}$$

For the case $t_2 \leq t_4$ this becomes

$$\begin{aligned}
&\int_0^1 \int_0^{t_4} \lambda e^{-\lambda t_4} \lambda e^{-\lambda t_2} \left(1 - e^{-\lambda(t_2+1)}\right) \mathrm{d}t_2 \mathrm{d}t_4 \\
&= \int_0^1 \lambda e^{-\lambda t_4} \left(\int_0^{t_4} \lambda e^{-\lambda t_2} - e^{-\lambda} \lambda e^{-2\lambda t_2} dt_2 \right) \mathrm{d}t_4 \\
&= \int_0^1 \left[\lambda e^{-\lambda t_4} - \frac{1}{2} e^{-\lambda} \lambda e^{-\lambda t_4} - \lambda e^{-2\lambda t_4} + \frac{1}{2} e^{-\lambda} \lambda e^{-3\lambda t_4} \right] \mathrm{d}t_4 \\
&= \frac{1}{2} - \frac{4}{3} e^{-\lambda} + e^{-2\lambda} - \frac{1}{6} e^{-4\lambda}.
\end{aligned} \tag{A.4}$$

By symmetry, we obtain the same result for $t_2 \geq t_4$, so that we end up with

$$P(T_2, T_4 < 1; T_3 < \min(T_2, T_4) + 1) = 1 - \frac{8}{3} e^{-\lambda} + 2e^{-2\lambda} - \frac{1}{3} e^{-4\lambda}. \tag{A.5}$$

An analogous calculation leads to

$$P(T_2, T_4 < 1; T_3 > \min(T_2, T_4) + 1) = \frac{2}{3} e^{-\lambda} - e^{-2\lambda} + \frac{1}{3} e^{-4\lambda}. \tag{A.6}$$

A.2 The case $k = 5$

Now, we calculate the probabilities of the case $k = 5$ for each possible number of mutations in more detail.

2) Two mutations:

$$\begin{aligned}
P(T_2 < 1; T_3 > T_2 + 1) P(T_4 > 2) P(T_5 > 1) &= e^{-3\lambda} \int_0^1 \int_{t_2+1}^{\infty} \lambda e^{-\lambda t_2} \lambda e^{-\lambda t_3} \mathrm{d}t_3 \mathrm{d}t_2 \\
&= \frac{1}{2} e^{-4\lambda} \left(1 - e^{-2\lambda}\right)
\end{aligned} \tag{A.7}$$

and

$$\begin{aligned}
&P(T_2 > 1)^2 P(T_3 < 2; T_4 > \min(2, T_3 + 1)) \\
&= e^{-2\lambda} \int_0^2 \int_{\min(2, t_3+1)}^{\infty} \lambda e^{-\lambda t_3} \lambda e^{-\lambda t_4} \mathrm{d}t_4 \mathrm{d}t_3 \\
&= \begin{cases} e^{-2\lambda} \int_0^1 \lambda e^{-\lambda t_3} \int_{t_3+1}^{\infty} \lambda e^{-\lambda t_4} \mathrm{d}t_4 \mathrm{d}t_3 & \text{for } t_3 \leq 1 \\ e^{-2\lambda} \int_1^2 \lambda e^{-\lambda t_3} \int_2^{\infty} \lambda e^{-\lambda t_4} dt_4 dt_3 & \text{for } t_3 > 1 \end{cases} \\
&= \begin{cases} \frac{1}{2} e^{-3\lambda} \left(1 - e^{-2\lambda}\right) & \text{for } t_3 \leq 1 \\ e^{-4\lambda} \left(e^{-\lambda} - e^{-2\lambda}\right) & \text{for } t_3 > 1. \end{cases}
\end{aligned} \tag{A.8}$$

Therefore, (2.14) becomes

$$e^{-3\lambda}\left(1-e^{-2\lambda}\right)\left(1+e^{-\lambda}\right)+2e^{-5\lambda}\left(1-e^{-\lambda}\right). \tag{A.9}$$

3) Three mutations:

$$\begin{aligned} P(T_2<1;T_3>T_2+1)P(T_5<1;T_4>T_5+1) &= \left(\int_0^1\int_{t_2+1}^{\infty}\lambda e^{-\lambda t_2}\lambda e^{-\lambda t_3}\mathrm{d}t_3\mathrm{d}t_2\right)^2 \\ &= \frac{1}{4}e^{-2\lambda}\left(1-e^{-2\lambda}\right)^2 \end{aligned} \tag{A.10}$$

and

$$\begin{aligned} &P(T_2<1;T_3<T_2+1;T_4>\min(2,T_3+1))P(T_5>1) \\ &= e^{-\lambda}\int_0^1\int_0^{t_2+1}\int_{\min(1,t_3)+1}^{\infty}\lambda e^{-\lambda t_2}\lambda e^{-\lambda t_3}\lambda e^{-\lambda t_4}\mathrm{d}t_4\mathrm{d}t_3\mathrm{d}t_2 \\ &= \begin{cases} e^{-\lambda}\int_0^1\int_0^1\int_{t_3+1}^{\infty}\lambda e^{-\lambda t_2}\lambda e^{-\lambda t_3}\lambda e^{-\lambda t_4}\mathrm{d}t_4\mathrm{d}t_3\mathrm{d}t_2 & \text{for } t_3\le 1 \\ e^{-\lambda}\int_0^1\int_1^{t_2+1}\int_2^{\infty}\lambda e^{-\lambda t_2}\lambda e^{-\lambda t_3}\lambda e^{-\lambda t_4}\mathrm{d}t_4\mathrm{d}t_3\mathrm{d}t_2 & \text{for } t_3>1 \end{cases} \\ &= \begin{cases} \frac{1}{2}e^{-2\lambda}\left(1-e^{-2\lambda}\right)\left(1-e^{-\lambda}\right) & \text{for } t_3\le 1 \\ e^{-4\lambda}\left(\frac{1}{2}-e^{-\lambda}+\frac{1}{2}e^{-2\lambda}\right) & \text{for } t_3>1 \end{cases} \end{aligned} \tag{A.11}$$

and

$$\begin{aligned} &P(T_2<1;T_4<2;T_3>\min(T_2,T_4)+1)P(T_5>1) \\ &= e^{-\lambda}\int_0^1\int_0^2\int_{\min(t_2,t_4)+1}^{\infty}\lambda e^{-\lambda t_2}\lambda e^{-\lambda t_4}\lambda e^{-\lambda t_3}\mathrm{d}t_3\mathrm{d}t_4\mathrm{d}t_2 \\ &= \begin{cases} e^{-\lambda}\int_0^1\int_0^{t_2}\int_{t_4+1}^{\infty}\lambda e^{-\lambda t_2}\lambda e^{-\lambda t_4}\lambda e^{-\lambda t_3}\mathrm{d}t_3\mathrm{d}t_4\mathrm{d}t_2 & \text{for } t_4\le t_2 \\ e^{-\lambda}\int_0^1\int_{t_2}^2\int_{t_2+1}^{\infty}\lambda e^{-\lambda t_2}\lambda e^{-\lambda t_4}\lambda e^{-\lambda t_3}\mathrm{d}t_3\mathrm{d}t_4\mathrm{d}t_2 & \text{for } t_4>t_2 \end{cases} \\ &= \begin{cases} \frac{1}{2}e^{-2\lambda}\left(\frac{2}{3}-e^{-\lambda}+\frac{1}{3}e^{-3\lambda}\right) & \text{for } t_4\le t_2 \\ e^{-2\lambda}\left(\frac{1}{3}-\frac{1}{2}e^{-2\lambda}-\frac{1}{3}e^{-3\lambda}+\frac{1}{2}e^{-4\lambda}\right) & \text{for } t_4>t_2 \end{cases} \end{aligned} \tag{A.12}$$

and

$$
\begin{aligned}
&P(T_2 > 1)^2 P(T_3 < \min(2, T_4+1); T_4 < \min(2, T_3+1)) \\
&= e^{-2\lambda} \int_0^{\min(2,t_4+1)} \int_0^{\min(2,t_3+1)} \lambda e^{-\lambda t_3} e^{-\lambda t_4} \mathrm{d}t_4 \mathrm{d}t_3 \\
&= \begin{cases}
e^{-2\lambda} \left(\int_1^2 \lambda e^{-\lambda t_3} \mathrm{d}t_3 \right)^2 & \text{for } t_3, t_4 \geq 1 \\
e^{-2\lambda} \left(\int_0^1 \lambda e^{-\lambda t_3} \mathrm{d}t_3 \right)^2 & \text{for } t_3, t_4 < 1 \\
e^{-2\lambda} \int_0^1 \int_1^{t_3+1} \lambda e^{-\lambda t_3} \lambda e^{-\lambda t_4} \mathrm{d}t_4 \mathrm{d}t_3 & \text{for } t_3 \leq 1 \leq t_4 \\
e^{-2\lambda} \int_0^1 \int_1^{t_4+1} \lambda e^{-\lambda t_3} \lambda e^{-t_4} \mathrm{d}t_3 \mathrm{d}t_4 & \text{for } t_4 \leq 1 \leq t_3
\end{cases} \\
&= \begin{cases}
e^{-4\lambda} \left(1 - e^{-\lambda}\right)^2 & \text{for } t_3, t_4 \geq 1 \\
e^{-2\lambda} \left(1 - e^{-\lambda}\right)^2 & \text{for } t_3, t_4 < 1 \\
e^{-3\lambda} \left(\frac{1}{2} - e^{-\lambda} + \frac{1}{2} e^{-2\lambda}\right) & \text{for } t_3 \leq 1 \leq t_4 \text{ and } t_4 \leq 1 \leq t_3.
\end{cases}
\end{aligned}
\tag{A.13}
$$

By inserting these results into (2.15), it yields for the probability that exactly three independent spontaneous mutations occur

$$
\frac{43}{12} e^{-2\lambda} - \frac{7}{2} e^{-3\lambda} - \frac{1}{2} e^{-4\lambda} - \frac{17}{6} e^{-5\lambda} + \frac{13}{4} e^{-6\lambda}. \tag{A.14}
$$

4) Four mutations:

$$
\begin{aligned}
&P(T_2, T_5 < 1; T_3 < T_2 + 1; T_4 > \min(T_3, T_5) + 1) \\
&= \int_0^1 \int_0^1 \int_0^{t_2+1} \int_{\min(t_3,t_5)+1}^{\infty} \lambda e^{-\lambda t_2} \lambda e^{-\lambda t_5} \lambda e^{-\lambda t_3} \lambda e^{-\lambda t_4} \mathrm{d}t_4 \mathrm{d}t_3 \mathrm{d}t_5 \mathrm{d}t_2 \\
&= \begin{cases}
\int_0^1 \int_0^1 \int_0^{t_5} \int_{t_3+1}^{\infty} \lambda e^{-\lambda t_2} \lambda e^{-\lambda t_5} \lambda e^{-\lambda t_3} \lambda e^{-\lambda t_4} \mathrm{d}t_4 \mathrm{d}t_3 \mathrm{d}t_5 \mathrm{d}t_2 & \text{for } t_3 \leq t_5 \\
\int_0^1 \int_0^1 \int_{t_5}^{t_2+1} \int_{t_5+1}^{\infty} \lambda e^{-\lambda t_2} \lambda e^{-\lambda t_5} \lambda e^{-\lambda t_3} \lambda e^{-\lambda t_4} \mathrm{d}t_4 \mathrm{d}t_3 \mathrm{d}t_5 \mathrm{d}t_2 & \text{for } t_3 > t_5
\end{cases} \\
&= \begin{cases}
\frac{1}{2} e^{-\lambda} \left(1 - e^{-\lambda}\right) \left(\frac{2}{3} - e^{-\lambda} + \frac{1}{3} e^{-3\lambda}\right) & \text{for } t_3 \leq t_5 \\
e^{-\lambda} \left(\frac{1}{3} \left(1 - e^{-\lambda}\right) \left(1 - e^{-3\lambda}\right) - \frac{1}{4} e^{-\lambda} \left(1 - e^{-2\lambda}\right)^2 \right) & \text{for } t_3 > t_5
\end{cases}
\end{aligned}
\tag{A.15}
$$

and

$$
\begin{aligned}
&P(T_5 > 1)P(T_2 < 1; T_3 < \min(T_2, T_4) + 1; T_4 < \min(2, T_3 + 1)) \\
&= e^{-\lambda} \int_0^1 \int_0^{\min(t_2,t_4)+1} \int_0^{\min(1,t_3)+1} \lambda e^{-\lambda t_2} \lambda e^{-\lambda t_3} \lambda e^{-\lambda t_4} \mathrm{d}t_4 \mathrm{d}t_3 \mathrm{d}t_2 \\
&= \begin{cases} e^{-\lambda} \int_0^1 \int_0^1 \int_0^{t_3+1} \lambda e^{-\lambda t_2} \lambda e^{-\lambda t_3} \lambda e^{-\lambda t_4} \mathrm{d}t_4 \mathrm{d}t_3 \mathrm{d}t_2 & \text{for } t_3 \leq 1 \\ e^{-\lambda} \int_0^1 \int_1^{\min(t_2,t_4)+1} \int_0^2 \lambda e^{-\lambda t_2} \lambda e^{-\lambda t_3} \lambda e^{-\lambda t_4} \mathrm{d}t_4 \mathrm{d}t_3 \mathrm{d}t_2 & \text{for } t_3 > 1 \end{cases} \\
&= \begin{cases} e^{-\lambda} \left(1 - e^{-\lambda}\right) \left(1 - \frac{3}{2} e^{-\lambda} + \frac{1}{2} e^{-3\lambda}\right) & \text{for } t_3 \leq 1 \\ e^{-\lambda} \int_0^1 \int_{t_2}^2 \int_1^{t_2+1} \lambda e^{-\lambda t_2} \lambda e^{-\lambda t_3} \lambda e^{-\lambda t_4} \mathrm{d}t_3 \mathrm{d}t_4 \mathrm{d}t_2 & \text{for } t_3 > 1, t_2 \leq t_4 \\ e^{-\lambda} \int_0^1 \int_0^{t_2} \int_1^{t_4+1} \lambda e^{-\lambda t_2} \lambda e^{-\lambda t_3} \lambda e^{-\lambda t_4} \mathrm{d}t_3 \mathrm{d}t_4 \mathrm{d}t_2 & \text{for } t_3 > 1, t_2 \geq t_4 \end{cases} \\
&= \begin{cases} e^{-\lambda} \left(1 - e^{-\lambda}\right) \left(1 - \frac{3}{2} e^{-\lambda} + \frac{1}{2} e^{-3\lambda}\right) & \text{for } t_3 \leq 1 - p \\ e^{-2\lambda} \left(\frac{1}{6} - e^{-2\lambda} + \frac{4}{3} e^{-3\lambda} - \frac{1}{2} e^{-4\lambda}\right) & \text{for } t_3 > 1, t_2 \leq t_4 \\ \frac{1}{2} e^{-2\lambda} \left(\frac{1}{3} - e^{-\lambda} + e^{-2\lambda} - \frac{1}{3} e^{-3\lambda}\right) & \text{for } t_3 > 1, t_2 \geq t_4. \end{cases}
\end{aligned}
\tag{A.16}
$$

Hence, (2.16) becomes

$$
-\frac{1}{6} \left(-1 + e^{-\lambda}\right)^3 e^{-\lambda} \left(20 + 17 e^{-\lambda} + 9 e^{-2\lambda}\right). \tag{A.17}
$$

5) Five mutations:

$$
\begin{aligned}
&P(T_2, T_5 < 1; T_3 < \min(T_2, T_4) + 1; T_4 < \min(T_3, T_5) + 1) \\
&= \int_0^1 \int_0^1 \int_0^{\min(t_2,t_4)+1} \int_0^{\min(t_3,t_5)+1} \lambda e^{-\lambda t_2} \lambda e^{-\lambda t_5} \lambda e^{-\lambda t_3} \lambda e^{-\lambda t_4} \mathrm{d}t_4 \mathrm{d}t_3 \mathrm{d}t_5 \mathrm{d}t_2 \\
&= \begin{cases} \int_0^1 \int_0^1 \int_0^{t_5} \int_0^{t_3+1} \lambda e^{-\lambda t_2} \lambda e^{-\lambda t_5} \lambda e^{-\lambda t_3} \lambda e^{-\lambda t_4} \mathrm{d}t_4 \mathrm{d}t_3 \mathrm{d}t_5 \mathrm{d}t_2 & \text{for } t_3 \leq t_5 \\ \int_0^1 \int_0^1 \int_{t_5}^{\min(t_2,t_4)+1} \int_0^{t_5+1} \lambda e^{-\lambda t_2} \lambda e^{-\lambda t_5} \lambda e^{-\lambda t_3} \lambda e^{-\lambda t_4} \mathrm{d}t_4 \mathrm{d}t_3 \mathrm{d}t_5 \mathrm{d}t_2 & \text{for } t_3 > t_5 \end{cases} \\
&= \begin{cases} \left(1 - e^{-\lambda}\right) \left(\frac{1}{2} - \frac{4}{3} e^{-\lambda} + e^{-2\lambda} - \frac{1}{6} e^{-4\lambda}\right) & \text{for } t_3 \leq t_5 \\ \int_0^1 \int_0^1 \int_0^{t_2} \int_{t_5}^{t_4+1} \lambda e^{-\lambda t_2} \lambda e^{-\lambda t_5} \lambda e^{-\lambda t_3} \lambda e^{-\lambda t_4} \mathrm{d}t_3 \mathrm{d}t_4 \mathrm{d}t_5 \mathrm{d}t_2 & \text{for } t_3 > t_5, t_2 \geq t_4 \\ \int_0^1 \int_0^1 \int_{t_5}^{t_2+1} \int_{t_2}^{t_5+1} \lambda e^{-\lambda t_2} \lambda e^{-\lambda t_5} \lambda e^{-\lambda t_3} \lambda e^{-\lambda t_4} \mathrm{d}t_4 \mathrm{d}t_3 \mathrm{d}t_5 \mathrm{d}t_2 & \text{for } t_3 > t_5, t_2 < t_4 \end{cases} \\
&= \begin{cases} \left(1 - e^{-\lambda}\right) \left(\frac{1}{2} - \frac{4}{3} e^{-\lambda} + e^{-2\lambda} - \frac{1}{6} e^{-4\lambda}\right) & \text{for } t_3 \leq t_5 \\ \frac{1}{2} \left(1 - e^{-\lambda}\right) \left(\frac{1}{2} - \frac{7}{6} e^{-\lambda} + \frac{1}{2} e^{-2\lambda} + \frac{1}{2} e^{-3\lambda} - \frac{1}{3} e^{-4\lambda}\right) & \text{for } t_3 > t_5, t_2 \geq t_4 \\ -\frac{2}{3} e^{-\lambda} \left(1 - e^{-\lambda}\right) \left(1 - e^{-3\lambda}\right) + \frac{1}{4} \left(1 - e^{-2\lambda}\right)^2 \left(1 + e^{-2\lambda}\right) & \text{for } t_3 > t_5, t_2 < t_4. \end{cases}
\end{aligned}
\tag{A.18}
$$

Therefore, it yields for (2.17)

$$
\frac{1}{12} \left(-1 + e^{-\lambda}\right)^4 \left(12 + 8e^{-\lambda} + 3e^{-2\lambda}\right). \tag{A.19}
$$

B Calculations related to the infinite, linear one-dimensional model

In this Appendix we explicitly calculate the probabilities $P(n_r = k_1, n_l = k_2)$ and $P(n_r = k)$ for the one-dimensional model which we derived in Chapter 3.

B.1 Calculation of $P(n_r = k_1, n_l = k_2)$

The joint probability of $P(n_r = k_1, n_l = k_2)$ for $k_1, k_2 \geq 0$ is given by the intersection of all events defined in Definition **??**. As some of these events are independent under the condition $T_0 = T$ and (3.19) holds, we obtain

$$
\begin{aligned}
&P(n_r = k_1, n_l = k_2) \\
&= \int_0^\infty \lambda e^{-\lambda T} P(n_r = k_1, n_l = k_2 | T_0 = T) \mathrm{d}T \\
&= \int_0^\infty \lambda e^{-\lambda T} P(N_r^1(k_1), N_l^1(k_2), B_r^1(k_1), B_l^1(k_2), C_r^1(k_1), C_l^1(k_2) | T_0 = T) \mathrm{d}T \\
&= \int_0^\infty \lambda e^{-\lambda T} P(N_r^1(k_1) | T_0 = T) \Big[P(C_r^1(k_1) | T_0 = T) + P(B_r^1(k_1) | T_0 = T) - 1 \Big] \\
&\qquad \cdot P(N_l^1(k_2) | T_0 = T) \Big[P(C_l^1(k_2) | T_0 = T) + P(B_l^1(k_2) | T_0 = T) - 1 \Big] \mathrm{d}T \\
&= \sum_{j=0}^\infty \int_j^{j+1} \lambda e^{-\lambda T (1 + 2(j + k_1 + k_2))} e^{-\lambda(-j(j+1) + k_1^2 + k_2^2)} \Big(1 - e^{-2\lambda T} e^{-\lambda(2k_1 + 3)} \Big) \\
&\qquad \cdot \Big(1 - e^{-2\lambda T} e^{-\lambda(2k_2 + 3)} \Big) \mathrm{d}T
\end{aligned}
$$

$$
\begin{aligned}
&= \sum_{j=0}^{\infty} \Big[\frac{1}{1+2(j+k_1+k_2)} e^{-\lambda(-j^2-j+k_1^2+k_2^2)} \Big(e^{-\lambda(1+2(j+k_1+k_2))j} - e^{-\lambda(1+2(j+k_1+k_2))(j+1)} \Big) \\
&- \frac{1}{3+2(j+k_1+k_2)} e^{-\lambda(-j^2-j+k_1^2+k_2^2+2k_1+3)} \Big(e^{-\lambda(3+2(j+k_1+k_2))j} - e^{-\lambda(3+2(j+k_1+k_2))(j+1)} \Big) \\
&- \frac{1}{3+2(j+k_1+k_2)} e^{-\lambda(-j^2-j+k_1^2+k_2^2+2k_2+3)} \Big(e^{-\lambda(3+2(j+k_1+k_2))j} - e^{-\lambda(3+2(j+k_1+k_2))(j+1)} \Big) \\
&+ \frac{1}{5+2(j+k_1+k_2)} e^{-\lambda(-j^2-j+k_1^2+2k_1+k_2^2+2k_2+6)} \Big(e^{-\lambda(5+2(j+k_1+k_2))j} - e^{-\lambda(5+2(j+k_1+k_2))(j+1)} \Big) \Big] \\
&= \sum_{j=0}^{\infty} \Big[\frac{1}{1+2(j+k_1+k_2)} e^{-\lambda(j^2+k_1^2+k_2^2+2j(k_1+k_2))} \Big(1 - e^{-\lambda(1+2(j+k_1+k_2))} \Big) \\
&- \frac{1}{3+2(j+k_1+k_2)} e^{-\lambda(j^2+2j+k_1^2+k_2^2+3+2j(k_1+k_2))} \Big(e^{-2\lambda k_1} + e^{-2\lambda k_2} \Big) \Big(1 - e^{-\lambda(3+2(j+k_1+k_2))} \Big) \\
&+ \frac{1}{5+2(j+k_1+k_2)} e^{-\lambda(j^2+4j+k_1^2+k_2^2+6+2(k_1+k_2)(j+1))} \Big(1 - e^{-\lambda(5+2(j+k_1+k_2))} \Big) \Big]. \qquad \text{(B.1)}
\end{aligned}
$$

B.2 Calculation of $P(n_r = k)$

The probability $P(n_r = k)$ can be calculated in a similar way by using exactly the same arguments. It yields

$$\begin{aligned}
P(n_r = k) &= \int_0^\infty \lambda e^{-\lambda T} P(n_r = k | T_0 = T) \mathrm{d}T \\
&= \int_0^\infty \lambda e^{-\lambda T} P(C_l^1(0), B_r^1(k), C_r^1(k), N_r(k)^1 | T_0 = T) \mathrm{d}T \\
&= \int_0^\infty \lambda e^{-\lambda T} P(C_l^1(0) | T_0 = T) P(N_r^1(k) | T_0 = T) \\
&\quad \cdot \Big[P(C_r^1(k) | T_0 = T) + P(B_r^1(k) | T_0 = T) - 1 \Big] \mathrm{d}T \\
&= \sum_{j=0}^\infty \int_j^{j+1} \lambda e^{-\lambda T (1+2(j+k))} e^{-\lambda(-j(j+1)+k^2)} \left(1 - e^{-2\lambda T} e^{-\lambda(2k+3)}\right) \mathrm{d}T \\
&= \sum_{j=0}^\infty \Bigg[\frac{1}{1+2(j+k)} e^{-\lambda(-j(j+1)+k^2)} \left(e^{-\lambda(1+2(j+k))j} - e^{-\lambda(1+2(j+k))(j+1)} \right) \\
&\quad - \frac{1}{3+2(j+k)} e^{-\lambda(-j(j+1)+k^2+2k+3)} \left(e^{-\lambda(3+2(j+k))j} - e^{-\lambda(3+2(j+k))(j+1)} \right) \Bigg] \\
&= \sum_{j=0}^\infty \Bigg[\frac{1}{1+2(j+k)} e^{-\lambda(j+k)^2} \left(1 - e^{-\lambda(1+2(j+k))}\right) \\
&\quad - \frac{1}{3+2(j+k)} e^{-\lambda(j^2+2j+k^2+2k+2kj)} \left(1 - e^{-\lambda(3+2(j+k))}\right) \Bigg] .
\end{aligned} \tag{B.2}$$

C Calculations related to the two-dimensional model

In this section we calculate in detail the conditioned probability for each event defined in Definition 4.1 given $T_0 = T$ for a fixed time $T \in \mathbb{R}_{\geq 0}$. All definitions and notations from Chapter 3 will be kept throughout this section.
We start with the probabilities for $B_r^2(k_1)$ and $B_l^2(k_2)$. For the first one we obtain

$$\begin{aligned}
&P(B_r^2(k_1)|T_0 = T) \\
&= P\left(T + k_1 + 1 > \min\left(T_{(k_1+v_1,\pm v_2)} + v - 1 | v_1 \geq 1, v_2 \geq 0, 1 \leq v \leq d_T + k_1 + 2\right)\right) \\
&= 1 - P\left(T + k_1 + 1 < \min\left(T_{(k_1+v_1,\pm v_2)} + v - 1 | v_1 \geq 1, v_2 \geq 0, 1 \leq v \leq d_T + k_1 + 2\right)\right) \\
&= 1 - \prod_{v=1}^{d_T+k_1+2} \prod_{v_1=1}^{v} P(T_{(k_1+v_1,\pm(v-v_1))} > T + k_1 + 2 - v) \\
&= 1 - \prod_{v=1}^{d_T+k_1+2} P(T_0 > T + k_1 + 2 - v)^{2v-1} \\
&= 1 - \prod_{v=1}^{d_T+k_1+2} e^{-\lambda(T+k_1+2-v)(2v-1)},
\end{aligned} \tag{C.1}$$

where the exponent $\frac{f_2(v)-2}{2} = 2v - 1$ corresponds to the number of v-step neighbors of deme $(k_1, 0)$ that are located to the right side of deme $(k_1, 0)$, which means that they are of the form (x, y) with $x > k_1$.
By symmetry, it immediately follows

$$P(B_l^2(k_2)|T_0 = T) = P(B_r^2(k_2)|T_0 = T). \tag{C.2}$$

Similarly, we get for $C_r^2(k_1)$ and $C_l^2(k_2)$:

$$\begin{aligned} P(C_r^2(k_1)|T_0 = T) &= P\left(T + k_1 < \min\left(T_{(k_1+v_1,\pm v_2)} + v | v_1 \geq 1, v_2 \geq 0, 1 \leq v \leq d_T + k_1\right)\right) \\ &= \prod_{v=1}^{d_T+k_1} \prod_{v_1=1}^{v} P(T_{(k_1+v_1,\pm(v-v_1))} > T + k_1 - v) \\ &= \prod_{v=1}^{d_T+k_1} P(T_0 > T + k_1 - v)^{2v-1} \\ &= \prod_{v=1}^{d_T+k_1} e^{-\lambda(T+k_1-v)(2v-1)} \end{aligned} \tag{C.3}$$

and by symmetry,

$$P(C_l^2(k_2)|T_0 = T) = P(C_r^2(k_2)|T_0 = T). \tag{C.4}$$

By making use of the same argument as in Chapter 3, we again set $P(N_r^2(k_1)|T_0 = T) = 1$ and $P(N_l^2(k_2)|T_0 = T) = 1$ for $k_1 = 0$ and $k_2 = 0$ respectively. For $k_1 > 0$ and $k_2 > 0$ we obtain

$$\begin{aligned} P(N_r^2(k_1)|T_0 = T) = P\big(&T + 1 < \min\left(T_{(1,0)}, T_{(1,\pm1)} + 1, ..., T_{(1,\pm(d_T+1))} + d_T + 1\right), \\ &T + 2 < \min\left(T_{(2,0)}, T_{(2,\pm1)} + 1, ..., T_{(2,\pm(d_T+2))} + d_T + 2\right), \\ &..., \\ &T + k_1 < \min\left(T_{(k_1,0)}, T_{(k_1,\pm1)} + 1, .., T_{(k_1,\pm(d_T+k_1))} + d_T + k_1\right)\big) \\ = \prod_{v_1=1}^{k_1} P\left(T + v_1 < \min\left(T_{(v_1,0)}, .., T_{(v_1,\pm(d_T+v_1))} + d_T + v_1\right)\right) \\ = \prod_{v_1=1}^{k_1} P(T_{(v_1,0)} > T + v_1) \prod_{v_2=1}^{d_T+v_1} P(T_{(v_1,\pm v_2)} > T + v_1 - v_2) \\ = \prod_{v_1=1}^{k_1} P(T_0 > T + v_1) \prod_{v_2=1}^{d_T+v_1} P(T_0 > T + v_1 - v_2)^2 \\ = \prod_{v_1=1}^{k_1} e^{-\lambda(T+v_1)} \prod_{v_2=1}^{d_T+v_1} e^{-2\lambda(T+v_1-v_2)} \end{aligned} \tag{C.5}$$

and by symmetry,

$$P(N_l^2(k_2)|T_0 = T) = P(N_r^2(k_2)|T_0 = T). \tag{C.6}$$

Finally, the probability for N_0^2 conditioned on $T_0 = T$ is given by

$$\begin{aligned} P(N_0^2|T_0 = T) &= P(T < \min\left(T_{(0,\pm v_2)} + v_2 | 1 \leq v_2 \leq d_T\right)) \\ &= \prod_{v_2=1}^{d_T} P(T_{(0,\pm v_2)} > T - v_2) \\ &= \prod_{v_2=1}^{d_T} P(T_0 > T - v_2)^2 \\ &= \prod_{v_2=1}^{d_T} e^{-2\lambda(T-v_2)}. \end{aligned} \tag{C.7}$$

D Calculations related to the three-dimensional model

In this section we show the detailed calculations for the conditioned probability of each event of Definition 5.3 given $T_0 = T$ for a fixed time $T \in \mathbb{R}_{\geq 0}$. The underlying idea for the definition of these events is the same as in the two-dimensional model and so are the calculations. The main difference between the calculations in two and three dimensions is the number of v-step neighbors that has to be taken into account.
We will keep all definitions and notations from Chapter 4 throughout this section.
Let us start with the probabilities for $B_r^3(k_1)$ and $B_l^3(k_2)$. For the first one we obtain

$$\begin{aligned}
&P(B_r^3(k_1)|T_0 = T) \\
&= P(T + k_1 + 1 > \min(T_{(k_1+v_1,\pm v_2,\pm v_3)} + v - 1 | v_1 \geq 1, v_2, v_3 \geq 0, 1 \leq v \leq d_T + k_1 + 2)) \\
&= 1 - P(T + k_1 + 2 < \min(T_{(k_1+v_1,\pm v_2,\pm v_3)} + v | v_1 \geq 1, v_2, v_3 \geq 0, 1 \leq v \leq d_T + k_1 + 2)) \\
&= 1 - \prod_{v=1}^{d_T+k_1+2} \prod_{v_1=1}^{v} \prod_{v_2=0}^{v-v_1} P(T_{(k_1+v_1,\pm v_2,\pm(v-v_1-v_2))} > T + k_1 + 2 - v) \\
&= 1 - \prod_{v=1}^{d_T+k_1+2} P(T_0 > T + k_1 + 2 - v)^{\frac{f_3(v)-f_2(v)}{2}} \\
&= 1 - \prod_{v=1}^{d_T+k_1+2} e^{-\lambda(T+k_1+2-v)(\frac{f_3(v)-f_2(v)}{2})},
\end{aligned} \tag{D.1}$$

where the exponent $\frac{f_3(v)-f_2(v)}{2}$ corresponds to the number of v-step neighbors to the right side of deme $(k_1, 0, 0)$, exactly as in the two-dimensional case.
Again, it immediately follows for $B_r^3(k_2)$ by symmetry:

$$P(B_l^3(k_2)|T_0 = T) = P(B_r^3(k_2)|T_0 = T). \tag{D.2}$$

Similarly, we obtain for $C_r^3(k_1)$ and $C_l^3(k_2)$

$$\begin{aligned}
&P(C_r^3(k_1)|T_0 = T) \\
&= P(T + k_1 < \min\left(T_{(k_1+v_1,\pm v_2,\pm v_3)} + v | v_1 \geq 1, v_2, v_3 \geq 0, 1 \leq v \leq d_T + k_1\right)) \\
&= \prod_{v=1}^{d_T+k_1} \prod_{v_1=1}^{v} \prod_{v_2=0}^{v-v_1} P(T_{(k_1+v_1,\pm v_2,\pm(v-v_1-v_2))} > T + k_1 - v) \\
&= \prod_{v=1}^{d_T+k_1} P(T_0 > T + k_1 - v)^{\frac{f_3(v)-f_2(v)}{2}} \\
&= \prod_{v=1}^{d_T+k_1} e^{-\lambda(T+k_1-v)(\frac{f_3(v)-f_2(v)}{2})}
\end{aligned} \tag{D.3}$$

as well as

$$P(C_l^3(k_2)|T_0 = T) = P(C_r^3(k_2)|T_0 = T). \tag{D.4}$$

As for the one- and the two-dimensional model we can again set $P(N_r^3(k_1)|T_0 = T) = 1$ and $P(N_l^3(k_2)|T_0 = T) = 1$ for $k_1 = 0$ and $k_2 = 0$ respectively. For $k_1 > 0$ and $k_2 > 0$ we obtain

$$\begin{aligned}
&P(N_r^3(k_1)|T_0 = T) \\
&= P\big(T + 1 < \min\left(T_{(1,\pm v_2,\pm v_3)} + v_2 + v_3 | v_2, v_3 \geq 0, 0 \leq v_2 + v_3 \leq d_T + 1\right), \\
&\qquad T + 2 < \min\left(T_{(2,\pm v_2,\pm v_3)} + v_2 + v_3 | v_2, v_3 \geq 0, 0 \leq v_2 + v_3 \leq d_T + 2\right), \\
&\qquad ..., \\
&\qquad T + k_1 < \min\left(T_{(k_1,\pm v_2,\pm v_3)} + v_2 + v_3 | v_2, v_3 \geq 0, 0 \leq v_2 + v_3 \leq d_T + k_1\right)\big) \\
&= \prod_{v_1=1}^{k_1} P\left(T + v_1 < \min\left(T_{(v_1,\pm v_2,\pm v_3)} + v_2 + v_3 | v_2, v_3 \geq 0, 0 \leq v_2 + v_3 \leq d_T + v_1\right)\right) \\
&= \prod_{v_1=1}^{k_1} P(T_{(v_1,0,0)} > T + v_1) \prod_{\substack{v_2+v_3 \geq 1, \\ v_2,v_3 \geq 0}}^{d_T+v_1} P(T_{(v_1,\pm v_2,\pm v_3)} > T + v_1 - v_2 - v_3) \\
&= \prod_{v_1=1}^{k_1} P(T_0 > T + v_1) \prod_{\substack{v_2+v_3 \geq 1, \\ v_2,v_3 \geq 0}}^{d_T+v_1} P(T_0 > T + v_1 - v_2 - v_3)^{f_2(v_2+v_3)} \\
&= \prod_{v_1=1}^{k_1} e^{-\lambda(T+v_1)} \prod_{\substack{v_2+v_3 \geq 1, \\ v_2,v_3 \geq 0}}^{d_T+v_1} e^{-\lambda(T+v_1-v_2-v_3) f_2(v_2+v_3)}
\end{aligned} \tag{D.5}$$

and

$$P(N_l^3(k_2)|T_0 = T) = P(N_r^3(k_2)|T_0 = T). \tag{D.6}$$

Finally, the probability for N_0^3 conditioned on $T_0 = T$ is given by

$$\begin{aligned} P(N_0^3|T_0 = T) &= P\left(T < \min\left(T_{(0,\pm v_2,\pm v_3)} + v_2 + v_3 | v_2, v_3 \geq 0, 1 \leq v_2 + v_3 \leq d_T\right)\right) \\ &= \prod_{\substack{v_2+v_3\geq 1,\\ v_2,v_3\geq 0}}^{d_T} P(T_{(0,\pm v_2,\pm v_3)} > T - (v_2 + v_3)) \\ &= \prod_{\substack{v_2+v_3\geq 1,\\ v_2,v_3\geq 0}}^{d_T} P(T_0 > T - (v_2 + v_3))^{f_2(v_2+v_3)} \\ &= \prod_{\substack{v_2+v_3\geq 1,\\ v_2,v_3\geq 0}}^{d_T} e^{-\lambda(T-(v_2+v_3))f_2(v_2+v_3)}. \end{aligned} \tag{D.7}$$

Bibliography

Durrett, R. *Probability models for DNA sequence evolution*. Springer Science & Business Media, 2008.

Evans, U. The laws of expanding circles and spheres in relation to the lateral growth of surface films and the grain-size of metals. *Transactions of the Faraday Society*, 41: 365–374, 1945.

Ewens, W. J. The sampling theory of selectively neutral alleles. *Theoretical population biology*, 3(1):87–112, 1972.

Ewens, W. and Tavaré, S. Ewens sampling formula. *Encyclopedia of Statistical Sciences*, 1998.

Ffrench Constant, R. H., Anthony, N., Aronstein, K., Rocheleau, T., and Stilwell, G. Cyclodiene insecticide resistance: from molecular to population genetics. *Annual review of entomology*, 45(1):449–466, 2000.

Greven, A., Pfaffelhuber, P., Pokalyuk, C., and Wakolbinger, A. The fixation time of a strongly beneficial allele in a structured population. *arXiv preprint arXiv:1402.1769*, 2014.

Griffiths, R. C. and Lessard, S. Ewens' sampling formula and related formulae: combinatorial proofs, extensions to variable population size and applications to ages of alleles. *Theoretical population biology*, 68(3):167–177, 2005.

Hartl, D. L., Clark, A. G., and Clark, A. G. *Principles of population genetics*, volume 116. Sinauer associates Sunderland, 1997.

Jeong, S., Rebeiz, M., Andolfatto, P., Werner, T., True, J., and Carroll, S. B. The evolution of gene regulation underlies a morphological difference between two drosophila sister species. *Cell*, 132(5):783–793, 2008.

Johnson, N. L., Kotz, S., and Balakrishnan, N. *Discrete multivariate distributions*, volume 165. Wiley New York, 1997.

Kimura, M. and Crow, J. F. The number of alleles that can be maintained in a finite population. *Genetics*, 49(4):725, 1964.

Kingman, J. F. C. *Mathematics of genetic diversity*, volume 34. SIAM, 1980.

Ralph, P. and Coop, G. Parallel adaptation: one or many waves of advance of an advantageous allele? *Genetics*, 186(2):647–668, 2010.

Zusammenfassung

Modelle, die die Anpassung einer Population an eine neue Umwelt beschreiben, basieren oft auf dem Auftreten einer einzelnen vorteilhaften Neumutation, welche dann in der Population fixiert wird. Doch es sind auch zahlreiche Beispiele bekannt, bei denen mehrere Mutationen mit ähnlichem phänotypischen Effekt eine parallele Anpassung ermöglichen. Es gibt bereits unterschiedliche Modelle, die einen kontinuierlichen Raum mit einem globalen Selektionsdruck annehmen, sowie auch Modelle mit einer räumlich strukturierten Population, die einer lokalen Selektion unterliegen. In beiden Fällen führt eine eingeschränkte Ausbreitung zu einem Muster abgegrenzter Gebiete, welche durch die Ausbreitung der Mutationen verschiedener Ursprünge entstehen.
In der vorliegenden Arbeit behandeln wir ein Modell mit diskreten Demes und betrachten eine deterministische Ausbreitung einer vorteilhaften Mutation aus einem beliebigen Deme zu allen benachbarten Demes. Dies existiert bereits im kontinuirlichen Raum, wo die deterministische Ausbreitung einer Anpassung analysiert wurde, jedoch nicht in einem Modell mit diskreten Demes.
Im Mittelpunkt dieser Arbeit steht das eindimensionale Modell, doch auch der zwei- bzw. dreidimensionale Fall werden behandelt. Abschließend wird ein Vergleich mit Resultaten der Ewens' Sampling Formel gezogen.

Manuela Geiß

Curriculum vitae

Personal Data

Name	Manuela Geiß
Date of birth	March 18th, 1990
Place of birth	Bad Soden am Taunus, Germany
Nationality	Germany

Education

08/00–06/09	**Secondary School**, *Albert-Einstein-Gymnasium Schwalbach am Taunus*, Graduation with distinction
10/09–08/12	**Bachelor Mathematics**, *University of Vienna*, Graduation with distinction
09/13–01/14	**ERASMUS Outgoing semester**, *Université Pierre et Marie Curie (Paris 6)*
since 10/12	**Master Mathematics**, *University of Vienna*
since 10/12	**Bachelor Biology**, *University of Vienna*

Scholarships

2010, 2012–2014	**Academic Excellence Scholarship**, *University of Vienna*

Languages

German	native
English	fluent
French	fluent
Latin	Latinum