Mean and variance of absorption time in a population genetics model: Comparison of Markov chain and diffusion process methods.

Alexander Koronen

March 15, 2019

Abstract

Finding the expected number of generations it takes before every individual in a population (of constant size) only has one type of gene at a particular locus of a chromosome, can be formulated as computing the mean absorption time in a Markov chain. However, the Markov chain approach is inappropriate when the population is large. This is commonly solved by approximating the Markov chain with a diffusion process, in which the mean absorption time is found by solving an ODE with boundary conditions. In this thesis, the formulas for the mean absorption time is derived in both cases. Using these formulas, the expected number of generations is computed numerically. The two different methods are compared, and we also discuss the genetic meaning of the results.

Acknowledgements

First, i would like to thank my supervisors Pelle Pettersson and Torbjörn Säll for introducing me to this interesting problem. It allowed me to use many different tools that i acquired during my mathematics education, and doing so for a practical purpose. It has given me a taste of applying abstract mathematical tools to concrete problems, which is something i really enjoy and would like to do in the future in one way or another. In particular, thanks to Pelle for many fun meetings, where sometimes we would both get stuck on small details and struggle with the numerics. Your humble attitude made it a pleasure to be your student. Thanks to Torbjörn for giving many and long explanations of the genetic aspects of this thesis subject.

Second, i want to thank my friends, family, and my girlfriend for always supporting me, despite not always understanding what i was doing, and for believing in me, even when i did not do so myself.

Contents

1 Genetical background

1.1 Introduction

To properly formulate the problem and purpose of this thesis in a non-mathematical way, we need to introduce some terminology from population genetics. Within a cell, there are *chromosomes*. A particular position on a chromosome is called a *locus*. On each locus of a chromosome there are, among other things, *genes*.

Hartl [3] describes a gene as:

"*Gene is a general term meaning, loosely, the physical entity transmitted from parent to offspring during the reproductive process that influences hereditary traits.*"

When there can be different versions of a gene at one particular locus, these different versions are called *alleles*. In set theoretic notation, this can be summarized as

$$\text{Allele (gene type)} \in \text{Locus} \subset \text{Chromosome} \subset \text{Cell}.$$

When organisms mate, there are two types of cells to consider. There are the "parents", called *haploid* cells or *gametes*, which contain one set of chromosomes. Then there is the "offspring", called *diploid* cells or *zygotes*, which has two copies of each type of chromosome, and hence two copies of each locus. The two copies of a particular locus can contain different alleles, and the different pairings of alleles thereby possible are called *genotypes*. The physical expression of a genotype (for example having blue eyes) is called *phenotype*.

In this thesis, we will consider two different alleles, A_1 and A_2, which leads to three[1] different genotypes: A_1A_1, A_1A_2 and A_2A_2. The number of a certain allele in a population depends on the number of that allele in the previous generation, and can be affected by *selection*, *mutation* and randomness (known as *drift*). These three forces will be properly defined once we get to the mathematical section.

Now, the purpose of this thesis is to compute the expected number of generations it will take until every individual of the population only has A_2 alleles at a particular locus, given that each individual of the first generation only has A_1 alleles at that locus, and that selection, mutation and drift are all at work. It is an important point that even though selection may work in favor of one of the alleles, that allele can still be lost from the population due to drift.

[1] We do not include A_2A_1 since it will have the same effect as A_1A_2. This symmetry will be evident in later formulas.

1.2 Hardy-Weinberg

The most fundamental result in the theory of population genetics is known as the Hardy-Weinberg theorem. Its validity relies on a number of assumptions:

- Large population size.
- Non-overlapping generations (constant population size).
- Random mating (uniformly).
- Equal allele frequency in both sexes.
- No mutation, selection or migration.

Theorem 1.1 (Hardy-Weinberg)**.** *Under the assumptions stated, a population having genotype frequencies X for A_1A_1, $2Y$ for A_1A_2 and Z for A_2A_2, achieves after one generation of random mating, stable genotype frequencies $x^2, 2x(1-x)$ and $(1-x)^2$, where*

$$x = X + Y,\ 1 - x = Y + Z.$$

If the initial frequencies are already of the form $x^2, 2x(1-x)$ and $(1-x)^2$, then those frequencies are stable for all generations.

To see this, let the genotypes A_1A_1, A_1A_2 and A_2A_2 have frequencies $X, 2Y$ and Z respectively in the first generation, and let $P(A_iA_j)$ denote the probability that a certain mating results in the genotype A_iA_j. Then the results of random mating can be summarized in the following table:

Mating	Frequency	$P(A_1A_1)$	$P(A_1A_2)$	$P(A_2A_2)$
$A_1A_1 \times A_1A_1$	X^2	1	0	0
$A_1A_1 \times A_1A_2$	$4XY$	1/2	1/2	0
$A_1A_1 \times A_2A_2$	$2XZ$	0	1	0
$A_1A_2 \times A_1A_2$	$4Y^2$	1/4	1/2	1/4
$A_1A_2 \times A_2A_2$	$4YZ$	0	1/2	1/2
$A_2A_2 \times A_2A_2$	Z^2	0	0	1

Note that the frequencies are correct since they sum up to $(X + 2Y + Z)^2 = 1$. By looking in this table, we can deduce that the genotype frequencies in the next generation are:

$$\begin{aligned}
X' &= X^2 + \frac{1}{2}4XY + \frac{1}{4}4Y^2 = X^2 + 2XY + Y^2 = (X+Y)^2, \\
2Y' &= \frac{1}{2}4XY + 2XZ + \frac{1}{2}4Y^2 + \frac{1}{2}4XZ = 2(X+Y)(Y+Z), \\
Z' &= \frac{1}{4}4Y^2 + \frac{1}{2}4YZ + Z^2 = Y^2 + 2YZ + Z^2 = (Y+Z)^2.
\end{aligned}$$

Now let us see what happens in the next generation, i.e. after another random mating. By following the above formulas, we get (after some algebra) that:

$$\begin{aligned}
X'' &= (X' + Y')^2 = (X(X + 2Y + Z) + Y(X + 2Y + Z))^2 = (X + Y)^2 = X', \\
2Y'' &= 2(X' + Y')(Y' + Z') = 2(X + 2Y + Z)(X + Y)(Y + Z) \\
&\quad = (X + Y)(Y + Z) = 2Y', \\
Z'' &= (Y' + Z')^2 = (Y(X + 2Y + Z) + Z(X + 2Y + Z))^2 = (Y + Z)^2 = Z'.
\end{aligned}$$

This shows that, with no forces other than drift at work, the genotype frequencies in a population become fixed after one generation.

1.3 Urn model

In this section, we develop a method to count the number of A_1 alleles in a generation. The total number of gametes will be denoted $N = 2M$, where M is the number of individuals in the population. To start with, the mating of two diploid individuals can intuitively be visualized as in Figure 1: The black and

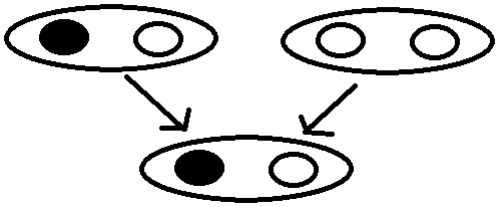

Figure 1: Two zygotes mate randomly

white dots are the two different alleles of a gene, and the larger ring is the locus[2] we are interested in. However, since we assume that the mating of diploid individuals is random, and the offspring obtains one gene from each of its parents, the mating of two zygotic individuals can be seen as a random collision of two gametes, as visualized in Figure 2. This can be abstracted one step further, to

Figure 2: Two gametes "collide" randomly, which results in a zygote.

what we will refer to as the urn model, visualized in Figure 3. In the urn model, we imagine that all the (huge number of) gametes produced by the N zygotes in one generation are thrown into a large urn, divided into N sections. The population in the next generation is then obtained by "drawing" two gametes (and hence producing a zygote) from each section of the urn. Under the assumptions of the Hardy-Weinberg theorem, the probability of drawing an allele is just the frequency of that allele in the previous generation.

[2]Or rather, a merging of the two copies of that locus.

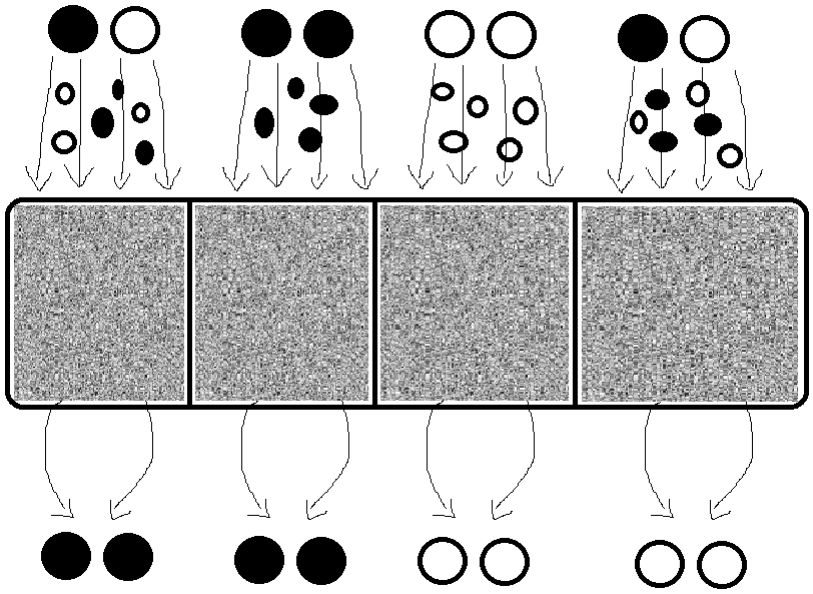

Figure 3: A visualization of the urn model

We can now ask: what is the probability that there are j A_1 alleles in generation t? To start, j A_1 alleles among the N genes can be chosen in $\binom{N}{j}$ ways. Due to our assumptions of random mating and absence of selection and mutation, the probability of drawing an A_1 allele from the urn is just $x_k = \frac{k}{N}$, the frequency of A_1 alleles in the previous generation. The probability of drawing j A_1 alleles is then just p_k^j. To keep the population size constant, we must at the same time draw $N - j$ A_2 alleles, and the probability of this is $(1 - x_k)^{N-j}$. In total, the probability we are looking for is

$$\binom{N}{j} x_k^j (1 - x_k)^{N-j}.$$

Now denote the number of A_1 alleles in generation t by $X(t)$. In the language of probability theory, we have shown that $X(t + 1)$, given that $X(t) = k$ and that drift is the only force at work, is a binomially distributed random variable with $p = x_k = \frac{k}{N}$:

$$X(t+1)\big|_{X(t)=k} \sim \text{Bin}(N, x_k). \tag{1.1}$$

To make things more interesting (and realistic), we can let go of the last of the assumptions made in the previous subsection, and introduce selection and mutation into our model. Selection means that some genotypes are more successful in producing offspring than others, and therefor more likely to survive into the next generation. Selection enters into our model in the form of two parameters. First, a selection coefficient, denoted by s, which describes how much "better"

one of the alleles is, in comparison to the other.[3] In this thesis we will only consider the cases $s = 0.02$ and $s = 0.1$.

The second parameter related to selection is called the degree of dominance, and is denoted by h. It models the dominance relation between the two alleles, and decides the phenotype of the genotype A_1A_2. The most common (and interesting) values for h are 0, $\frac{1}{2}$ and 1. In the case $h = 0$, A_1 is said to be *dominant* to A_2, which means that the phenotype of A_1A_2 will be the same as that of A_1A_1. In the case $h = 1$, A_1 is said to be *recessive* to A_2, which means that the phenotype of A_1A_2 will be the same as that of A_2A_2. In the intermediate case $h = \frac{1}{2}$, the effects of the alleles on the genotype A_1A_2 is said to be *additive*.

The selection parameters s and h are introduced into the model in the form of relative weights w_{ij} for the three different genotypes, indicating how much more successful some of the genotypes are at surviving and reproducing compared to the others.

Genotype	Relative weight	Meaning
A_1A_1	$w_{11} = 1$	Reference. Others are compared to A_1A_1.
A_1A_2	$w_{12} = 1 + hs$	h decides relation between A_1A_2 and the rest.
A_2A_2	$w_{22} = 1 + s$	A_2 has selective advantage.

Table 1: Genotypes and their relative weights.

We emphasize that it is the relationship (quotient) between the weights that matters, and not their specific value.[4] Also note that in the case $h = \frac{1}{2}$, we have that $w_{12} = \frac{w_{11}+w_{22}}{2}$, which explains the choice of the word "additive".

Lastly, mutation means that given allele can, with a certain, usually very low probability, turn into another allele. We will only consider the case when A_1 can mutate to A_2.[5] This enters into our model (naturally) in the form of a mutation probability, denoted by u. The mutation probability is usually around 10^{-6} to 10^{-5}, and throughout this thesis we will, unless otherwise noted, let $u = 5 \cdot 10^{-6}$.

[3] For example, if A_2 has a selection coefficient of 0.02, it means that it is 2 percent better than A_1.

[4] For example, Ewens [2] uses the weights $1 + \tilde{s}$, $1 + \tilde{h}\tilde{s}$ and 1 respectively (the tilde was added for clarity). For the quotients to be equal, we must have that $s = -\tilde{s} + \mathcal{O}(N^{-1})$ and $h = 1 - \tilde{h}$. After these adjustments, the formulas given in [2] and the ones given in this thesis become equivalent.

[5] Many authors have also considered a parameter v, being the probability of mutation $A_2 \mapsto A_1$. Many results are known in the cases $u = v = 0$ and $u > 0,\ v > 0$, but we will not consider those cases here. For details and references see [2].

The parameters of our model are summarized in the following table:

Parameter	Meaning	Values
M	Number of individuals in the population.	$0.5 \cdot 10^k,\ k \in \mathbb{N}$
N	Number of gametes (gene copies). $N = 2M$	$10^k,\ k \in \mathbb{N}$
s	Models how much "better" the A_2 allele is.	$0.02,\ 0.1$
h	Models the dominance between A_1 and A_2.	$0,\ 0.5,\ 1$
u	Probability of mutation $A_1 \mapsto A_2$.	$5 \cdot 10^{-6}$

Table 2: Explanation of the parameters.

We emphasize that the parameter values given in the table are not absolute. However, the values given are the most typically occurring ones in the literature, and therefore the ones we will consider.

Now with all the parameters and weights properly introduced, we redefine the probability of "drawing" an A_1 allele in generation $t+1$ as

$$\begin{aligned} p_k &= \mathbb{P}(\text{drawing an } A_1 \text{ allele at } t+1 \mid k \ A_1 \text{ alleles at } t) \\ &= \frac{(w_{11}x_k^2 + w_{12}x_k(1-x_k))(1-u)}{w_{11}x_k^2 + 2w_{12}x_k(1-x_k) + w_{22}(1-x_k)}. \end{aligned} \tag{1.2}$$

The expression $w_{11}x_k^2 + w_{12}x_k(1-x_k)$ is commonly called the fitness of the A_1 allele. Note that the assumptions made in the previous case amounts to setting $w_{11} = w_{12} = w_{22} = 1$ and $u = 0$, so that

$$p_k = \frac{x_k^2 + x_k(1-x_k)}{x_k^2 + 2x_k(1-x_k) + (1-x_k)^2} = \frac{x_k}{(x_k + 1 - x_k)^2} = x_k,$$

as stated.

The numerator of (1.2) has the following interpretation: An A_1 allele can come from a A_1A_1 genotype (which has frequency $w_{11}x_k^2$), or as one of the two possibilities from a A_1A_2 genotype (which has frequency $w_{12}x_k(1-x_k)$), and we only count the fraction of A_1 alleles that do not mutate into A_2. The denominator describes all possible outcomes, and is commonly called the fitness of the population.

In this setting, it is clear that $X(t+1)$ conditional on $X(t)$ still has a binomial distribution, but now with $p = p_k$ given by (1.2). In mathematical notation:

$$X(t+1)|_{X(t)=k} \sim \text{Bin}(N, p_k). \tag{1.3}$$

It turns out that the assumptions of constant population size and random mating makes it natural to model changes in the number of A_1 alleles, $X(t)$, at the locus as a so-called *Markov chain*, regardless if $X(t)$ follows (1.1) or (1.3). The theory of Markov chains allows us compute many quantities of interest using linear algebra, and it is the subject of the next section.

2 Markov chain on $\mathbb{N}$

2.1 General theory

Definition 2.1. *A Markov chain is a stochastic process $\{X(t)\}$ in discrete time, with discrete state space, and which has the property:*

$$\mathbb{P}(X(t+1) = i_{n+1} \mid X(t) = i_n, \cdots, X(0) = i_0) = \mathbb{P}(X(t+1) = i_{n+1} \mid X(t) = i_n).$$

This property is known as the (weak) Markov property.

Intuitively, the Markov property means that the next step of the process only depends on the current state, and not on the previous history of the process. Most Markov chains of interest also satisfy the following:

Definition 2.2. *Let $\{X(t)\}_{t\in\mathbb{N}}$ be a Markov chain. If the conditional probabilities*

$$p_{j,i} = \mathbb{P}(X(t+1) = j \mid X(t) = i)$$

do not depend on t, the process is called time-homogeneous.

The conditional probabilities in the definition above are called transition probabilities, since they describe the probability that the Markov chain makes the transition from state i to state j. The matrix whose elements are transition probabilities is called the transition matrix, and is denoted P. Since the chain will always move to *some* state in each time step, the row sums of P are all 1. The next definition says that it is also possible to compute the probability that the Markov chain is in a certain state, several steps from now.

Definition 2.3. *The probability*

$$p_{j,i}^{(m)} = \mathbb{P}(X(t+m) = j \mid X(t) = i)$$

is called the transition probability of order m.

From the law of total probability, we easily obtain the following important result, which allows us to compute higher order transition probabilities in term of lower order ones by conditioning on an intermediate step:

Theorem 2.1 (Chapman-Kolmogorov)**.** *Let P be the transition matrix of a time-homogeneous Markov chain. Then the transition probabilities of order $m+n$ satisfy*

$$p_{j,i}^{(m+n)} = \sum_k p_{j,k}^{(m)} p_{k,i}^{(n)}, \tag{2.1}$$

and the matrix $\mathsf{P}^{(m)}$ of transition probabilities of order m satisfy $\mathsf{P}^{(m)} = \mathsf{P}^m$.

To "start" a Markov process, we need an initial distribution, a vector containing the probabilities of starting in each of the states. This is usually denoted $\boldsymbol{\pi}(0)$. The probability that the process is in state i at time t is then given by the i:th

element of the vector $\boldsymbol{\pi}(t) = \boldsymbol{\pi}(0)\mathsf{P}^t$.

Under some conditions, a Markov chain can have a so-called stationary distribution: an initial distribution that does not change with time, i.e. satisfies $\boldsymbol{\pi}(0) = \boldsymbol{\pi}(0)\mathsf{P}^t$ for all t.[6] We will not consider stationary distributions here (see for example section VIII §4 in [8]), but instead focus on another important property that some Markov chains possess:

Definition 2.4. *State j is absorbing if*

$$p_{i,j} = \delta_{ij},$$

where δ_{ij} is the Kronecker delta. In other words: if the process stays in state j forever once it has entered it.

Intimately related to the notion of an absorbing state is the notion of an absorbing Markov chain:

Definition 2.5. *A Markov chain is called absorbing if it is possible to reach the absorbing state from any other state.*

A related, but more general property is that of being irreducible:

Definition 2.6. *A Markov chain is irreducible if it is possible to reach any state from any other state.*

The Markov chains considered in this thesis will be absorbing, and the main focus of the thesis is to compute the mean time it takes for the chain to reach its absorbing state. If the absorbing state (or one of them) has index j, this is defined as the expected value of the random variable

$$T_j = \min\{t \mid t > 0,\ X(t) = j\}.$$

Assume that a Markov chain has n states, m of which are absorbing. Then the transition matrix of the process, P, can (possibly after a relabelling of states) be factored as

$$\mathsf{P} = \begin{pmatrix} \mathsf{I} & \mathsf{0} \\ \bar{\mathsf{P}} & \hat{\mathsf{P}} \end{pmatrix}, \tag{2.2}$$

where I is an $m \times m$ identity matrix consisting of the (trivial) transition probabilities between the absorbing states, $\mathsf{0}$ is an $m \times (n-m)$ zero matrix, $\bar{\mathsf{P}}$ is an $(n-m) \times m$ matrix consisting of the transition probabilities from the non-absorbing states to the absorbing states, and $\hat{\mathsf{P}}$ consists of the transition probabilities between the non-absorbing states. Observe that in order for P to be absorbing, $\bar{\mathsf{P}}$ must contain at least one nonzero element. $\hat{\mathsf{P}}$ will be irreducible in this thesis, but it doesn't need to be in general. It turns out that in order to analyze the absorption time T_j, only the matrix $\hat{\mathsf{P}}$ is needed.

[6] In fact, in the case $u > 0,\ v > 0$ briefly mentioned in the previous section, the resulting Markov chain will have a stationary distribution instead of absorbing states.

To simplify notation a bit, let us denote $e_i = (0, \ldots, 1, \ldots, 0)$, and $\mathbf{1} = (1, 1, \ldots, 1)$. Also let $(.)_{j,i}$ denote the element on row i and column j of a matrix. The main result of this section is:

Theorem 2.2. *The mean absorption time in a Markov chain, given that it starts in state i, is given by*

$$\mathbb{E}(T_j \mid X(0) = i) = e_i^T (\mathsf{I} - \hat{\mathsf{P}})^{-1} \mathbf{1}. \tag{2.3}$$

Hence, it is the i:th row sum of the matrix $(\mathsf{I} - \hat{\mathsf{P}})^{-1}$. Some work needs to be done before we can actually prove this. We need to show that $(\mathsf{I} - \hat{\mathsf{P}})^{-1}$ always exists, and then figure out how to find it. First, we recall some matrix theory.

Given a vector norm $\|v\|$ on $\mathbb{R}^n$, the corresponding *induced matrix norm* on $\mathbb{R}^{n \times n}$ (space of real quadratic matrices) is given by

$$\|\mathsf{A}\| := \sup_{\|v\| \neq 0} \frac{\|\mathsf{A}v\|}{\|v\|} = \sup_{\|v\|=1} \|\mathsf{A}v\|.$$

Such a matrix norm has the *submultiplicative property*: $\|\mathsf{AB}\| \leqslant \|\mathsf{A}\|\|\mathsf{B}\|$. The following lemma can now solve our problems:

Lemma 2.3. *Let $\mathsf{A} \in \mathbb{R}^{n \times n}$ with $\|\mathsf{A}\| < 1$. Then $\mathsf{I} - \mathsf{A}$ is invertible and*

$$\sum_{k=0}^{\infty} \mathsf{A}^k = (\mathsf{I} - \mathsf{A})^{-1}.$$

Proof. First we show invertibility. Let u be an eigenvector of A corresponding to an arbitrary eigenvalue λ. Then

$$\|\mathsf{A}\|\|u\| = \sup_{\|v\| \neq 0} \frac{\|\mathsf{A}v\|}{\|v\|} \|u\| \geqslant \frac{\|\mathsf{A}u\|}{\|u\|} \|u\| = \|\mathsf{A}u\| = |\lambda| \|u\|.$$

Since $\|u\| \neq 0$ (because it is an eigenvector), this means that $|\lambda| \leqslant \|\mathsf{A}\| < 1$. Hence 1 is not an eigenvalue of A, so that $\mathsf{A}x = x$ only holds if $x = 0$, and this is equivalent to $(\mathsf{I} - \mathsf{A})$ being invertible.

By the triangle inequality, submultiplicativity and the assumption,

$$\| \sum_{k=0}^{\infty} \mathsf{A}^k \| \leqslant \sum_{k=0}^{\infty} \|\mathsf{A}\|^k = \frac{1}{1 - \|\mathsf{A}\|} < \infty.$$

Hence the left hand side is well defined. It is equally easy to see that

$$(\mathsf{I} - \mathsf{A})(\mathsf{I} + \mathsf{A} + \cdots + \mathsf{A}^{n-1}) = \mathsf{I} - \mathsf{A}^n, \tag{2.4}$$

so it only remains to show that $\lim_{n \to \infty} \mathsf{A}^n = 0$. But this follows easily from continuity and submultiplicativity of the norm:

$$\| \lim_{n \to \infty} \mathsf{A}^n \| = \lim_{n \to \infty} \|\mathsf{A}^n\| \leqslant \lim_{n \to \infty} \|\mathsf{A}\|^n = 0.$$

The result now follows by taking limits on both sides of (2.4). □

Every row sum of P is 1, and since we assume that our Markov chain is absorbing, there is a nonzero probability to reach the absorbing state(s) from any other state. Hence the row sums of $\hat{\mathsf{P}}$ must all be strictly less than 1, which makes it is natural to consider the matrix norm

$$\|\mathsf{A}\|_\infty = \max_{1 \leqslant i \leqslant n} \sum_{j=1}^{n} |a_{ij}|, \tag{2.5}$$

which is induced by the vector norm

$$\|v\|_\infty = \max_{1 \leqslant i \leqslant n} |v_i|.$$

With the norm (2.5), $\hat{\mathsf{P}}$ satisfies the hypothesis of lemma 2.4, and so we have proved

$$\sum_{k=0}^{\infty} \hat{\mathsf{P}}^k = (\mathsf{I} - \hat{\mathsf{P}})^{-1}, \tag{2.6}$$

which is the first step towards proving the main result. Next, let n_j denote the number of times a Markov process is in state j.[7] The next lemma shows us how to find the conditional expectation of n_j:

Lemma 2.4. $\mathbb{E}(n_j \mid X(0) = i) = \left((\mathsf{I} - \hat{\mathsf{P}})^{-1}\right)_{j,i}$

Proof. Consider the indicator random variable

$$I_{\{X(k)=j\}} = \begin{cases} 1, & \text{if } X(k) = j \\ 0, & \text{otherwise} \end{cases}.$$

It is clear that

$$n_j = \sum_{k=0}^{\infty} I_{\{X(k)=j\}},$$

and the conditional expectation is by definition

$$\mathbb{E}(I_{\{X(k)=j\}} \mid X(0) = i) = p_{j,i}^{(k)}.$$

[7] Note that n_j is a random variable (since it is a function of a random variable), so that it makes sense to consider its conditional expectations. Also note that we must either set $n_i = 0$ if i is an absorbing state, or only define n_j for non-absorbing states. However, our calculations will not be affected by which of these approaches we choose.

Hence, by a limit theorem for conditional expectations (see [8] p. 218):

$$\begin{aligned}\mathbb{E}(n_j \mid X(0)=i) &= \mathbb{E}\left(\sum_{k=0}^{\infty} I_{\{X(k)=j\}} \;\middle|\; X(0)=i\right)\\ &= \sum_{k=0}^{\infty} \mathbb{E}(I_{\{X(k)=j\}} \mid X(0)=i)\\ &= \sum_{k=0}^{\infty} p_{j,i}^{(k)}\\ &= \left(\sum_{k=0}^{\infty} \hat{\mathsf{P}}^k\right)_{j,i}\\ &= \left((\mathsf{I}-\hat{\mathsf{P}})^{-1}\right)_{j,i},\end{aligned}$$

where (2.6) was used in the last equality. □

With this lemma at hand, it is now an easy task to prove Theorem 2.2:

Proof. For simplicity, we consider the case when 0 is the only absorbing state (if there are several absorbing states, then they can be merged into one by rearranging P). Since an absorbing Markov chain will reach its absorbing state with probability one (see for example [4] thm 3.1.1), the left hand side of (2.3) can be computed by summing the mean time the process spends in every non-absorbing state. This gives:

$$\begin{aligned}\mathbb{E}(T_0 \mid X(0)=i) &= \mathbb{E}\left(\sum_{j=1}^{N} n_j \;\middle|\; X(0)=i\right)\\ &= \sum_{j=1}^{N} \mathbb{E}(n_j \mid X(0)=i)\\ &= \sum_{j=1}^{N} \left((\mathsf{I}-\hat{\mathsf{P}})^{-1}\right)_{j,i}\\ &= e_i^T(\mathsf{I}-\hat{\mathsf{P}})^{-1}\mathbf{1},\end{aligned} \tag{2.7}$$

and we are done. □

Next, we deal with the variance.

Theorem 2.5. *The variance of the absorption time in a Markov chain, given that it starts in state i, is given by*

$$\mathbb{V}(T_0 \mid X(0)=i) = e_i^T\left(2(\mathsf{I}-\hat{\mathsf{P}})^{-2}-(\mathsf{I}-\hat{\mathsf{P}})^{-1}\right)\mathbf{1} - \left(e_i^T(\mathsf{I}-\hat{\mathsf{P}})^{-1}\mathbf{1}\right)^2 \tag{2.8}$$

In words, the variance is the i:th row sum of the matrix $2(\mathsf{I}-\hat{\mathsf{P}})^{-2}-(\mathsf{I}-\hat{\mathsf{P}})^{-1}$ minus the squared i:th row sum of the matrix $(\mathsf{I}-\hat{\mathsf{P}})^{-1}$.

Proof. For simplicity, let us denote

$$\mathbf{E} = \Big(\mathbb{E}(T_0 \mid X(0) = 1), \ldots, \mathbb{E}(T_0 \mid X(0) = N)\Big).$$

From the previous result we know that

$$\mathbf{E} = (\mathsf{I} - \hat{\mathsf{P}})^{-1}\mathbf{1}.$$

Similarly, let us denote

$$\mathbf{E_2} = \Big(\mathbb{E}(T_0^2 \mid X(0) = 1), \ldots, \mathbb{E}(T_0^2 \mid X(0) = N)\Big).$$

With this notation, the variance is

$$\mathbb{V}(T_0 \mid X(0) = i) = e_i^T(\mathbf{E_2} - \mathbf{E} \circ \mathbf{E}),$$

where $\circ$ denotes the elementwise product between two vectors (also known as the Hadamard product). Hence, we need to find the vector $\mathbf{E_2}$.

The i:th element of $\mathbf{E_2}$ can be found by conditioning on the first step of the process, and using the law of total probability:

$$\mathbb{E}(T_0^2 \mid X(0) = i) = \sum_{k=0}^{N} \mathbb{E}(T_0^2 \mid X(0) = i \ , \ X(1) = k)p_{k,i}. \tag{2.9}$$

If $k = 0$, the process cannot visit any non-absorbing states except the starting state. If $k \neq 0$, the starting state is still non-absorbing, and by the Markov property, we can restart the count after the first step.[8] In mathematical notation, this means that

$$T_0\big|_{X(0)=i, X(1)=k} = \begin{cases} 1, & \text{if } k = 0 \\ 1 + T_0, & \text{if } k \neq 0 \end{cases}.$$

Consequently,

$$\mathbb{E}(T_0^2 \mid X(0) = i \ , \ X(1) = k) = \begin{cases} 1, & \text{if } k = 0 \\ \mathbb{E}((1 + T_0)^2 \mid X(1) = k), & \text{if } k \neq 0 \end{cases}.$$

Observe that we now condition on k and not i (see the footnote).

[8] We can think of this as considering T_0 for the process $Y(t) = X(t + 1)$, which has the same properties as $X(t)$ due to the Markov property. For $Y(t)$ we are then only conditioning on $Y(0) = k$.

Hence the sum (2.9) is

$$\begin{aligned}\sum_{k=0}^{N} \mathbb{E}(T_0^2 \mid X(0)=i, X(1)=k)p_{k,i} &= 1p_{0,i} + \sum_{k=1}^{N} \mathbb{E}((1+T_0)^2 \mid X(1)=k)p_{k,i} \\ &= 1 + 2\sum_{k=1}^{N} \mathbb{E}(T_0 \mid X(1)=k)p_{k,i} \\ &\quad + \sum_{k=1}^{N} \mathbb{E}(T_0^2 \mid X(1)=k)p_{k,i}.\end{aligned}$$

This sum can be identified as the i:th element of the vector

$$\mathbf{1} + 2\hat{\mathsf{P}}\mathbf{E} + \hat{\mathsf{P}}\mathbf{E_2}.$$

Hence, if we tie things together and recall the previous result, we get

$$(\mathsf{I}-\hat{\mathsf{P}})\mathbf{E_2} = \mathbf{1} + 2\hat{\mathsf{P}}(\mathsf{I}-\hat{\mathsf{P}})^{-1}\mathbf{1} = (\mathsf{I} + 2\hat{\mathsf{P}}(\mathsf{I}-\hat{\mathsf{P}})^{-1})\mathbf{1},$$

which is equivalent to

$$\begin{aligned}\mathbf{E_2} &= (\mathsf{I}-\hat{\mathsf{P}})^{-1}(\mathsf{I} + 2\hat{\mathsf{P}}(\mathsf{I}-\hat{\mathsf{P}})^{-1})\mathbf{1} \\ &= (\mathsf{I}-\hat{\mathsf{P}})^{-1}(\mathsf{I} + 2((\mathsf{I}-\hat{\mathsf{P}})^{-1} - \mathsf{I}))\mathbf{1} \\ &= (\mathsf{I}-\hat{\mathsf{P}})^{-1}(2(\mathsf{I}-\hat{\mathsf{P}})^{-1} - \mathsf{I})\mathbf{1} \\ &= (2(\mathsf{I}-\hat{\mathsf{P}})^{-2} - (\mathsf{I}-\hat{\mathsf{P}})^{-1})\mathbf{1},\end{aligned}$$

since $(\mathsf{I}-\hat{\mathsf{P}})^{-1}\hat{\mathsf{P}} = (\mathsf{I}-\hat{\mathsf{P}})^{-1} - \mathsf{I}$ (recall (2.6)).

We can now conclude that

$$\begin{aligned}\mathbb{V}(T_0 \mid X(0)=i) &= e_i^T(\mathbf{E_2} - \mathbf{E}\circ\mathbf{E}) \\ &= e_i^T\big(2(\mathsf{I}-\hat{\mathsf{P}})^{-2} - (\mathsf{I}-\hat{\mathsf{P}})^{-1}\big)\mathbf{1} - \left(e_i^T(\mathsf{I}-\hat{\mathsf{P}})^{-1}\mathbf{1}\right)^2.\end{aligned}$$

□

Remark 2.6. *The expectation could have been computed using the same approach as in the proof above. Also, the above variance could have been computed in a similar way to the expectation: by first computing the variance of n_j, and then summing the result over all non-absorbing states (assuming of course that the n_j's are independent). However, that approach leads to a sum that is hard to manage.*

2.2 Our case

Recall from section 1.3 that the number of A_1 alleles in generation $t+1$, $X(t+1)$, conditional on $X(t) = k$ and that selection, mutation and drift are all at work, follows a binomial distribution,

$$X(t+1)|_{X(t)=k} \sim \text{Bin}(N, p_k), \tag{2.10}$$

where

$$p_k = \frac{(w_{11}x_k^2 + w_{12}x_k(1-x_k))(1-u)}{w_{11}x_k^2 + 2w_{12}x_k(1-x_k) + w_{22}(1-x_k)^2},$$

and $x_k = \frac{k}{N}$ is the frequency of A_1 alleles at the locus in the previous generation.

Using the terminology introduced in the previous subsection, can now model the changes in the number of A_1 alleles from generation to generation as as a time-homogeneous, absorbing Markov chain $\{X(t)\}_{t\in\mathbb{N}}$ with transition probabilities

$$p_{j,k} = \mathbb{P}(X(t+1) = j | X(t) = k) = \binom{N}{j} p_k^j (1-p_k)^{N-j}. \tag{2.11}$$

The absorbing state of this Markov chain is state 0, corresponding to the event that there are no more A_1 alleles in the population. The mean and variance of the absorption time can now be computed by (2.2) and (2.8) respectively.

Although we have now found formulas for the desired quantities, those formulas are not suitable for numerical computation if the state space is large (which is the case in practical applications!). For one thing, the binomial coefficients included in the transition probabilities quickly become very large. For example, the number $\binom{100}{50} \approx 10^{29}$ is already much bigger than the estimated number of stars in the universe.[9] A way to get around this is to take the natural logarithm of the transition probabilities (2.11), and then compute them recursively according to

$$\begin{cases} \log p_{k,0} = N \log(1-p_k) \\ \log p_{k,j} = \log p_{k,j-1} + \log\left(\frac{N-j+1}{j}\right) + \log\left(\frac{p_k}{1-p_k}\right) \end{cases} \tag{2.12}$$

The transition matrix of the process will then be the (elementwise) exponential of the resulting matrix. However, even with this simplification, one has to be careful when building the transition matrix numerically for chains with state spaces of size $N = 10^4$ or larger. Using some more sophisticated tricks (sparse matrix methods among other things) it is possible to do the computations up to $N = 10^6$, but even then it takes considerable time and computational effort. Therefore, some other strategy is needed to do these computations for large

[9] It is of course impossible to know the exact number of stars, and the estimated number varies depending on who you ask. A rough estimate is 10^{21}, see for example http://curious.astro.cornell.edu/the-universe/stars-and-star-clusters/78-the-universe/stars-and-star-clusters/general-questions/345-is-it-possible-to-count-the-stars-beginner

populations in practice.

We end this subsection by providing formulas for the mean and variance of the change in the value of the Markov process in one time step. They will be important later on.

Theorem 2.7. *For a time-homogeneous Markov chain with binomial transition probabilities, the conditional mean and variance of the change in one time step is given by*

$$\mathbb{E}(X(t+1) - X(t) \mid X(t) = k) = Np_k - k. \tag{2.13}$$

$$\mathbb{V}(X(t+1) - X(t) \mid X(t) = k) = Np_k(1 - p_k). \tag{2.14}$$

Proof. This follows directly from (2.10) and elementary properties of mean and variance. □

2.2.1 Scaling

The first step towards approximating our Markov chain with a diffusion process is to scale both time and space by $\frac{1}{N}$ (so that we consider the frequency of A_1 alleles rather than the number of them).[10] This will not affect the transition probabilities (2.11), since the event

$$\frac{1}{N} X\left(\frac{t}{N}\right) = \frac{k}{N}$$

is equivalent to the event $X(t) = k$. However, (2.13) and (2.14) will look a bit different, as we shall see.

In what follows, we assume that the parameters s and u are both $\mathcal{O}(N^{-1})$ (this will be necessary for our later approximations to be valid). First, let's see what happens with the probability p_k after scaling. If we expand (1.2), and write x instead of x_k, we obtain

$$\begin{aligned}
p_k &= \frac{(x^2 + (1+hs)x(1-x))(1-u)}{x^2 + 2(1+hs)x(1-x) + (1+s)(1-x)^2} \\
&= \frac{(x + hsx(1-x))(1-u)}{1 + 2hsx(1-x) + s(1-x)^2} \\
&= (x - ux + hsx(1-x) + \mathcal{O}(N^{-2})) \\
&\qquad \times \left(1 - 2hsx(1-x) - s(1-x)^2 + \mathcal{O}(N^{-2})\right) \\
&= x - ux + hsx(1-x) - 2hsx^2(1-x) - sx(1-x)^2 + \mathcal{O}(N^{-2}) \\
&= x - ux + sx(1-x)(x - 1 + h(1-2x)) + \mathcal{O}(N^{-2}),
\end{aligned}$$

where we used a Taylor expansion in the third equality.

[10] For the time scale, maybe "relabelling" is a better word, since the scaling function $t \mapsto \frac{t}{N}$ is just a bijection between the sets $\mathbb{N}$ and $\{0, \frac{1}{N}, \frac{2}{N}, \ldots\}$.

To simplify notation a bit, let us denote $\tau = \frac{t}{N}$ and $\delta\tau = \frac{1}{N}$. The expected change in one time step, corresponding to (2.13), can then be computed as:

$$\begin{aligned}
\mathbb{E}\left(\frac{X(\tau+\delta\tau)-X(\tau)}{N} \;\middle|\; \frac{X(\tau)}{N} = \frac{k}{N}\right) &= \frac{1}{N}\mathbb{E}\left(X(\tau+\delta\tau)-X(\tau) \;\middle|\; \frac{X(\tau)}{N} = \frac{k}{N}\right) \\
&= \frac{1}{N}\mathbb{E}\left(X(t+1)-X(t) \mid X(t)=k\right) \\
&= \frac{1}{N}\left(Np_k - k\right) \\
&= p_k - x \\
&= -ux + sx(1-x)(x-1+h(1-2x)) \\
&\qquad + \mathcal{O}(N^{-2}).
\end{aligned}$$

By a similar computation, the variance corresponding to (2.14), is seen to be:

$$\begin{aligned}
\mathbb{V}\left(\frac{X(\tau+\delta\tau)-X(\tau)}{N} \;\middle|\; \frac{X(\tau)}{N} = \frac{k}{N}\right) &= \frac{1}{N^2}\mathbb{V}(X(t+1) \mid X(t)=k) \\
&= \frac{1}{N}p_k(1-p_k) \\
&= \frac{1}{N}(x+\mathcal{O}(N^{-1}))(1-x+\mathcal{O}(N^{-1})) \\
&= \frac{1}{N}x(1-x) + \mathcal{O}(N^{-2}).
\end{aligned}$$

In the third equality we used the fact that $x = \mathcal{O}(N^{-1})$ unless $x = 0$ or 1 (and in those cases, the mean and variance above is zero anyway). We will return to these quantities in section 4.

The next step is to see what happens when $N \to \infty$, so that the times become positive real numbers and the state space becomes the closed interval $[0, 1]$. The process thereby obtained is called a *diffusion* process, and it is the subject of the next section.

3 Diffusion process in [0,1]

3.1 General theory

In this section, we will study a diffusion process which is the limit of the scaled Markov process from the previous subsection as $N \to \infty$. For convenience, we allow ourselves to abuse notation and denote this process by $X(t)$ as well. It is not obvious that such a limit process exists (and many authors simply assume it does), but Watterson [10] proved the following

Theorem 3.1. *Let $F_N(x,t) = \mathbb{P}(X(t) \leqslant x)$. Under the time-scale transformation $t = Nu$ and certain sufficient conditions,*

$$\lim_{N\to\infty} F_N(x, Nu) = F(x,u)$$

where $F(x,u)$ is a distribution function uniquely determined by a diffusion equation subject to boundary conditions.

The "certain sufficient conditions" stated by Watterson are rather technical, but the point of them is that the moment generating function of the difference

$$\theta(X(t+1) - X(t)),$$

where θ is a continuous parameter, should have a certain form. Since the proof is rather long and complicated, we refer to the original source [10] for details. Observe that the u used by Watterson should not be confused with the mutation probability u used otherwise in this thesis.

Throughout this chapter, we let δx denote the change in the random variable

$$\lim_{N\to\infty} \frac{1}{N} X\left(\frac{t}{N}\right)$$

from the previous subsection in one time unit δt, and we use Ewens [2] assumptions that

$$\mathbb{E}(\delta x) = a(x)\delta t + o(\delta t), \tag{3.1}$$

$$\mathbb{V}(\delta x) = b(x)\delta t + o(\delta t), \tag{3.2}$$

$$\mathbb{E}(|\delta x|^3) = o(\delta t), \tag{3.3}$$

where both the expectation and variance is of course conditional on a current value x of the random variable.

3.1.1 The forward and backward equations

This subsection concerns the so called forward and backward equations. They are both of fundamental importance in the theory of diffusion processes, but we will only need the backward equation for our purposes. Despite this, derivations of both are included in this thesis for the sake of completeness. Also, we will not solve the backward equation explicitly, but instead use it as a tool in further derivations.

Let $f : [0,1] \times [0,1] \times [0,\infty) \to [0,1]$ be a 3-dimensional distribution function which is $\mathcal{C}^2$ in the first and second arguments, and $\mathcal{C}^1$ in the third argument. Our discussion then begins with the Chapman-Kolmogorov equation, which takes the form:

$$f(z,p,t+\delta t) = \int_0^1 f(x,p,t) f(z,x,\delta t)\, dx. \tag{3.4}$$

This should be interpreted as the probability density of the process changing value from p to z in time $t+\delta t$. Let $Q \in \mathcal{C}^2([0,1])$ be a function that satisfies $Q(0) = Q(1) = Q'(0) = Q'(1) = 0$.[11] If we multiply this equation with such a Q on both sides, and integrate over $[0,1]$, we obtain

$$\int_0^1 Q(z) f(z,p,t+\delta t)\, dz = \int_0^1 \int_0^1 Q(z) f(x,p,t) f(z,x,\delta t)\, dx\, dz. \tag{3.5}$$

If we now do a Taylor expansion of $Q(z)$ (recalling that $z = x + \delta x$) and using the linearity of the integral, we see that this is equal to

$$\begin{aligned} &\int_0^1 \int_0^1 Q(x) f(x,p,t) f(z,x,\delta t)\, dx\, dz \\ &\quad + \int_0^1 \int_0^1 Q'(x) \delta x f(x,p,t) f(z,x,\delta t)\, dx\, dz \\ &\quad + \int_0^1 \int_0^1 Q''(x) \frac{(\delta x)^2}{2} f(x,p,t) f(z,x,\delta t)\, dx\, dz \\ &\quad + \int_0^1 \int_0^1 \mathcal{O}(|\delta x|^3) f(x,p,t) f(z,x,\delta t)\, dx\, dz \end{aligned} \tag{3.6}$$

We will now deal with the integrals one by one, assuming throughout that the orders of integration can be interchanged. Since f is a density function, it is clear that the first one is equal to

$$\int_0^1 Q(x) f(x,p,t) \int_0^1 f(z,x,\delta t)\, dz\, dx = \int_0^1 Q(x) f(x,p,t)\, dx.$$

[11] A simple example of such a function is $Q(x) = x^2(x-1)^2$.

The second integral is similar, but now we also need to use integration by parts and the assumption (3.1):

$$\begin{aligned}
\int_0^1 Q'(x) f(x,p,t) \int_0^1 \delta x f(z,x,\delta t)\,dz\,dx & \\
&= \int_0^1 \mathbb{E}(\delta x) Q'(x) f(x,p,t)\,dx \\
&= \delta t \int_0^1 a(x) Q'(x) f(x,p,t)\,dx + o(\delta t) \\
&= \delta t \left[Q(x)a(x)f(x,p,t)\right]_0^1 \\
&\qquad - \delta t \int_0^1 Q(x) \frac{\partial}{\partial x}(a(x)f(x,p,t))\,dx\big) + o(\delta t) \\
&= -\delta t \int_0^1 Q(x) \frac{\partial}{\partial x}(a(x)f(x,p,t))\,dx + o(\delta t).
\end{aligned}$$

The third integral needs to be integrated by parts twice:

$$\begin{aligned}
\frac{1}{2}\int_0^1 Q''(x) f(x,p,t) \int_0^1 (\delta x)^2 f(z,x,\delta t)\,dz\,dx & \\
&= \frac{1}{2}\int_0^1 \mathbb{E}((\delta x)^2) Q''(x) f(x,p,t)\,dx \\
&= \frac{1}{2}\int_0^1 \mathbb{V}(\delta x) Q''(x) f(x,p,t)\,dx + o(\delta t) \\
&= \frac{1}{2}\delta t \int_0^1 b(x) Q''(x) f(x,p,t)\,dx + o(\delta t) \\
&= \frac{1}{2}\delta t \Bigg(\left[Q'(x)b(x)f(x,p,t)\right]_0^1 \\
&\qquad - \int_0^1 Q'(x) \frac{\partial}{\partial x}(b(x)f(x,p,t))\,dx \Bigg) + o(\delta t) \\
&= -\frac{1}{2}\delta t \Bigg(\left[Q(x) \frac{\partial}{\partial x}(b(x)f(x,p,t))\right]_0^1 \\
&\qquad - \int_0^1 Q(x) \frac{\partial^2}{\partial x^2}(b(x)f(x,p,t))\,dx \Bigg) + o(\delta t) \\
&= \frac{1}{2}\delta t \int_0^1 Q(x) \frac{\partial^2}{\partial x^2}(b(x)f(x,p,t))\,dx + o(\delta t).
\end{aligned}$$

In the second equality we used the fact that

$$\begin{aligned}
\mathbb{V}(\delta x) &= \mathbb{E}((\delta x)^2) - \mathbb{E}(\delta x)^2 \\
&= \mathbb{E}((\delta x)^2) - (a(x)\delta t + o(\delta t))^2 \\
&= \mathbb{E}((\delta x)^2) + o(\delta t),
\end{aligned}$$

which is true since $(\delta t)^2 = o(\delta t)$. Also, assumption (3.2) was (obviously) used in the third equality. Finally, observe that the last integral is bounded by a constant times $\mathbb{E}(|\delta x|^3)$, which is $o(\delta t)$ by assumption. Hence, putting all the pieces together, we have that

$$\begin{aligned}\int_0^1 Q(z)f(z,p,t+\delta t)\,dz = \int_0^1 Q(x)f(x,p,t)\,dx \\ - \delta t\int_0^1 Q(x)\frac{\partial}{\partial x}(a(x)f(x,p,t))\,dx \\ + \frac{1}{2}\delta t\int_0^1 Q(x)\frac{\partial^2}{\partial x^2}(b(x)f(x,p,t))\,dx + o(\delta t),\end{aligned}$$

which simplifies to

$$\begin{aligned}&\int_0^1 Q(x)(f(x,p,t+\delta t) - f(x,p,t))\,dx \\ &= \delta t\int_0^1 Q(x)\left(-\frac{\partial}{\partial x}(a(x)f(x,p,t)) + \frac{1}{2}\frac{\partial^2}{\partial x^2}(b(x)f(x,p,t))\right)dx + o(\delta t).\end{aligned}$$

Dividing both sides by δt and letting $\delta t \to 0$, we arrive at

$$\int_0^1 Q(x)\frac{\partial f}{\partial t}(x,p,t)\,dx = \int_0^1 Q(x)H(x)\,dx, \tag{3.7}$$

where

$$H(x) = \left(-\frac{\partial}{\partial x}(a(x)f(x,p,t)) + \frac{1}{2}\frac{\partial^2}{\partial x^2}(b(x)f(x,p,t))\right).$$

For the final step, we will need a lemma, known as the fundamental lemma of calculus of variations. The proof presented here is basically the same as the one in [1].

Lemma 3.2. *Let $h \in \mathcal{C}^1([0,1])$, and $Q \in \mathcal{C}^2([0,1])$ be a function such that $Q(0) = Q(1) = Q'(0) = Q'(1) = 0$. If for every such Q,*

$$\int_0^1 Q(x)h(x)\,dx = 0,$$

then $h(x) = 0$ for all $x \in [0,1]$.

Proof. We prove the contrapositive statement, which reads as follows: Assume that there exists a point x_0 where $h(x) \neq 0$ (we may without loss of generality assume that $h(x) > 0$ there). Then there also exists a function $Q \in \mathcal{C}^2([0,1])$ which satisfies the conditions, but makes the integral nonzero.

Since h is continuous, there exists some $\epsilon > 0$ such that $h(x) > 0$ for $|x - x_0| < \epsilon$. For this ϵ, consider the function

$$Q(x) = \begin{cases}(x - x_0 + \epsilon)^2(x - x_0 - \epsilon)^2, & |x - x_0| < \epsilon \\ 0, & |x - x_0| \geqslant \epsilon\end{cases}$$

Clearly, this Q satisfies our conditions and is positive in $|x - x_0| < \epsilon$. However,

$$\int_0^1 Q(x)h(x)\,dx = \int_{x_0-\epsilon}^{x_0+\epsilon} (x - x_0 + \epsilon)^2(x - x_0 - \epsilon)^2 h(x)\,dx > 0$$

since the integrand is positive. □

Since the identity (3.7) is assumed to hold for any Q with the given properties, we can use lemma 3.2 with $h = \frac{\partial f}{\partial t} - H$ to finally arrive at:

Theorem 3.3 (Kolmogorov Forward Equation)**.** *Under the assumptions made in the beginning of this section, the following equation holds*

$$\frac{\partial f}{\partial t}(x, p, t) = -\frac{\partial}{\partial x}(a(x)f(x, p, t)) + \frac{1}{2}\frac{\partial^2}{\partial x^2}(b(x)f(x, p, t)). \tag{3.8}$$

If we focus on the starting position p rather than the current position x (so that $z = p + \delta p$), the Chapman-Kolmogorov equation becomes

$$f(x, p, t + \delta t) = \int_0^1 f(x, z, t)f(z, p, \delta t)\,dz. \tag{3.9}$$

Note that some authors use a different notation for this case (see for example [2] p.118). If we do a Taylor expansion of $f(x, z, t)$ with respect to $z = p + \delta p$ and use the linearity of the integral, we obtain

$$\begin{aligned} f(x, p, t + \delta t) = &\int_0^1 f(x, p, t)f(z, p, \delta t)\,dz + \int_0^1 \delta p \frac{\partial f}{\partial p}(x, p, t)f(z, p, \delta t)\,dz \\ &+ \frac{1}{2}\int_0^1 (\delta p)^2 \frac{\partial^2 f}{\partial p^2}(x, p, t)f(z, p, \delta t)\,dz + \mathcal{O}(|\delta p|^3). \end{aligned}$$

Performing calculations similar to the ones in the case of the forward equation, we see that this is equal to

$$\begin{aligned} &f(x, p, t) + \frac{\partial f}{\partial p}(x, p, t)\mathbb{E}(\delta p) + \frac{1}{2}\frac{\partial^2 f}{\partial p^2}(x, p, t)\big(\mathbb{V}(\delta p) + o(\delta t)\big) + o(\delta t) \\ &\quad = f(x, p, t) + \frac{\partial f}{\partial p}(x, p, t)a(p)\delta t + \frac{1}{2}\frac{\partial^2 f}{\partial p^2}(x, p, t)b(p)\delta t + o(\delta t) \\ &\quad = f(x, p, t) + \delta t\left(a(p)\frac{\partial f}{\partial p}(x, p, t) + \frac{1}{2}b(p)\frac{\partial^2 f}{\partial p^2}(x, p, t)\right) + o(\delta t). \end{aligned}$$

In the second equality we used (3.1) and (3.2) again and merged all the $o(\delta t)$ terms into one. Combining this with (3.9), rearranging, dividing by δt and letting $\delta t \to 0$, we arrive at

Theorem 3.4 (Kolmogorov Backward Equation)**.** *Under the assumptions made in the beginning of this section, the following equation holds*

$$\frac{\partial f}{\partial t}(x, p, t) = a(p)\frac{\partial f}{\partial p}(x, p, t) + \frac{1}{2}b(p)\frac{\partial^2 f}{\partial p^2}(x, p, t). \tag{3.10}$$

3.1.2 Equations for F, P_0 and P_1

For our purposes, it will be useful to have an equation involving

$$F(x,p,t) = \int_0^x f(y,p,t)\,dy$$

rather than f. To find such an equation, we need the following lemma concerning differentiation under the integral sign. A version of this lemma can be found in [6]:

Lemma 3.5. *Let*

$$I(t) = \int_a^b f(y,p,t)\,dy,$$

and assume that f is $\mathcal{C}^1$ in the third argument. Then

$$I'(t) = \int_a^b \frac{\partial f}{\partial t}(y,p,t)\,dy.$$

Proof. Define

$$g(x,p,t) = \frac{f(x,p,t) - f(x,p,s)}{t-s}$$

for $|t-s| < \delta$. By the mean value theorem, there is a $u \in [s,t]$ such that $g(x,p,t) = \frac{\partial f}{\partial t}(x,p,u)$. Since $|u-s| \leqslant |t-s| < \delta$, our assumption gives that

$$\left| g(x,p,t) - \frac{\partial f}{\partial t}(x,p,s) \right| = \left| \frac{\partial f}{\partial t}(x,p,u) - \frac{\partial f}{\partial t}(x,p,s) \right| < \epsilon.$$

Hence $g(x,p,t) \to \frac{\partial f}{\partial t}(x,p,s)$ uniformly as $t \to s$.

Using the uniform convergence of g, we can now compute

$$\begin{aligned} I'(s) = \lim_{t\to s} \frac{I(t)-I(s)}{t-s} = \lim_{t\to s} \int_a^b g(y,p,t)\,dy &= \int_a^b \lim_{t\to s} g(y,p,t)\,dy \\ &= \int_a^b \frac{\partial f}{\partial t}(y,p,s)\,dy. \end{aligned}$$

Since this holds for all s in $|t-s| < \delta$, it holds in particular for t, and we are done. □

With this lemma at hand, we are justified in computing

$$\begin{aligned} \frac{\partial F}{\partial t}(x,p,t) &= \frac{\partial}{\partial t}\left(\int_0^x f(y,p,t)\,dy \right) = \int_0^x \frac{\partial f}{\partial t}(y,p,t)\,dy \\ &= \int_0^x \left(a(p)\frac{\partial f}{\partial p}(y,p,t) + \frac{1}{2}b(p)\frac{\partial^2 f}{\partial p^2}(y,p,t) \right) dy \\ &= a(p)\frac{\partial}{\partial p}\left(\int_0^x f(y,p,t)\,dy \right) + \frac{1}{2}b(p)\frac{\partial^2}{\partial p^2}\left(\int_0^x f(y,p,t)\,dy \right) \\ &= a(p)\frac{\partial F}{\partial p}(x,p,t) + \frac{1}{2}b(p)\frac{\partial^2 F}{\partial p^2}(x,p,t). \end{aligned}$$

Hence F solves the backward equation (3.10) as well. Note that lemma 3.5 was used again on the p-variable in the second to last equality.

Recall from earlier sections that a state x is absorbing if the probability to get out of it is 0. In the language of diffusion processes, this means that $f(z, x, t) = 0$ for all $z \neq x$. Now assume that $x = 0$ and $x = 1$ are absorbing states and consider the probability that absorption has occurred at or before time t for the respective states:

$$\begin{aligned} \mathbf{P}_0(p,t) &= \mathbb{P}(X(T) = 0,\ T \in [0,t] \mid X(0) = p) = F(0,p,t), \\ \mathbf{P}_1(p,t) &= \mathbb{P}(X(T) = 1,\ T \in [0,t] \mid X(0) = p) = F(1,p,t). \end{aligned} \tag{3.11}$$

From this definition it is clear that $\mathbf{P}_0$ and $\mathbf{P}_1$ also solve the backward equation

$$a(p)\frac{\partial \mathbf{P}_0}{\partial p}(p,t) + \frac{1}{2}b(p)\frac{\partial^2 \mathbf{P}_0}{\partial p^2}(p,t) = \frac{\partial \mathbf{P}_0}{\partial t}(p,t). \tag{3.12}$$

Now define $P_0(p) = \lim_{t\to\infty} \mathbf{P}_0(p,t)$. Since $\mathbf{P}_0(p,t)$ satisfies the above equation for all non-negative t, we can take the limit in t on both sides and conclude that the equation is satisfied for $P_0(p)$ as well. Note that this process makes the right hand side equal to zero. Carrying out the same argument for P_1, we can summarize this with the two equations

$$a(p)\frac{dP_0}{dp}(p) + \frac{1}{2}b(p)\frac{d^2P_0}{dp^2}(p) = 0, \tag{3.13}$$

and

$$a(p)\frac{dP_1}{dp}(p) + \frac{1}{2}b(p)\frac{d^2P_1}{dp^2}(p) = 0. \tag{3.14}$$

We also have the intuitively obvious boundary conditions

$$P_0(0) = 1, P_0(1) = 0, P_1(0) = 0, P_1(1) = 1.$$

Now we will solve both equations at once, beginning with the observation that $\frac{dP_0}{dp}(p) = -\frac{dP_1}{dp}(p)$, which is true since $P_0(p) + P_1(p) = 1$ for all p.

With this in mind, set $q(p) = \frac{dP_0}{dp}(p) = -\frac{dP_1}{dp}(p)$. This simplifies equations (3.13) and (3.14) to

$$\frac{dq}{dp}(p) + 2\frac{a(p)}{b(p)}q(p) = 0.$$

Multiplying both sides by the integrating factor, this becomes

$$\frac{d}{dp}\left(\exp\left(2\int_0^p \frac{a(z)}{b(z)}\,dz\right)q(p)\right) = 0,$$

or equivalently

$$q(p) = C\exp\left(-2\int_0^p \frac{a(z)}{b(z)}\,dz\right).$$

The boundary conditions give us that

$$\int_0^1 q(t)\,dt = P_0(1) - P_0(0) = -1,$$

(which means that $C < 0$), and also

$$\int_0^p q(t)\,dt = -P_1(p),\ \int_p^1 q(t)\,dt = -P_0(p).$$

From this, we conclude that

$$P_0(p) = \frac{\int_p^1 q(t)\,dt}{\int_0^1 q(t)\,dt} = \frac{\int_p^1 \psi(t)\,dt}{\int_0^1 \psi(t)\,dt},\ P_1(p) = \frac{\int_0^p q(t)\,dt}{\int_0^1 q(t)\,dt} = \frac{\int_0^p \psi(t)\,dt}{\int_0^1 \psi(t)\,dt}, \tag{3.15}$$

where

$$\psi(t) = \exp\left(-2\int_0^t \frac{a(z)}{b(z)}\,dz\right).$$

3.1.3 Mean absorption time

The goal of this section is to give a formula for the expected time until absorption occurs, starting with the case when both $x = 0$ and $x = 1$ are absorbing states. For this purpose, first consider the density function giving the probability of absorption (in any of the two states) at time t:

$$\begin{aligned}\phi(p,t) &= \mathbb{P}(X(t) = 0 \text{ or } X(t) = 1 \mid X(0) = p)\\ &= \mathbb{P}(X(t) = 0 \mid X(0) = p) + \mathbb{P}(X(t) = 1 \mid X(0) = p)\\ &= f(0,p,t) + f(1,p,t)\end{aligned} \tag{3.16}$$

(for each t, absorption in 0 and 1 are disjoint events). From this definition it is clear that ϕ also solves the backward equation (3.10). Now, the expected time until absorption occurs, given that the process starts in p, is given by the function

$$\bar{t}(p) = \int_0^\infty t\phi(p,t)\,dt. \tag{3.17}$$

Observe that we must have that $\lim_{t\to\infty} t\phi(p,t) = 0$ for the integral to make sense.

Theorem 3.6. *$\bar{t}(p)$ satisfies the boundary value problem*

$$a(p)\frac{d\bar{t}}{dp}(p) + \frac{1}{2}b(p)\frac{d^2\bar{t}}{dp^2}(p) = -1,\ \bar{t}(0) = \bar{t}(1) = 0. \tag{3.18}$$

Proof. Assume that ϕ is nice enough to allow interchange of differentiation and integration (this could be justified for example by the dominated convergence

theorem, see [7]). Then we can use integration by parts and the fact that ϕ solves the backward equation (3.10) to compute:

$$\begin{aligned}
-1 &= -\int_0^\infty \phi(p,t)\,dt = -1 \cdot \left([t\phi(p,t)]_0^\infty - \int_0^\infty t\frac{\partial \phi}{\partial t}(p,t)\,dt \right) \\
&= \int_0^\infty t\left(a(p)\frac{\partial \phi}{\partial p}(p,t) + \frac{1}{2}\frac{\partial^2 \phi}{\partial p^2}(p,t) \right) dt \\
&= a(p)\frac{d}{dp}\left(\int_0^\infty t\phi(p,t)\,dt \right) + \frac{1}{2}b(p)\frac{d^2}{dp^2}\left(\int_0^\infty t\phi(p,t)\,dt \right) \\
&= a(p)\frac{d\bar{t}}{dp}(p) + \frac{1}{2}b(p)\frac{d^2\bar{t}}{dp^2}(p).
\end{aligned}$$

The boundary conditions are intuitively clear: if the process starts in one of the absorbing states, it will not take any additional time to get there! □

Before we actually solve this problem, we recall some facts from the theory of ordinary differential equations:

Definition 3.1. *The Wronskian $W(y_1, y_2)$ of two solutions to a second order ordinary differential equation is defined as*

$$W(y_1,y_2)(x) = \det \begin{pmatrix} y_1(x) & y_2(x) \\ y_1'(x) & y_2'(x) \end{pmatrix} = y_1(x)y_2'(x) - y_1'(x)y_2(x).$$

For the Wronskian, we have the following useful identity:

Lemma 3.7 (Abel)**.** *Let y_1, y_2 be solutions of the ordinary differential equation*

$$y''(x) + p(x)y'(x) + q(x)y(x) = 0$$

for $x \in I \subset \mathbb{R}$. Then

$$W(y_1,y_2)(x) = W(y_1,y_2)(x_0)\exp\left(-\int_{x_0}^x p(t)\,dt \right)$$

for every point $x_0 \in I$.

Proof. We write $W(x)$ instead of $W(y_1, y_2)(x)$ for simplicity. Using that y_1 and y_2 solves the differential equation, we can find a differential equation for $W(x)$:

$$\begin{aligned}
W'(x) &= y_1'(x)y_2'(x) + y_1(x)y_2''(x) - y_1''(x)y_2(x) - y_1'(x)y_2'(x) \\
&= -y_1(x)(p(x)y_2'(x) + q(x)y_2(x)) + y_2(x)(p(x)y_1'(x) + q(x)y_1(x)) \\
&= -p(x)(y_1(x)y_2'(x) - y_1'(x)y_2(x)) + q(x)(-y_1(x)y_2(x) + y_1(x)y_2(x)) \\
&= -p(x)W(x).
\end{aligned}$$

This can be solved using the method of integrating factor, which gives us that

$$W(x) = C\exp\left(-\int_{x_0}^x p(z)\,dz \right).$$

Setting $x = x_0$, we conclude that $C = W(x_0)$. □

Lemma 3.7 shows that the Wronskian is either always zero (in which case the solutions y_1, y_2 are linearly dependent) or never zero. The following two results can be phrased more elegantly if we introduce the notation

$$\mathcal{L} = \left(a_2(x)\frac{d^2}{dx^2} + a_1(x)\frac{d}{dx} + a_0(x) \right),$$

$$\mathcal{B} = \begin{pmatrix} b_{11}^a & b_{12}^a & b_{11}^b & b_{12}^b \\ b_{21}^a & b_{22}^a & b_{21}^b & b_{22}^b \end{pmatrix},$$

$$\bar{y}_{a,b} = (y(a), y'(a), y(b), y'(b))^T.$$

A general boundary value problem can then be written

$$\begin{cases} \mathcal{L}y = f \\ \mathcal{B}\bar{y}_{a,b} = \bar{c} \end{cases}, \tag{3.19}$$

where $\bar{c} = (c_1, c_2)$. Typically, most elements of $\mathcal{B}$ is 0. For example, if $b_{11}^a = b_{21}^b = 1$ and the rest is 0, we get Dirichlet conditions $y(a) = y(b) = 0$, and if $b_{12}^a = b_{22}^b = 1$ and the rest is 0, we get Neumann conditions $y'(a) = y'(b) = 0$.

Definition 3.2. *The Green function of the boundary value problem* $\mathcal{L}y = f, \mathcal{B}\bar{y}_{a,b} = \bar{0}$ *is defined as*

$$G(x,\xi) = \begin{cases} \dfrac{y_2(\xi)y_1(x)}{a_2(\xi)W(y_1,y_2)(\xi)}, & a \leqslant \xi \leqslant x \leqslant b \\ \dfrac{y_2(x)y_1(\xi)}{a_2(\xi)W(y_1,y_2)(\xi)}, & a \leqslant x \leqslant \xi \leqslant b \end{cases}, \tag{3.20}$$

where y_1 *and* y_2 *solve the corresponding homogeneous equation* $\mathcal{L}y = 0$ *with one boundary condition each and* $W(y_1, y_2)$ *is their Wronskian.*

The next result will be crucial for our computations.

Theorem 3.8. *The boundary value problem* (3.19) *has a unique solution for every* $f \in \mathcal{C}([a,b])$ *if and only if the homogeneous problem* ($f \equiv 0$) *only has the trivial solution* $y \equiv 0$. *In that case, the solution is given by*

$$y(x) = \int_a^b G(x,\xi)f(\xi)\,d\xi. \tag{3.21}$$

Going back to our diffusion process, here is the main result of this section, which is stated in [2] without any proof or reference:

Theorem 3.9. *Equation* (3.18) *has a unique solution, given by*

$$\bar{t}(p) = \int_0^1 t(x,p)\,dx,$$

where

$$t(x,p) = \begin{cases} \dfrac{2P_0(p)}{b(x)\psi(x)} \int_0^x \psi(y)\,dy, & 0 \leqslant x \leqslant p \leqslant 1 \\ \dfrac{2P_1(p)}{b(x)\psi(x)} \int_x^1 \psi(y)\,dy, & 0 \leqslant p \leqslant x \leqslant 1 \end{cases}. \tag{3.22}$$

Proof. The result follows easily from Theorem 3.8 once we make the right identifications. It is clear that $\frac{1}{2}b(x)$ corresponds to the function $a_2(x)$ in the definition on the Green function. From (3.13) and (3.14) we see that P_1 and P_0 solves the homogeneous version of (3.18), with the required boundary conditions $P_1(0) = 0, P_0(1) = 0$. Hence P_1 and P_0 correspond to the functions y_1 and y_2.

It remains to compute the Wronskian of P_1 and P_0. We have that

$$\begin{aligned} W(P_1, P_0)(0) &= \det \begin{pmatrix} P_1(0) & P_0(0) \\ P_1'(0) & P_0'(0) \end{pmatrix} \\ &= \det \begin{pmatrix} 0 & 1 \\ (\int_0^1 \psi(y)\,dy)^{-1} & -(\int_0^1 \psi(y)\,dy)^{-1} \end{pmatrix} \\ &= -\left(\int_0^1 \psi(y)\,dy \right)^{-1}, \end{aligned}$$

since $P_1'(x) = \psi(x)(\int_0^1 \psi(y)\,dy)^{-1}$ and $P_1'(x) = -P_0'(x)$. Hence by lemma 3.7, the Wronskian is

$$\begin{aligned} W(P_1, P_0)(x) &= -\left(\int_0^1 \psi(y)\,dy \right)^{-1} \exp\left(-2\int_0^x \frac{a(z)}{b(z)}\,dz \right) \\ &= -\psi(x) \left(\int_0^1 \psi(y)\,dy \right)^{-1}. \end{aligned}$$

With this, we conclude that

$$\begin{aligned} G(x,\xi) &= \begin{cases} \dfrac{P_1(\xi)P_0(x)}{\frac{b(\xi)}{2}W(P_1,P_0)(\xi)}, & 0 \leqslant \xi \leqslant x \leqslant 1 \\ \dfrac{P_0(\xi)P_1(x)}{\frac{b(\xi)}{2}W(P_1,P_0)(\xi)}, & 0 \leqslant x \leqslant \xi \leqslant 1 \end{cases} \\ &= \begin{cases} -\dfrac{2P_0(x)}{b(\xi)\psi(\xi)} \int_0^\xi \psi(y)\,dy, & 0 \leqslant \xi \leqslant x \leqslant 1 \\ -\dfrac{2P_1(x)}{b(\xi)\psi(\xi)} \int_\xi^1 \psi(y)\,dy, & 0 \leqslant x \leqslant \xi \leqslant 1 \end{cases} \end{aligned}$$

is the Green function of the boundary value problem (3.18). Since $f(x) \equiv -1$ in our case, the solution of the boundary value problem is given by

$$\begin{aligned} \bar{t}(x) &= \int_0^1 G(x,\xi) f(\xi)\,d\xi \\ &= \int_0^x \frac{2P_0(x)}{b(\xi)\psi(\xi)} \int_0^\xi \psi(y)\,dy\,d\xi + \int_x^1 \frac{2P_1(x)}{b(\xi)\psi(\xi)} \int_\xi^1 \psi(y)\,dy\,d\xi, \end{aligned} \tag{3.23}$$

which was to be proven. This is equivalent to the solution stated in [2] after renaming the variables.

To prove the uniqueness of the solution, Theorem 3.8 says that all we need to do is to show that the homogeneous problem only has the trivial solution $\bar{t}(p) \equiv 0$.

Setting $q(p) = \frac{d\bar{t}}{dp}(p)$, we recover the equation

$$\frac{dq}{dp}(p) + 2\frac{a(p)}{b(p)}q(p) = 0,$$

with the solution

$$q(p) = C\psi(p).$$

This, together with the condition $\bar{t}(0) = 0$ gives that

$$\bar{t}(p) = C\int_0^p \psi(t)\,dt.$$

However, since $\psi > 0$, the other condition $\bar{t}(1) = 0$ gives that $C = 0$, so that $\bar{t}(p) \equiv 0$. □

In the case of only one absorbing state, many of the same arguments can be used as in the case of two absorbing states. We will only go through the case where $x = 0$ is the absorbing state here, since that will be the case of interest later on. In this case, $\bar{t}(p)$ will still solve equation (3.18), but needs to be redefined as

$$\bar{t}(p) = \int_0^\infty t f(0, p, t)\,dt,$$

and have the new boundary conditions $\bar{t}(0) = 0, \bar{t}'(1) = 0$. To find the solution to this problem, we will once again use Theorem 3.8.

Just as in the previous case, the function $\frac{1}{2}b(x)$ corresponds to $a_2(x)$. Also, the function P_1 still solves the homogeneous version of (3.18) and the boundary condition $P_1(0) = 0$ (of course, we no longer interpret P_1 as a probability, since it would have to be 0 in this case!). For the second solution, we simply pick the constant function $1(x) \equiv 1$ (any constant function would work, since the derivative of such a function is zero everywhere). The Wronskian becomes

$$W(P_1, 1)(x) = -P_1'(x) = -\psi(x)\left(\int_0^1 \psi(y)\,dy\right)^{-1}.$$

Hence the Green function of this problem is given by

$$\begin{aligned} G(x,\xi) &= \begin{cases} \dfrac{P_1(\xi)}{\frac{b(\xi)}{2}W(P_1,1)(\xi)}, & 0 \leqslant \xi \leqslant x \leqslant 1 \\ \dfrac{P_1(x)}{\frac{b(\xi)}{2}W(P_1,1)(\xi)}, & 0 \leqslant x \leqslant \xi \leqslant 1 \end{cases} \\ &= \begin{cases} -\dfrac{2}{b(\xi)\psi(\xi)}\int_0^\xi \psi(y)\,dy, & 0 \leqslant \xi \leqslant x \leqslant 1 \\ -\dfrac{2}{b(\xi)\psi(\xi)}\int_0^x \psi(y)\,dy, & 0 \leqslant x \leqslant \xi \leqslant 1 \end{cases}, \end{aligned} \tag{3.24}$$

and the solution of the boundary value problem is given by

$$\begin{aligned}\bar{t}(x) &= \int_0^1 G(x,\xi) f(\xi)\, d\xi \\ &= \int_0^x \frac{2}{b(\xi)\psi(\xi)} \int_0^\xi \psi(y)\, dy\, d\xi + \int_x^1 \frac{2}{b(\xi)\psi(\xi)} \int_0^x \psi(y)\, dy\, d\xi,\end{aligned} \tag{3.25}$$

where $f(x) \equiv -1$ as before. As in the previous case, this solution is equivalent to the one stated in [2] once we rename the variables.

Showing that the solution is unique is no harder in this case. Arguing as in the previous case, we obtain

$$\bar{t}(p) = C \int_0^p \psi(t)\, dt,$$

and the new condition $\bar{t}'(1) = 0$ gives that $C = 0$.

3.1.4 Higher moments

One can use the same approach as in the derivation of (3.18) to compute higher moments of the absorbtion time. Under the (increasingly strong) assumption that

$$\lim_{t\to\infty} t^n \phi(p,t) = 0,$$

the n:th moment is given by the function

$$\bar{t}^{(n)}(p) = \int_0^\infty t^n \phi(p,t)\, dt,$$

and it satisfies the differential equation

$$a(p)\frac{d\bar{t}^{(n)}}{dp}(p) + \frac{1}{2}b(p)\frac{d^2\bar{t}^{(n)}}{dp^2}(p) = -n\bar{t}^{(n-1)}(p), \tag{3.26}$$

plus the boundary conditions $\bar{t}^{(n)}(0) = \bar{t}^{(n)}(1)$ (in the case of two absorbing states) or $\bar{t}^{(n)}(0) = \bar{t}^{(n)\prime}(1) = 0$ (in the case of one absorbing state). The boundary conditions for the higher moments come from the boundary conditions for $\bar{t}(p)$, since for example $\bar{t}(0) = 0$ requires that $\phi(0,t) = 0$, which in turn affects the value of $\bar{t}^{(n)}(0)$.

For example, the variance is given by

$$\mathbb{V}(t \mid p) = \bar{t}^{(2)}(p) - \bar{t}(p)^2,$$

where the second moment $\bar{t}^{(2)}(p)$ solves

$$a(p)\frac{d\bar{t}^{(2)}}{dp}(p) + \frac{1}{2}b(p)\frac{d^2\bar{t}^{(2)}}{dp^2}(p) = -2\bar{t}(p), \tag{3.27}$$

plus one of the two types of boundary conditions.

Note that the boundary value problems for higher moments will all have the same Green function as the boundary value problem for $\bar{t}(p)$, since the left hand sides of their differential equations do not change with n. The difference in the solutions of the boundary problems comes from the different left hand sides, the f in definition 3.2. Therefor, we can easily follow a similar procedure as in the previous subsection, and deduce that the second moment has the form:

$$\begin{aligned}\bar{t}^{(2)}(x) &= \int_0^1 G(x,\xi) f(\xi)\, d\xi \\ &= 4\left(\int_0^x \frac{\bar{t}(\xi)}{b(\xi)\psi(\xi)} \int_0^\xi \psi(y)\, dy\, d\xi + \int_x^1 \frac{\bar{t}(\xi)}{b(\xi)\psi(\xi)} \int_0^x \psi(y)\, dy\, d\xi\right),\end{aligned} \tag{3.28}$$

where $\bar{t}(x)$ has the form (3.25) or (3.23) depending on whether the process has one or two absorbing states respectively.

Although in theory it is possible to find formulas for these higher moments by the above procedure, the formulas become increasingly complicated, since the n:th moment depends on complete knowledge of the previous $n-1$ moments, and is given by a very complicated looking $2n$-dimensional integral.

3.2 Our case

3.2.1 Mean

The quantity we are ultimately interested in is the function (3.25) given in the previous section, evaluated at $x = 1$ (so that we start with nothing but A_1 alleles) and multiplied by N (so that the answer is in number of generations). Hence, our original mean absorbtion time can be approximated by

$$N\bar{t}(1) = N \int_0^1 \frac{2}{b(x)\psi(x)} \int_0^x \psi(y)\, dy\, dx \tag{3.29}$$

(note that the second half of the Green function vanishes).

The first step towards expanding (3.29) is to find the functions $a(x)$ and $b(x)$ in our model. In section 2.2 we found that

$$\mathbb{E}(\delta x) = -ux + sx(1-x)(x-1+h(1-2x)) + \mathcal{O}(N^{-2}),$$

$$\mathbb{V}(\delta x) = \frac{1}{N}x(1-x) + \mathcal{O}(N^{-2}).$$

This, together with the assumptions (3.1) and (3.2), reveals that

$$a(x) = -Nux + Nsx(1-x)(x-1+h(1-2x)), \tag{3.30}$$

$$b(x) = x(1-x). \tag{3.31}$$

Using (3.30) and (3.31), we can now also find our $\psi(x)$ as

$$\begin{aligned}
\psi(x) &= \exp\left(-2\int_0^x \frac{-Nuz + Nsz(1-z)(h-1+z(1-2h))}{z(1-z)}\, dz\right) \\
&= \exp\left(-2\int_0^x \left(-\frac{Nu}{1-z} + Ns(h-1+z(1-2h))\right) dz\right) \\
&= \exp\left(-2Nu\log(1-x) - 2Nshx + 2Nsx - Ns(1-2h)x^2\right) \\
&= (1-x)^{-2Nu}\exp(2Ns(1-h)x - Ns(1-2h)x^2).
\end{aligned}$$

This allows us to expand $N\bar{t}(1)$ in all of its glory:

$$\begin{aligned}
N\bar{t}(1) &= N\int_0^1 \frac{2}{x(1-x)}(1-x)^{2Nu}\exp(-2Ns(1-h)x + Ns(1-2h)x^2) \\
&\qquad \times \int_0^x (1-y)^{-2Nu}\exp(2Ns(1-h)y - Ns(1-2h)y^2)\, dy\, dx \\
&= 2N\int_0^1 \frac{1}{x}(1-x)^{2Nu-1}\int_0^x (1-y)^{-2Nu} e(x,y)\, dy\, dx,
\end{aligned}$$

where $e(x,y) = \exp\big(-2Ns(x-y)\big((1-h)-(\frac{1}{2}-h)(x+y)\big)\big)$.

From this formula we see why $h = \frac{1}{2}$ and $h = 1$ are interesting cases. More importantly, we note that the integrand appears to be singular in both endpoints. However, the singularity at 0 is removable, since

$$\lim_{x\to 0}\frac{1}{x}\int_0^x \psi(y)dy = \lim_{x\to 0}\psi(x) = \psi(0) = 1,$$

by l'Hôpitals rule. It is nonetheless tempting to do the change of variables

$$\xi = x,\ \eta = \frac{y}{x},$$

which transforms the integral to

$$N\bar{t}(1) = 2N\int_0^1 (1-\xi)^{2Nu-1}\int_0^1 (1-\eta\xi)^{-2Nu}e(\xi,\eta\xi)\,d\eta\,d\xi. \tag{3.32}$$

The numerical evaluation of this integral is one of the main points of this thesis, and the results can be found in subsection 4.2.2.

The computation of (3.32) simplifies significantly if we set $s = 0$. Mathematically, this provides an upper bound of the integral, since a quick analysis reveals that $0 \leqslant e(x,y) \leqslant 1$ and $e(x,y) = 1$ only if $s = 0$. In genetic terms, $s = 0$ means that the A_2 allele has no selective advantage over A_1, so it can only go to fixation by mutation. With a very small mutation probability, it is expected to take a very large number of generations before fixation occurs, and the largest number of generations (in mean) is $N\bar{t}(1)\big|_{s=0}$. We state the result as a theorem:

Theorem 3.10.

$$N\bar{t}(1)|_{s=0} = \begin{cases} \frac{2N}{2Nu-1}(\Psi(2Nu)-\Psi(1)) & \text{if } 2Nu \neq 1 \\ 2N\Psi'(1) & \text{if } 2Nu = 1 \end{cases}, \tag{3.33}$$

where $\Psi(x) = \frac{d}{dx}\log(\Gamma(x))$ *is the digamma function.*

Proof. We start by simply evaluating the integral:

$$\begin{aligned}
N\bar{t}(1)\big|_{s=0} &= 2N\int_0^1 (1-\xi)^{2Nu-1}\int_0^1 (1-\eta\xi)^{-2Nu}\,d\eta\,d\xi \\
&= 2N\int_0^1 (1-\xi)^{2Nu-1}\left[\frac{1}{2Nu-1}\frac{1}{\xi}(1-\eta\xi)^{1-2Nu}\right]_0^1 d\xi \\
&= \frac{2N}{2Nu-1}\int_0^1 \frac{1}{\xi}\left(1-(1-\xi)^{2Nu-1}\right)\,d\xi \\
&= \frac{2N}{2Nu-1}\int_0^1 \sum_{k=1}^{2Nu-1}\binom{2Nu-1}{k}(-1)^{k-1}\xi^{k-1}\,d\xi \\
&= 2N\sum_{k=0}^{2Nu-2}\binom{2Nu-2}{k}\frac{(-1)^k}{k+1}\int_0^1 \xi^k\,d\xi \\
&= 2N\sum_{k=0}^{2Nu-2}\binom{2Nu-2}{k}\frac{(-1)^k}{(k+1)^2}.
\end{aligned}$$

A sum on this form can be computed explicitly with the aid of the digamma function $\Psi(x) = \frac{d}{dx}\log\Gamma(x)$ and the harmonic numbers $H_n = \sum_{k=1}^{n}\frac{1}{k}$. First note that the digamma function follows the recursion

$$\begin{aligned}
\Psi(x+1) = \frac{d}{dx}\left(\log\Gamma(x+1)\right) &= \frac{d}{dx}\left(\log x\Gamma(x)\right) \\
&= \frac{d}{dx}\left(\log(x)+\log\Gamma(x)\right) = \frac{1}{x}+\Psi(x),
\end{aligned}$$

so that, for integer n,

$$\Psi(n+1) = \Psi(n)+\frac{1}{n} = \ldots = \Psi(1)+H_n.$$

Next, note that H_n can be written as:

$$H_n = \sum_{k=0}^{n-1}\frac{1}{k+1} = \sum_{k=0}^{n-1}\int_0^1 x^k\,dx = \int_0^1\sum_{k=0}^{n-1}x^k\,dx = \int_0^1\frac{1-x^n}{1-x}\,dx,$$

which after the change of variables $x \mapsto 1-x$ becomes

$$\int_0^1 \frac{1}{x}\left(1-(1-x)^n\right)\,dx.$$

This integral is very similar to what we had after the third equality in the computation of $N\bar{t}(1)\big|_{s=0}$, so if we do a similar manipulation, we see that

$$H_n = n\sum_{k=0}^{n-1}\binom{n-1}{k}\frac{(-1)^k}{(k+1)^2}.$$

Now, if we set $n = 2Nu - 1$ and put all the pieces together, we get

$$\Psi(2Nu) - \Psi(1) = (2Nu - 1) \sum_{k=0}^{2Nu-2} \binom{2Nu-2}{k} \frac{(-1)^k}{(k+1)^2},$$

so that, finally

$$\bar{t}(1)\big|_{s=0} = \frac{2N}{2Nu-1}(\Psi(2Nu) - \Psi(1)). \tag{3.34}$$

Observe that if $2Nu = 1$, this needs to be interpreted as the limit

$$\lim_{x\to 1} \frac{2N}{x-1}(\Psi(x) - \Psi(1)) = 2N\Psi^{'}(1),$$

where $\Psi'(x)$ is known as the polygamma function of order one. □

The expression (3.33) has the advantage of being much easier to compute numerically than the integral form (or the sum for that matter).

Remark 3.11. *In fact, $\Psi(1) = -\gamma$, where γ is the Euler-Mascheroni constant*

$$\gamma = \lim_{n\to\infty} (H_n - \log(n)) \approx 0.577.$$

Since $\Gamma(n+1) = n!$, the easiest way to prove this is to use the identity

$$\Psi(1) = \Psi(n+1) - H_n, \tag{3.35}$$

and Stirling's approximation

$$n! \sim \sqrt{2\pi n}\left(\frac{n}{e}\right)^n.$$

For large n, the right hand side of (3.35) is approximately

$$\begin{aligned}\frac{d}{dn}\left(\log\left(\sqrt{2\pi n}(\frac{n}{e})^n\right)\right) - H_n &= \frac{d}{dn}\left(\log(\sqrt{2\pi}) + (n+\frac{1}{2})\log(n) - n\right) - H_n \\ &= \log(n) + \frac{1}{2n} - H_n.\end{aligned}$$

The identity becomes exact when $n \to \infty$, and then the above expression tends to the definition of $-\gamma$.

3.2.2 Variance

Recall from section 3.1.4 that it is also (in theory) possible to compute higher moments of the absorbtion time using similar methods. Here we will only consider the variance of the absorbtion time, given that we start with only A_1 alleles. This is given by

$$\mathbb{V}(T_0) = \bar{t}^{(2)}(1) - (\bar{t}(1))^2, \tag{3.36}$$

where

$$\bar{t}^{(2)}(1) = 4 \int_0^1 \frac{\bar{t}(\xi)}{b(\xi)\psi(\xi)} \int_0^\xi \psi(y)\, dy\, d\xi \tag{3.37}$$

is (3.28) evaluated in $x = 1$, and $\bar{t}(\xi)$ is given by (3.25). In theory, (3.36) should serve as an approximation of (2.8), but due to its complicated nature, the author was not able to compute it numerically for any values of N.[12] It is therefore not clear how good of an approximation it is.

One possible approach to compute (3.36) could be to approximate $\bar{t}(x)$ by a simpler function, for example a polynomial. That would transform (3.37) into a sum of integrals, each of which would be similar to (3.32), which we know how to integrate numerically. However, to approximate $\bar{t}(x)$ with a polynomial might be hard since the graph of $\bar{t}(x)$ is highly skewed (see figure 4). Another approach could be to only consider the case $s = 0$, which hopefully leads to some simplifications.[13] In conclusion, more knowledge of numerical methods, or dramatic simplifications, is required before the formulas (3.36) and (3.37) can become useful in practice.

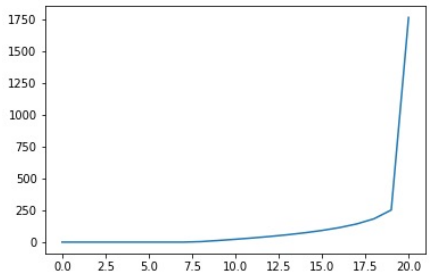

Figure 4: $\bar{t}(x)$ evaluated at 20 equally spaced points, with $N = 10^6$, $s = 0.02,\ h = 0.5,\ u = 5 \cdot 10^{-6}$.

[12] This is not that surprising, since numerical integration requires one to evaluate the function at a large amount of points, and our function $\bar{t}(x)$ is hard enough to evaluate at only a handful of points.

[13] In fact, in the case $s = 0$, the first part of $\bar{t}(x)$ can be written in terms of the generalized hypergeometric function as $2xF(1, 1, 1 - 2Nu, 2, 2, x)$, which can easily be integrated numerically. However, for the second part, the author has not found any simplifying expression.

4 Numerical Experiments

This section concerns numerical computations of the quantities derived in earlier sections. The numerical computations have two purposes: one genetic and one mathematical. The genetic purpose of the computations is to get some idea of the distribution of the absorbtion time, and see how it depends on the parameters N, s, h and u. The mathematical purpose is to see how well the mean absorption time, computed via the diffusion process, approximates the same quantity computed via the Markov process. It is also interesting to see how large the population size N can be before the computations become troublesome or impossible.

4.1 Method

For the numerical computations, Python 3.7.0 was used with Spyder, on a 2017 Lenovo ideapad 320 (unless otherwise noted). The packages scipy.linalg, scipy.integrate and scipy.special was imported. In the Markov chain case, (2.3) was used for computation of the mean, and the square root of (2.8) was used for the standard deviation. For both quantities, the solve() method was used to compute the required inverses. In the diffusion process case, (3.32) was used, and computed using the quad() method twice (for $N \leqslant 10^5$) and dblquad() (for $N \geqslant 10^6$). Missing values in the tables means either that the number could not be computed at all (within reasonable time), or that the result was nonsense due to numerical issues.

4.2 Results

To make the results more readable, recall the table from section 1 explaining the parameters:

Parameter	Meaning	Values
M	Number of individuals in the population.	$0.5 \cdot 10^k,\ k \in \mathbb{N}$
N	Number of gametes (gene copies). $N = 2M$	$10^k,\ k \in \mathbb{N}$
s	Models how much "better" the A_2 allele is.	$0.02,\ 0.1$
h	Models the dominance between A_1 and A_2.	$0,\ 0.5,\ 1$
u	Probability of mutation $A_1 \mapsto A_2$.	$5 \cdot 10^{-6}$

Table 3: Explanation of the parameters.

4.2.1 Markov Chain Computations

We now present some reults of computing the mean absorbtion time

$$\mathbb{E}(T_0|X(0) = 1)$$

numerically. The first two tables were computed by Pelle Pettersson [5] using Matlab. For $N \geqslant 10^5$, more sophisticated numerical methods (sparse matrix methods), as well as a more powerful computer was used, since the very large matrices involved causes memory problems. We choose to provide his numbers instead of the thesis authors own results, since more available numbers makes it easier to compare the two different methods later on.

N	$h = 0$	$h = \frac{1}{2}$	$h = 1$
10	187930	183233	178738
10^2	120163	87957	66114
10^3	40235	10850	6192
10^4	13841	2179	2950
10^5	6828	1648	6884
10^6	5814	1828	20183

Table 4: Mean absorbtion time with $s = 0.02$.

N	$h = 0$	$h = \frac{1}{2}$	$h = 1$
10	151662	134708	120944
10^2	56676	21600	12278
10^3	17987	2349	1514
10^4	6111	521	1150
10^5	2937	408	3065
10^6	-	-	-

Table 5: Mean absorbtion time with $s = 0.1$.

As expected, a larger selection coefficient leads to shorter absorption time in general, but apart from this, we can see similar patterns in the two different cases. Since we start with only A_1 alleles, we need to wait until two mutations have occurred before selection starts working in favor of A_2A_2. If there has been only one mutation, there is a nonzero probability that the single A_2 thus obtained is ruled out by drift, in which case we are back where we started and need to wait for two more mutations. This explains why the numbers are larger for smaller populations, since mutations are less frequent then.

Now let us first look at $h = 1$, in which case the A_1A_2 and A_2A_2 genotypes have equal selective advantage. One needs to wait until the first two mutations have occurred, but after that, selection kicks in and quickly increases the number of A_2 alleles. With increasing N, we can expect more mutations and

hence decreased absorption times. However, we can also notice another effect after $N = 10^4$, when the numbers starts to increase. Note that selection only works against the A_1 alleles that come in pairs. This means that two individuals having the A_1 allele need to mate and produce offspring carrying the A_1A_1 genotype before those two A_1 alleles can be lost from the population by selection. In larger populations, mating of two such individuals quickly becomes rare, since the fraction of A_1 alleles in the population quickly decreases.

Apart from selection working against the A_1A_1 genotype, A_1 alleles can only be lost from the population by mutation or drift. To sum up: in the case $h = 1$ and in populations of size 10^4 or larger, there will for a long time remain a small number of A_1 alleles which are hard to get rid of completely. We will gather more numerical evidence for this phenomenon later.

In the case $h = 0$, the genotypes A_1A_1 and A_1A_2 are equally bad in comparison to A_2A_2. Since we start with only A_1 alleles, the first A_2 alleles to enter the population does so in the form of a A_1A_2 genotype. However, a single such A_2 allele might be ruled out by selection, since selection is now working equally against the A_1A_1 and A_1A_2 genotypes! In other words: selection, which works in favor of the A_2A_2 genotype, might also get rid of single A_2 alleles "by mistake". Also, as we mentioned earlier, single A_2 alleles obtained through mutation could also be ruled out by drift. These factors together explain why we see the largest numbers in this case, up until $N = 10^5$ (at which point another phenomenon starts to increase the numbers in the case $h = 1$). The numbers are at their peak when $N = 10$, because for such a small population, the loss of only one A_2 allele makes a big difference.

Lastly, in the case $h = \frac{1}{2}$, the numbers lie somewhere in between the two other cases. This is because the relative weight of the A_1A_2 is the average of the other two weights, and so the absorption time is affected by all the effects previously mentioned.

Next we provide tables of the standard deviation (computed using (2.8)), and also include the mean, to hopefully gain some more insight on the distribution. The standard deviations turned out to be more convenient and illustrative to use, since the variances turned out to be very large.

N	$h = 0$	$h = \frac{1}{2}$	$h = 1$
10	(187930, 187913)	(183233, 183215)	(178738, 178720)
10^2	(120163, 119992)	(87957, 87770)	(66114, 65907)
10^3	(40235, 39541)	(10850, 10148)	(6192, 5293)
10^4	(13842, 12032)	(2179, 1030)	(2950, 1139)

Table 6: Mean and standard deviation of absorbtion time, with $s = 0.02$.

N	$h = 0$	$h = \frac{1}{2}$	$h = 1$
10	(151662, 151645)	(134708, 134691)	(120944, 120926)
10^2	(56676, 56576)	(21600, 21488)	(12278, 12142)
10^3	(17987, 17705)	(2349, 2136)	(1514,1152)
10^4	(6111, 5376)	(521, 217)	(1150, 471)

Table 7: Mean and standard deviation of absorbtion time, with $s = 0.1$.

We note that in most cases, the means and standard deviations are very similar, especially for small populations. This suggests that in those cases, the absorbtion time follows a geometric distribution, $T_0 \sim Ge(p)$, since for such a distribution,

$$\mu = \frac{1}{p},\ \sigma = \frac{\sqrt{1-p}}{p},$$

and these numbers are very close if p is small, which is certainly true in our case, with $p = u$. Since we start with only A_1 alleles, the fixation of the A_2 allele can only happen after a mutation has occurred. Hence most of the time before absorption is spent waiting for a first mutation to occur, which is exactly the type of situation which the geometric distribution is used to model. We should not say too much, though. If the absorption time has geometric distribution with $p = u$, the mean would be $\frac{1}{u} = 200000$, which we see is not the case. This makes it clear that the mutation probability is not the only thing that affects the absorption time. Also, as N increases, the standard deviations decrease and no longer stay close to the mean, which means that the distribution becomes more centered around its mean. Hence the mean says more about the distribution in those cases.

To gain even more insight about the distribution of the absorption time, we can try to see what happens with the mean and standard deviation when the mutation probability u is increased by powers of 10. Intuitively, with a higher mutation probability, the first two mutations are likely to occur faster, and as we mentioned earlier, once the first two mutations have occurred, the remaining

time before absorption is rather short. These two things together suggests that the mean absorption time should be smaller in this case. Apart from this, we expect to see the same patterns that we discussed earlier.

N	$h = 0$	$h = \frac{1}{2}$	$h = 1$
10	(18809, 18791)	(18340, 18322)	(17891, 17873)
10^2	(12156, 11986)	(8985, 8799)	(6837, 6633)
10^3	(4504, 3836)	(1708, 1038)	(1429,630)
10^4	(2392, 982)	(1179, 188)	(2406, 989)

Table 8: Mean and standard deviation of absorption time, with $s = 0.02$ and $u = 5 \cdot 10^{-5}$.

N	$h = 0$	$h = \frac{1}{2}$	$h = 1$
10	(15181, 15164)	(13488, 13471)	(12114, 12096)
10^2	(5742, 5642)	(2262, 2151)	(1351, 1217)
10^3	(1981, 1709)	(424, 217)	(484, 191)
10^4	(996, 443)	(311, 39)	(1132, 452)

Table 9: Mean and standard deviation of absorption time, with $s = 0.1$ and $u = 5 \cdot 10^{-5}$.

N	$h = 0$	$h = \frac{1}{2}$	$h = 1$
10	(1896, 1879)	(1850, 1832)	(1806, 1788)
10^2	(1349, 1189)	(1077, 906)	(897, 914)
10^3	(853, 340)	(694, 194)	(856, 341)
10^4	(862, 108)	(865, 132)	(1949, 777)

Table 10: Mean and standard deviation of absorption time, with $s = 0.02$ and $u = 5 \cdot 10^{-4}$.

N	$h = 0$	$h = \frac{1}{2}$	$h = 1$
10	(1533, 1516)	(1366, 1349)	(1231, 1213)
10^2	(646, 551)	(326, 221)	(256, 138)
10^3	(355, 142)	(213, 40)	(364, 147)
10^4	(329, 38)	(251, 28)	(935, 403)

Table 11: Mean and standard deviation of absorption time, with $s = 0.1$ and $u = 5 \cdot 10^{-4}$.

The results are as expected. The means have decreased by roughly a factor of 10, which is the same factor that we increased u with. This is another indication that the value u plays an important role in the value of the mean absorption time.

4.2.2 Diffusion Process Computations

Now we provide computed values of

$$N\bar{t}(1) = 2N \int_0^1 (1-\xi)^{2Nu-1} \int_0^1 (1-\eta\xi)^{-2Nu} e(\xi, \eta\xi)\, d\eta\, d\xi$$

from section 3.2, which should serve as an approximation of $\mathbb{E}(T_0|X(0) = N)$, and should be computable for larger N. The following two tables should be compared with tables 4 and 5 from the previous subsection.

N	$h = 0$	$h = \frac{1}{2}$	$h = 1$
10	187450	181290	175372
10^2	119786	86680	64255
10^3	40178	10707	6047
10^4	13831	2159	2919
10^5	6822	1635	6822
10^6	5807	1753	19652
10^7	2826	1019	58989
10^8	544	9985	167524
10^9	-	1093	406404
10^{10}	-	-	778228

Table 12: Mean absorbtion time when $s = 0.02$.

N	$h = 0$	$h = \frac{1}{2}$	$h = 1$
10	149383	126446	107642
10^2	56131	20112	10743
10^3	17927	2207	1365
10^4	6099	496	1094
10^5	2928	391	2834
10^6	2354	220	8769
10^7	101	203	27101
10^8	109	217	81459
10^9	-	-	223312
10^{10}	-	-	507096

Table 13: Mean absorbtion time when $s = 0.1$.

From these tables it is clear that the mean absorption time can be computed for larger N if it is done via the diffusion process. However, it is hard to say much about the reliability of the numbers for $N > 10^6$ since there simply aren't anything to compare them to! Also, for large N, different numerical methods gives different answers. Given that these numbers are reliable though, the genetic interpretation of them is the same as that of tables 4 and 5, since the same

patterns that we saw in those tables only seem to continue with larger N.

For $N \leqslant 10^6$ we see that the numbers in tables 12 and 13 are close to the ones in tables 4 and 5. This what we hoped for, but it is also remarkable, since one of the assumptions made in the diffusion approximation was that $s = \mathcal{O}(N^{-1})$, and that doesn't hold for most N we have considered! To see just how well the diffusion process approximates the Markov process, a table of the relative errors is provided:

N	$h = 0$	$h = \frac{1}{2}$	$h = 1$
10	0.26	1.06	1.88
10^2	0.31	1.45	2.88
10^3	0.14	1.32	2.34
10^4	0.07	0.92	1.05
10^5	0.09	0.79	0.90
10^6	0.12	4.10	2.63

Table 14: Relative error (percent) of the diffusion approximation when $s = 0.02$.

N	$h = 0$	$h = \frac{1}{2}$	$h = 1$
10	1.50	6.13	11.00
10^2	0.96	6.89	12.50
10^3	0.33	6.05	9.84
10^4	0.20	4.80	4.87
10^5	0.31	4.17	7.54

Table 15: Relative error (percent) of the diffusion approximation when $s = 0.1$.

We see that, with few exceptions, the errors are very small, so the diffusion process approximates the Markov process remarkably well. All the additional mathematical work has payed off!

To gather more numerical evidence of some of the genetic effects mentioned in the previous subsection, we can compute the mean time until 95 percent of the population is A_2 and compare it to the mean absorption time (when 100 percent of the population is A_2). We compute this by (3.32), with the modification that the outer integral now is from 0.05 to 1. The following two tables should be compared with tables 12 and 13.

N	$h = 0$	$h = \frac{1}{2}$	$h = 1$
10	187449	181289	175371
10^2	119777	86670	64246
10^3	40108	10626	5946
10^4	13649	1866	1997
10^5	6523	1112	1584
10^6	5393	365	1077
10^7	2315	36	805
10^8	-	-	445

Table 16: Mean time until 95 percent A_2, when $s = 0.02$.

N	$h = 0$	$h = \frac{1}{2}$	$h = 1$
10	149381	126445	107641
10^2	56121	20103	10733
10^3	17895	2162	1268
10^4	6046	405	421
10^5	2852	255	232
10^6	2172	21	180
10^7	-	-	86

Table 17: Mean time until 95 percent A_2, when $s = 0.1$.

When comparing tables 16 and 17 with tables 12 and 13, we see several interesting things. For small N, the numbers are almost the same in both cases, and this indicates that absorption happens rather quickly once the first mutation has occurred.

In the case $h = 1$, we note two things. For one thing, the numbers no longer increase after $N = 10^4$. This means that in a population of that size, there are still A_1 alleles in the population that are hard to get rid of, but they are very few, namely at most 5 percent of the population. The other thing is that the numbers have decreased dramatically for larger N in comparison to tables 12 and 13. This proves the point mentioned earlier, that although the A_1 alleles are very few, those few remaining A_1's are very hard to get rid of completely. In genetic terms, this explains why there are still many bad genes with low frequencies in the human population, giving rise to diseases and malformations.

In the case $h = 0$, the difference in mean time between 95 and 100 percent fixation is very small, which indicates that the last 5 percent A_1 alleles is no harder to get rid of than the previous 95. This makes sense, because with $h = 0$, selection can work against all the A_1 alleles, regardless if they are part of A_1A_1 or A_1A_2, and independent of the remaining frequency of A_1.

Again, the case $h = \frac{1}{2}$ is intermediate.

Lastly, we consider the case $s = 0$ (which makes the parameter h redundant). We compare the means computed via the Markov and diffusion methods to the exact value given by (3.33). The numbers in the Markov case for $N \geqslant 10^5$ were computed by Pettersson [5].

N	Markov Chain	Diffusion	Exact form
10	200017	200020	200020
10^2	200197	200200	200200
10^3	201986	201987	201987
10^4	218985	218812	218812
10^5	327960	328987	328987
10^6	497392	628660	628660
10^7	-	1045424	1045935
10^8	-	1492470	1498393
10^9	-	1904733	1957697
10^{10}	-	2143666	2418051

Table 18: Mean absorbtion time when $s = 0$.

These numbers impliy yet again that for smaller N, the absorption time follows a geometric distribution, $T_0 \sim Ge(u)$, because if that was the case, then $\mathbb{E}(T_0) = \frac{1}{u} = 200000$, which is close to the numbers in the table.

It is no surprise that we get the same result up to $N = 10^7$, regardless if we use the diffusion approximation or the exact form, since the exact formula was derived using the diffusion approximation. The advantage of the exact form is that it is much faster to compute, and can be computed for (practically) arbitrarily large N. We also see that the diffusion computation is starting to fall behind after $N = 10^7$. This might indicate a computational limitation of the diffusion approximation (3.32), or at least a limitation of the methods by which the author chose to compute it.

Also note that in the case $N = 10^6$, the number computed using the Markov chain method is quite far off. This shows that the dramatic simplifications and special methods required to make the computation possible at all may have a negative effect on the end result in some cases.

References

[1] Bolza, Oscar "Lectures on the Calculus of Variations", third edition, Chelsea Publishing Company, 1978

[2] Ewens, Warren J. "Mathematical Population Genetics", Springer Verlag Berlin-Heidelberg, 1979

[3] Hartl, Daniel L. "A Primer of Population Genetics", Third Edition, Sinauer Associates Inc., 2000

[4] Kemeny, John G. Snell, J. Laurie "Finite Markov Chains" D. Van Nostrand Company, 1965

[5] Pettersson, Pelle, Personal Communication, 2018

[6] Rudin, Walter "Principles of Mathematical Analysis", 3rd edition, McGraw Hill Inc. 1976

[7] Rudin, Walter "Real and Complex Analysis", McGraw Hill Inc., 1966

[8] Shiryaev, Albert N. "Probability", 2nd English Edition, Springer, 1996

[9] Wahlén, Erik, Lecture notes in Ordinary Differential Equations 2, 16th february 2018. Pdf available at http://www.ctr.maths.lu.se/media11/MATM27/2018vt2018/Feb16.pdf

[10] Watterson, G.A. "Some Theoretical Aspects of Diffusion Theory in Population Genetics", Annals of Mathematical Statistics, Volume 33, Number 3, p.939-957 (1962)