
# MAT-3921

MASTER'S THESIS IN INDUSTRIAL MATHEMATICS

# Two Models of Population Genetics

*Jørgen Skancke*

June 2008

Faculty of Science
Department of Mathematics and Statistics
University of Tromsø

**Abstract**

The production of proteins in a cell is a regulated process. This means that the cell will only produce a type of protein when that type is needed. A fundamental step where this regulation occurs is at gene transcription. It is observed that transcription is regulated differently for different genes, and the question is therefore asked: why has evolution come up with different modes of transcriptional regulation for different genes?

Mathematical models of biological evolution are important for two reasons: 1) aiding researchers in understanding how complex biological systems have emerged and 2) enabling modelers to predict future outcomes of evolution. In this work, models of evolution of natural populations are applied to better understand the mechanisms of gene regulation in *E. coli* by investigating two predictor arguments of gene regulatory mode, namely the demand rule and the rule of minimal error load.

Two models of population genetics are derived: the Wright-Fisher model and the Moran model. These discrete stochastic models are approximated to continuous stochastic models and to continuous deterministic mean field models. The continuous stochastic models are used to investigate the demand rule, while the continuous deterministic models are used to investigate the rule of minimal error load.

In the continuous limits it is found that both discrete Wright-Fisher and Moran models can be described by the same equations. Two special cases are investigated in the model derivations: variable population size for the Wright-Fisher model and non-zero selection coefficients for continuous approximation of the Moran model. The models show that the demand rule describes well the evolution for the most basic mode of gene regulation, and that the rule of minimal error load describes the evolution for a larger group of gene regulation modes.

It is concluded that one should use the rule of minimal error load to investigate advanced systems of gene regulation. The demand rule is correct only as a special case for the most basic mode of gene regulation.

# Acknowledgments

First and foremost I would like to thank my supervisor Tor Flå for many stimulating conversations and for guiding me through my final year at the University of Tromsø as well as giving me an idea of what research is about.

I would like to thank my parents for always making everything so easy for me. Especially in these last months your support has been invaluable.

I also want to thank my friends who have helped me relax and unwind during these stressful times. Thank you for putting up with me!

A last thank you goes to my teachers at the Department of Mathematics and Statistics who have taught me about mathematics, statistics and physics for making it so that it was never difficult to ask for help.

# Contents

# Part I
# Models of Population Genetics

## 1 Introduction

Population genetics is a field within theoretical biology that began its development in the first half of the 20th century. Using mathematics, it connected the ideas of Mendelian inheritance with Darwin's ideas of natural selection and thus contributed to what is called the modern evolutionary synthesis. In the following paragraphs a presentation of some of the underlying biology of population genetics will be given.

Darwin spoke of **evolution** as 'descent with modification'. In this work, evolution is the change of gene frequencies in a population over the course of generations. If there are $n$ individuals with a certain gene in a population of total size $N$, then that gene has the **frequency** $x = n/N$ in that population. The frequencies of genes in a population are always changing. As individuals reproduce they pass on their genes to their offspring; since not all individuals have the same number of offspring gene frequencies will **change** over time. The main task of population genetics is to study in detail what underlies these fluctuations.

To understand how gene frequencies change, one must understand the function of genes and how they are passed on. The role of a **gene** in an organism is to serve as a blueprint for the production of one or several proteins in the organism's cells. Genes are for the most part located on **chromosomes**, which are long helical DNA molecules made up of base pairs of the nucleotides G, T, C and A. An organism can have just one or several chromosomes. Humans, for example, have $2 \times 23$ chromosomes in each normal cell. Humans inherit one chromosome of the same kind from each of their parents - one from their mother and one from their father, and thus humans are diploid organisms. In the models examined in this work the organisms are assumed to be **haploid**, which means they have only one chromosome of each kind. It is further assumed that they reproduce asexually by **cell division**. An example of a haploid, asexually reproducing organism is the bacterium *E. coli*, which in this work will be used as a model organism.

The specific location of a gene on a chromosome is called a **locus**. Different variants of the gene at a locus are called **alleles**. Different alleles are identified by their different **characteristic effects** on their host organisms. For example, consider a locus on an animal's chromosome where the gene for fur color resides. Different alleles at the locus will give different fur color for the animal. On the molecular level, different alleles always have different sequences of DNA nucleotides. Yet one allele may be represented by different DNA sequences *as long as the sequences produce the same charac-*

*teristic effect*. Hence it is the characteristic, and not the specific molecular sequence, that identifies the allele.

Evolutionary biologists have identified several mechanisms that drive evolution. In the population genetics models considered in this work, three fundamental evolutionary mechanisms will be described mathematically: genetic drift, mutation and natural selection.

**Genetic drift.** In reproduction and hence in evolution there is an element of chance. The term that is used to describe chance in evolution is genetic drift. As an example, assume a population of 100 individuals where half of the population carries an allele for pink fur and the other half an allele for brown fur. They are haploid individuals, reproduce by self-insemination and have only one offspring each. For simplicity, assume that this species reproduces seasonally and that the individuals live only one season. It is then possible that, for some reason, in one season only 50 individuals reproduce, and by chance all of them are individuals with pink fur. Thus all the offspring that season will have pink fur, and the allele for brown fur in the population is lost, for no other reason than chance. This is the mechanism of genetic drift. This mechanism will come into the population genetics models from the fundamental assumption of random mating.

**Mutation.** DNA molecules in living organisms continuously suffer from mutations. A mutation could be one base pair that changes into another, it could be the deletion or insertion of a base pair, or it could be larger events such as the duplication or deletion of whole stretches of DNA, like a cut and paste operation on the chromosome. Mutations can be caused by external factors, such as radiation, or by internal factors, such as mistakes made by the cellular copying machinery. There is a certain probability that as mutations change an allele's base pairs, that allele will change into another allele. This is the mechanism of mutation in evolution. Coefficients of mutation rates will be used in the population genetics models to quantify this effect. Although the rates of mutation will in general fluctuate in time, for instance during environmental stress, mutation rates are assumed constant in this work.

**Natural selection.** The different characteristic effects of the different alleles invite to a ranking between them. Suppose an organism with pink fur is easier to detect by predators than an organism of the same species with brown fur. This will likely lead to more pink-furred individuals being killed before reaching reproductive age. Hence the allele that gives brown fur is more likely to be passed on to the following generations. Another way of saying this is that the allele for brown fur has higher fitness than the allele for pink fur. Since nature thus 'selects' the most fit organisms, this is called the mechanism of natural selection. There are many ways by which selection can affect an organism; biologists differentiate between sexual selection, survival selection and several others. In the end however, the selection value of an allele may be identified with the relative reproductive rate of an organism

carrying that allele compared to its fellow species members bearing different alleles in the same environment. From this definition, it is clear that the selection value - often called the selection coefficient - of an allele depends on the environment. In the population genetics models, a selection coefficient will be given to each allele to quantify the fitness difference between them. Since an allele can be beneficial to the organism in one environment but detrimental in another, the selection coefficients of the alleles will be time dependent. For instance, two alleles can affect an organism equally until a stressful situation arises. Again considering the species with pink and brown fur, the color of the fur of the organism may not matter unless there are predators around. The alleles for pink and brown fur can thus be assumed to be selectively neutral during those periods (perhaps while sleeping in the nest or at night) while the allele for pink fur may have a selective disadvantage during other periods of time (such as while foraging for food during the day).

Having described the biological background necessary to understand the parameters of the population genetics models, the objective of this work will now be stated. The aim of this work is twofold. The first aim is to derive mathematical models that describe the change of allele frequencies in a population over time, incorporating the above mentioned evolutionary mechanisms. These models are the Wright-Fisher and Moran models of population genetics. They will first be introduced as discrete time, stochastic models. Starting from these, continuous time stochastic models and continuous time mean field models will be derived. The second aim is to apply both kinds of continuous time models to shed light on a problem in biology. The problem is the identification of factors that drive the evolution of gene regulation. Gene regulation put simply is how a cell decides when to turn a gene ON and OFF. One explanation of these factors is the Savageu demand rule, another is the rule of minimal error load. The validity of the demand rule and the rule of minimal error load will be tested using the population genetics models.

## 2 The Wright-Fisher model

The Wright-Fisher model is named after Sewall Wright and Ronald Fisher, two of the pioneers of population genetics. The model will first be introduced as a discrete stochastic Markov process and then approximated to a continuous Fokker-Planck equation.

At the heart of the Wright-Fisher model is a population that reproduces with non-overlapping generations. This means that individuals from two different generations can not exist together. At each time step in the model all the individuals of the current generation reproduce randomly and then die. Their offspring constitute the next generation.

The traditional way of introducing this model is with a constant popu-

lation size, $N(t) = N$. In this work a Wright-Fisher model that includes a time depended population size will be introduced as well.

It is the assumed random mating of the parents that determines the distribution alleles in the offspring. It permits thinking of the individuals, or alleles[1], in the offspring generation as having been randomly sampled from the parent generation. This thinking lends itself naturally to the binomial distribution.

## 2.1 The binomial distribution

The basic example of a binomial distribution is the distribution of the result of a series of coin tosses. The coin has two sides, the heads(H) side and the tails(T) side. The probability that a toss gives H is given by $p$ and the probability that the toss gives T is given by $1 - p$. If the coin is flipped $N$ times, the number of Hs, a stochastic variable $X$, is then binomially distributed:

$$P(X = n) = \binom{N}{n} p^n (1-p)^{N-n}. \tag{2.1}$$

From real-world coin-flips one initially expects $p = 0.5$. The outcome is then said to be random. Returning to the example with fur color, associate H with the event that an individual with pink fur is born into the offspring generation. Then, flipping 50 successive Hs in 50 flips would correspond to only individuals with pink fur reproducing one season. The allele for brown fur is lost since no Ts were flipped. This is how genetic drift comes into the Wright-Fisher model.

One can imagine effects that could influence the value of $p$. If occasionally a coin that landed on H changed into a T, the value of $p$ would decrease. In terms of evolution, this would be a mutation during reproduction in the allele for pink fur that changed it into an allele for brown fur. And if there was some inherent quality of T that caused it to land face up more often than H, the value of $p$ would decrease further. This would correspond to the allele for brown fur having a higher fitness than the allele for pink fur.

## 2.2 The Wright-Fisher model for two alleles

Assume that two alleles, called A and B, can occupy a chromosomal locus in a population of size $N(t)$. In generation $t$ allele A has the frequency $x' = n'/N(t)$. The $'$-notation is used to identify the number and frequency of alleles in generation $t$, as opposed to $n$ or $x$ which denote the number and frequency of alleles in generation $t+1$. Later, when the model is clearly defined, the $'$-notation will be dropped for the sake of readability. The model

[1]Since the organisms are haploid, one can identify an organism with its alleles in a 1-1 fashion.

will be specified in terms of $x = x_A$, the frequency of allele A, but of course $x_B$ can always be found from the relation $x_B = 1 - x$.

The variable of interest in the continuous approximation of the Wright-Fisher model is the allele frequency $x$. However, the model will initially be defined in terms of the discrete allele number $n$. The variable $x$ will be introduced when the approximation of the model to continuous gene frequencies is made.

The probability that out of $N(t+1)$ offspring there will be $n$ with allele A is equal to the probability of selecting $n$ parents that will give birth to a child with allele A out of $N(t+1)$ reproduction events, or samples. This probability is given by the binomial distribution as

$$P(n; t+1|n'; t) = \binom{N(t+1)}{n} p_1^n (1-p_1)^{N(t+1)-n}, \tag{2.2}$$

where $p_1$ is the probability of sampling a parent that gives birth to a child with allele A. In general, $p_1$ will be a function of $x'$, the frequency of allele A in the parent generation, as well as mutation rates and selection coefficients. The parents are assumed to be sampled with replacement, which means that one or more parents may reproduce several times.

To see how genetic drift, mutations and selection are introduced to $p_1$, each mechanism will be introduced one at a time. First, assume that there are no fitness differences between the two alleles and that mutations do not occur during reproduction. Then $p_1$ is simply the probability of selecting an A individual in a population where $n'$ out of $N(t)$ are A:

$$p_1 = x'. \tag{2.3}$$

If this value is used in equation (2.2), that equation describes the allele evolution in the population under influence of genetic drift only.

Next, consider the occurrence of mutations. Genetic drift happens as before but mutations are now assumed to occur as one generation changes into the next. Allele A is assumed to mutate[2] into allele B with probability $\nu$ and allele B into allele A with probability $\mu$. Each allele can theoretically mutate into an arbitrary DNA molecule; so for consistency in the two allele model one can either assume 1) that no other allele than A or B can occupy the locus in question or 2) that allele A is one specific allele and allele B is the set of all other possible alleles. In this work the interpretation 2) will be used.

With mutation in the picture there are now two pathways by which a child with allele A can be born from the parent generation. Either the parent carried allele A and no mutation occurred or the parent carried allele

[2]Mutation rates in bacteria are typically $\mathcal{O}(10^{-8})$ per base pair per cell division. The values of $\mu$ and $\nu$ depend on how many base pairs the alleles consist of (how long the gene is) and how many base pair changes that are capable of changing one allele into another.

B which mutated into allele A during reproduction. This gives the following modified expression for $p_1$:

$$p_1 = (1-\nu)x' + \mu(1-x'). \tag{2.4}$$

Third, assume that one of the alleles increases its bearer's chance to reproduce compared to bearers of the other allele. Thus one allele has a selection advantage over the other, or alternatively one allele has a selection disadvantage compared to the other. A selection disadvantage for one allele is assumed to reduce the probability for that allele to be sampled for reproduction. To quantify this effect, allele A is given a selection coefficient of 1 and allele B a selection coefficient of $(1-s)$. Thus if $s$ is positive allele A has a selection advantage and if $s$ is negative allele A has a selection disadvantage. In general, $s$ will be taken to be positive, so that allele B will be considered the less fit of the two. With selection in the picture, the probabilities that allele A and allele B are sampled for reproduction are respectively given by

$$\phi_1(x',t) = \frac{p_1}{z} \tag{2.5}$$

$$\phi_2(x',t) = (1-s(t))\frac{p_2}{z}. \tag{2.6}$$

$z$ is the partition sum given by

$$z = p_1 + (1-s(t))p_2, \tag{2.7}$$

to ensure that the sum of the two probabilities is equal to one, and $p_1$ is given by (2.4). Writing the probability that a child is born with allele A as $\phi_1(x',t)$ will from now on be beneficial, since aspects of this probability as a mathematical function will be considered. Since focus is on allele A, the notation $\phi(x',t) = \phi_1(x',t)$ will be used. The full expression for $\phi(x',t)$ is given by

$$\phi(x',t) = \frac{\mu + (1-\nu-\mu)x'}{1 - s(t)(1-\mu) + s(t)(1-\nu-\mu)x'}. \tag{2.8}$$

When the approximation of the Wright-Fisher model to continuous $x$ is made, this probability must be given an approximate value. The approximation is obtained by assuming that the model parameters $\mu$, $\nu$ and $s$ are $\mathcal{O}(\epsilon)$. It is assumed throughout this work that $\epsilon$ is a number less or equal to $10^{-2}$, and it is also assumed that $N(t)$ is large enough so that $1/N(t) = \mathcal{O}(\epsilon)$ for all $t$. As shown in the appendix, the approximation is given by

$$\phi(x',t) = x' + \epsilon M(x',t) + \mathcal{O}(\epsilon^2), \tag{2.9}$$

where

$$M(x',t) = \mu + s(t)(1-x')x' - (\nu+\mu)x'.$$

For the most part of this work $s(t)$ will be written as $s$ for simplicity of notation.

With $\phi(x,t)$ replacing $p_1$, equation (2.2) becomes

$$P(n|n') = \binom{N}{n} \phi_1^n (1-\phi_1)^{N-n}. \tag{2.10}$$

Here $N = N(t+1)$ and $P(n;t+1|n';t) = P(n|n')$ are used to simplify notation. Equation (2.10) gives the transition probabilities between the different possible numbers of individuals with allele A during the reproduction event. Thinking of the numbers $(0,1,\ldots,N)$ as the states the population can occupy, the reproduction event can be seen as a process that changes the state of the population. This process is Markov, since it is the allele frequencies, and the model parameters, of the parent population (the state at time $t$) that determines the allele frequency of the offspring population (the state at time $t+1$).

One can calculate the exact probability that the population has $n$ allele A at time $t+1$ from the marginal distribution

$$P(n,t+1) = \sum_{n'=0}^{N(t)} P(n|n')P(n',t), \tag{2.11}$$

using (2.10) to find the values of $P(n|n')$. The interpretation of $P(n,t+1)$ is that for a large ensemble of populations, a fraction $P(n,t+1)$ will have $n$ individuals with allele A in generation $t+1$.

In terms of computation, the probabilities for all values of $n$ at time $t+1$ can be found from the matrix equation

$$\underline{P}(t+1) = \boldsymbol{P}\underline{P}(t). \tag{2.12}$$

Thus, given an initial distribution of alleles $\underline{\mathrm{P}}(0)$, one can find the probability distribution for $n$ in the population for all time. This determines the evolution of the system and thereby defines the discrete Wright-Fisher model.

From now on the variable of interest will change from the allele number $n$ to the allele frequency $x = n/N$. The reason is that this change of variable will mediate the diffusion approximation of the Markov model, as will be presented in the following section. The coordinate change does not change the probability values, since $P(n,t) = \hat{P}(x,t)$, where $x = n/N$.

In the two following sections the derivation of the diffusion approximation will be sketched, first for $N(t)$ constant and then for $N(t)$ variable. The complete investigations are given in the appendix.

### 2.2.1 The diffusion approximation for $N(t)$ constant

In this section a continuous Fokker-Planck differential equation will be derived from the discrete Markov model. This derivation yields the classic

diffusion equation for population genetics where population size $N(t) = N$ is constant.

The idea behind the diffusion approximation is to assume that the population number is large enough so that the marginal distribution (2.11) may be substituted for its continuous counterpart. Rewrite

$$P(n, t+1) = \sum_{n'=0}^{N} P(n|n')P(n', t) \tag{2.13}$$

as

$$P\left(Nx, t+1\right) = \sum_{n'=0}^{N} P\left(Nx|Nx'\right) P\left(Nx', t\right). \tag{2.14}$$

This equation can in turn be written as

$$\bar{P}(x_n, t+1) = \sum_{n'=0}^{N} \bar{P}(x_n|x_{n'})\bar{P}(x_{n'}, t). \tag{2.15}$$

Now scale the equation by multiplication with $N$ and $N/\delta x$ so that

$$N\bar{P}(x_n, t+1) = \sum_{n'=0}^{N} N\bar{P}(x_n|x_{n'})N\bar{P}(x_{n'}, t)\delta x, \tag{2.16}$$

where $\delta x = 1/N$. Rewrite this as

$$\hat{f}(x_n, t+1) = \sum_{n'=0}^{N} \hat{f}(x_n|x'_n)\hat{f}(x'_n, t)\delta x. \tag{2.17}$$

Letting $N \to \infty$ this can be approximated by

$$f(x, t+1) = \int_S f(x|x')f(x', t)\mathrm{d}x', \tag{2.18}$$

where $S \in [0, 1]$. The allele number $n$ is now considered to be a continuous variable and $f(x, t)$ is taken to be a continuous probability distribution with the same moments as $P(n, t)/N$. Let $Q(x)$ be a continuous function with compact support on $S = [0, 1]$ with the following boundary values

$$Q^{(k)}(0) = Q^{(k)}(1) = 0 \tag{2.19}$$

for all non-negative integer values of $k$. The set of such functions make up the linear test space $\mathcal{D}$. By multiplying equation (2.18) with $Q(x)$ and integrating over $x$, the following identification is made for the left hand side of the ensuing equation:

$$(f^{t+1}, Q) = \int_S f(x, t+1)Q(x)\mathrm{d}x. \tag{2.20}$$

By the linearity of the integral, this is a linear functional, and by assumption it is also continuous, and thus $f$ can be considered as a distribution. The motivation for interpreting $f$ in this way comes from a paper by McKane and Waxman [1] which shows that $f$ will in general have singularities at the boundaries. This interpretation also comes naturally since $f$ is a continuous probability distribution. Inserting (2.11) on the right hand side of (2.20) results in

$$(f^{t+1}, Q) = \iint_S f(x|x')f(x',t)Q(x)\mathrm{d}x'\mathrm{d}x. \tag{2.21}$$

On the right hand side of this equation an expansion of $Q(x)$ about $x'$ given by

$$Q(x) = Q(x') + \frac{\partial Q(x')}{\partial x}\Delta x + \frac{1}{2}\frac{\partial^2 Q(x')}{\partial x^2}(\Delta x)^2 + \mathcal{O}((\Delta x)^3),$$

is inserted, where $\Delta x = x - x'$. By rearranging terms it is shown in the appendix that this gives

$$\begin{aligned} = \int_S f(x',t)Q(x')\mathrm{d}x' + \int_S f(x',t)\mathrm{E}[\Delta x|x']\frac{\partial Q(x')}{\partial x}\mathrm{d}x' \\ + \frac{1}{2}\int_S f(x',t)\mathrm{E}[(\Delta x)^2|x']\frac{\partial^2 Q(x')}{\partial x^2}\mathrm{d}x' + \mathcal{O}(\epsilon^2). \end{aligned} \tag{2.22}$$

The term $\mathcal{O}(\epsilon^2)$ comes from higher order moments of $\Delta x|x'$ as seen from equation (2.23) below. The moments of $\Delta x$ are found using (2.10):

$$\begin{aligned} \mathrm{E}[\Delta x|x'] &= \epsilon M(x',t) + \mathcal{O}(\epsilon^2) \\ \mathrm{E}[(\Delta x)^2|x'] &= \mathrm{Var}(\Delta x|x') + \mathrm{E}[\Delta x|x']^2 \\ &= \mathrm{Var}(\Delta x|x') + \mathcal{O}(\epsilon^2) \\ \mathrm{E}[(\Delta x)^3|x'] &= \mathcal{O}(\epsilon^2) \end{aligned} \tag{2.23}$$

$$\mathrm{Var}(\Delta x|x') = \epsilon\frac{D(x')}{N} + \mathcal{O}(\epsilon^2), \tag{2.24}$$

where $D(x') = x'(1-x')$. The approximate values of the moments come from using the approximation for $\phi(x',t)$ given in equation (2.9) when deriving them. Inserting the the moments $\mathrm{E}[\Delta x|x']$ and $\mathrm{E}[(\Delta x)^2|x']$ in (2.22), and recognizing that the differentiated $Q$ functions represent differentiation of $f$ in the distributional sense, one finds

$$(f^{t+1} - f^t, Q) = -\epsilon(\frac{\partial}{\partial x}[f(x,t)M(x,t)], Q) + (\frac{\epsilon}{2N}\frac{\partial^2}{\partial x^2}[f(x,t)D(x)], Q) + \mathcal{O}(\epsilon^2). \tag{2.25}$$

Since there is no longer a need separate between $x$ and $x'$, the variable $x'$ has changed name to $x$ for convenience. From (2.25), the following difference

equation is satisfied in the distributional sense:

$$f(x,t+1)-f(x,t)=-\epsilon\frac{\partial}{\partial x}[f(x,t)M(x,t)] + \frac{\epsilon}{2N}\frac{\partial^2}{\partial x^2}[f(x,t)D(x)]+\mathcal{O}(\epsilon^2). \quad (2.26)$$

To go from difference to differential equation, it will be necessary to scale the time variable. The following change of coordinates on the time variable is introduced:

$$\tau=\frac{t}{N}. \quad (2.27)$$

Thus

$$f(x,t)=f(x,N\tau)=\hat{f}(x,\tau) \quad (2.28)$$

and

$$f(x,t+\Delta t)=f(x,N\tau+\Delta\tau)=\hat{f}(x,\tau+\Delta\tau), \quad (2.29)$$

where $\Delta t=1$ and $\Delta\tau=\frac{1}{N}$. Introducing the coordinate change and dividing (2.26) by $\Delta\tau$, one finds

$$\frac{\hat{f}(x,\tau+\Delta\tau)-\hat{f}(x,\tau)}{\Delta\tau}=-\epsilon\frac{\partial}{\partial x}[\hat{f}(x,\tau)\hat{M}(x,\tau)] + \frac{\epsilon}{2}\frac{\partial^2}{\partial x^2}[\hat{f}(x,\tau)D(x)]+\mathcal{O}(\epsilon^2), \quad (2.30)$$

where $\hat{M}(x,\tau)=NM(x,\tau)$ indicates that $N$ has been absorbed into the parameters of $M(x,\tau)$. By letting $N\to\infty$, the above difference equation becomes a differential equation,

$$\frac{\partial f(x,t)}{\partial t}=-\frac{\partial}{\partial x}[f(x,t)M(x,t)]+\frac{1}{2}\frac{\partial^2}{\partial x^2}[f(x,t)D(x)], \quad (2.31)$$

to $\mathcal{O}(\epsilon^2)$. Here abuse of notation has been used to rewrite $\hat{f}(x,\tau)$ and $\hat{M}(x,\tau)$ as $f(x,t)$ and $M(x,t)$.

Equation (2.31) is by physicists called a Fokker-Planck equation, while in mathematical literature it is more often referred to as the forward Kolmogorov equation [2]. When used in the context of evolutionary theory it is called the diffusion equation of population genetics. This equation describes the time evolution of the allele frequency probability distribution for two alleles to $\mathcal{O}(\epsilon^2)$, where $\epsilon$ is the order of the model parameters $\mu$, $\nu$, $s$ and $1/N$. In physics literature on the Fokker-Planck equation $M(x,t)$ is called the drift coefficient and $D(x)$ is called the diffusion coefficient; they represent the deterministic and the stochastic part of the equation respectively. Thus, intuitively, what contributes to genetic diffusion is the variance of the change of gene frequency, captured in $D(x)$, while the mutation rates and the selection coefficient in $M(x,t)$ give a deterministic direction to this drift.

The labeling of $M(x,t)$ as the drift coefficient is not convenient in the context of population genetics since it is when $M(x,t) = 0$ that one describes what in evolutionary biology is called pure genetic drift. If $M(x,t) = 0$ then (2.31) is more famously known as the heat equation. Mathematically, (2.31) is a linear second-order partial differential equation of parabolic type [3].

In (2.31) the diffusion coefficient $D(x)$ is given by $D(x) = x(1-x)$ and $M(x,t)$ is given by

$$M(x,t) = U + S(t)(1-x)x - (V+U)x,$$

where

$$U = N\mu, \quad V = N\nu, \quad S(t) = Ns(t), \tag{2.32}$$

The interpretation of $f(x,t)$ in the Fokker-Planck equation is that for large values of $N$

$$\int_a^b f(x,t)\mathrm{d}x \tag{2.33}$$

is a good approximation of the fraction of populations in a large ensemble of populations where the allele A frequency $x$ is between $a$ and $b$ at time $t$ [1].

It should be noted that the diffusion approximation has traditionally not been derived from (2.11) as was done here. For example in [2] and [4], the diffusion equation is derived by just assuming that a continuous Markovian stochastic process for the change of allele frequency exists, independently of - but surely motivated by - the discrete Wright-Fisher model.

### 2.2.2 The diffusion approximation for $N(t)$ variable

When the population size can change over generations, the conditional probability for $P(n, t+1)$ is given by

$$P(n,t+1) = \sum_{n'=0}^{N(t)} P(n|n')P(n',t). \tag{2.34}$$

From this equation one arrives at

$$N(t+1)\bar{P}(x_n,t+1) = \sum_{n'=0}^{N(t)} N(t+1)\bar{P}(x_n|x_{n'})N(t)\bar{P}(x_{n'},t)\delta x, \tag{2.35}$$

where $\delta x = 1/N(t)$. Rewrite this as

$$\hat{f}(x_n,t+1) = \sum_{n'=0}^{N(t)} \hat{f}(x_n|x_{n'})\hat{f}(x_{n'},t)\delta x. \tag{2.36}$$

Letting $N(t) \to \infty$ equation (2.36) be approximated by

$$f(x, t+1) = \int_S f(x|x')f(x', t)\mathrm{d}x', \tag{2.37}$$

where $S \in [0, 1]$. The allele A number $n$ is considered to be a continuous variable and $f(x, t)$ is taken to be a continuous probability distribution with the same moments as $P(n, t)/N$. Define the space $\mathcal{D}$ of functions $Q(x) \in C^\infty[0, 1]$ with compact support on $S$ and boundary values

$$Q^{(k)}(0) = Q^{(k)}(1) = 0 \tag{2.38}$$

for all $k \in Z$. $\mathcal{D}$ is as such a linear test space. Multiplying both sides of (2.37) by $Q(x)$ and integrating over $x$ one obtains

$$\int_S f(x, t+1)Q(x)\mathrm{d}x = \iint_S f(x|x')f(x', t)Q(x)\mathrm{d}x'\mathrm{d}x. \tag{2.39}$$

The left hand side of this equation is now associated with the distribution $f^{t+1}$:

$$(f^{t+1}, Q) = \int_S f(x, t+1)Q(x)\mathrm{d}x \tag{2.40}$$

A Taylor expansion of $Q(x)$ about the value $x'$ gives

$$Q(x) = Q(x') + \frac{\partial Q(x')}{\partial x}\Delta x + \frac{1}{2}\frac{\partial^2 Q(x')}{\partial x^2}(\Delta x)^2 + O((\Delta x)^3),$$

where $\Delta x = x - x'$. $Q(x')$ and its derivatives are in $\mathcal{D}$ since $x' \in S$. Proceeding similarly from here on as for the case with $N(t)$ constant, one arrives at the difference equation

$$f(x, t+1) - f(x, t) = H(x, t) + \mathcal{O}(\epsilon^2), \tag{2.41}$$

satisfied in the distributional sense, where

$$H(x, t) = -\epsilon\frac{\partial}{\partial x}[f(x, t)M(x, t)] + \frac{\epsilon}{2N(t+1)}\frac{\partial^2}{\partial x^2}[f(x, t)D(x)] \tag{2.42}$$

Define a new time variable $\tau$ as

$$\tau = \int_0^{t+1} \frac{\mathrm{d}s}{N(s)}. \tag{2.43}$$

The generation time step in the new variable is given by

$$\Delta\tau = \frac{1}{N(t+1)} \tag{2.44}$$

Applying the variable change from $t$ to $\tau$ and dividing by $\Delta\tau$, (2.41) becomes

$$\frac{\hat{f}(x,\tau+\Delta\tau)-\hat{f}(x,\tau)}{\Delta\tau}=N(t+1)\hat{H}(x,\tau)+\mathcal{O}(\epsilon^2), \tag{2.45}$$

Letting $N(t)\to\infty$ for all $t$ one finds

$$\frac{\partial f(x,t)}{\partial t}=-\frac{\partial}{\partial x}[f(x,t)M(x,t)]+\frac{1}{2}\frac{\partial^2}{\partial x^2}[f(x,t)D(x)] \tag{2.46}$$

to $\mathcal{O}(\epsilon^2)$, after applying abuse of notation. The difference between this equation and the one for $N(t)$ constant is that the model parameters in $M(x,t)$ are now multiplied by the time dependent population size $N(t+1)$.

### 2.2.3 The mean allele frequency equation

By solving the above derived Fokker-Planck equation one can find the time evolution of the probability distribution for the allele frequency $x$. From the probability distribution one gets a full description of the Wright-Fisher model. However, the mean of $x$, denoted by $\langle x\rangle$, is an interesting parameter to investigate on its own, and this can be done without having to solve the Fokker-Planck equation first. In this section a differential equation that describes the time evolution of the mean allele frequency will be derived for $N(t)$ constant.

Let $n(t)$ denote that allele A number at time $t$. The mean of this number, given by $\mathrm{E}[n(t)]$, will be found from the identity

$$\mathrm{E}[n(t+1)]=\mathrm{E}[\mathrm{E}[n(t+1)|n(t)]]. \tag{2.47}$$

$\mathrm{E}[n(t+1)]$ will from now on be written as $\langle n\rangle(t+1)$ and $\mathrm{E}[n(t+1)|n(t))]$ as $\langle n|n'\rangle$. The value of $\langle n|n'\rangle$ is given by the binomial distribution (2.10). Thus,

$$\langle n\rangle(t+1)=N\langle\phi(x',t)\rangle. \tag{2.48}$$

Dividing this equation by $N$ gives

$$\langle x\rangle(t+1)=\langle x\rangle(t)+\epsilon\langle M(x,t)\rangle+\mathcal{O}(\epsilon^2), \tag{2.49}$$

where abuse of notation has been used by removing the $'$-notation and the usual approximation for $\phi(x,t)$ has been used. Seeking the $\langle M(x,t)\rangle$ function on the right hand side of this equation to be a function of $\langle x\rangle$, an expansion of $M(x,t)$ about $\langle x\rangle$ is performed. The expansion is given by

$$M(x,t)=M(\langle x\rangle,t)+\frac{\partial M(\langle x\rangle,t)}{\partial x}(x-\langle x\rangle)+\frac{1}{2}\frac{\partial^2 M(\langle x\rangle,t)}{\partial x^2}(x-\langle x\rangle)^2, \tag{2.50}$$

since $\partial^3 M(x,t)/\partial x^3=0$. Differentiating and inserting for $M(x,t)$, one finds

$$M(x,t)=M(\langle x\rangle,t)+s(1-2\langle x\rangle-(\mu+\nu))(x-\langle x\rangle)-s(x-\langle x\rangle)^2. \tag{2.51}$$

Taking the mean on both sides of this equation, $\langle M(x,t)\rangle$ is found to be

$$\langle M(x,t)\rangle = M(\langle x\rangle, t) - s\mathrm{Var}(x). \tag{2.52}$$

Inserting the right hand side of this equation in equation (2.49), one finds

$$\langle x\rangle(t+1) = \langle x\rangle(t) + \epsilon M(\langle x\rangle, t) - \epsilon s\mathrm{Var}(x) + \mathcal{O}(\epsilon^2) \tag{2.53}$$

Subtracting $\langle x\rangle(t)$ from both sides of this equation gives

$$\langle x\rangle(t+1) - \langle x\rangle(t) = \epsilon M(\langle x\rangle, t) - \epsilon s\mathrm{Var}(x) + \mathcal{O}(\epsilon^2). \tag{2.54}$$

By applying the coordinate change

$$\tau = \frac{1}{N} \tag{2.55}$$

and dividing the equation by $\Delta\tau = 1/N$, one finds

$$\frac{\langle x\rangle(\tau+\Delta\tau) - \langle x\rangle(\tau)}{\Delta\tau} = \epsilon\hat{M}(\langle x\rangle, \tau) + \epsilon S\mathrm{Var}(x) + \mathcal{O}(\epsilon^2), \tag{2.56}$$

where where $S = Ns$. Letting $N \to \infty$ results in a differential equation for the mean allele frequency:

$$\frac{\mathrm{d}\langle x\rangle}{\mathrm{d}t} = M(\langle x\rangle, t) + S\mathrm{Var}(x), \tag{2.57}$$

where the familiar abuse of notation has been used. Similarly as for the Fokker-Planck equation, the parameters in $M(\langle x\rangle, t)$ have been scaled by multiplication with $N$. Equation (2.57) describes the mean allele A frequency in the population to $\mathcal{O}(\epsilon^2)$.

## 2.3 The Wright-Fisher model for $K$ alleles

By deriving the continuous approximation of the Wright-Fisher model for two alleles one finds a Fokker-Planck equation that can readily be solved on a computer. However, in nature it is not realistic that only two alleles can reside at a locus. Therefore the Wright-Fisher model for $K$ alleles is now introduced. The discussion in section 2.2 can be generalized by assuming that there are $K$ possible alleles at a locus instead of just two. To each of the $K$ alleles a selection coefficient is assigned, $s_i$ for allele $i$. During reproduction each allele can mutate into one of the others with a specified mutation probability. $U_{ij}$ is the mutation probability from allele $i$ to allele $j$.

The probability of change of allele number is given by the multinomial distribution:

$$P(\boldsymbol{n}|\boldsymbol{n}') = \binom{N(t)}{n_1 \dots n_{K-1}} \phi_1^{n_1} \cdots \phi_{K-1}^{n_{K-1}} (1 - \sum_{k=1}^{K-1} \phi_k)^{N(t) - \sum_k n_k}, \tag{2.58}$$

where $\boldsymbol{n} = (n_1, \ldots, n_{K-1})$ contains the number of alleles 1 through $K-1$ in the population. $n_K$ is left out since it can be found from the first $K-1$. It is shown in the appendix that $\phi_i(\boldsymbol{x}', t)$ is given by

$$\phi_i(\boldsymbol{x}', t) = x_i' + \epsilon \mathcal{M}(\boldsymbol{x}', t) + \mathcal{O}(\epsilon^2), \tag{2.59}$$

where

$$\mathcal{M}_i(\boldsymbol{x}', t) = A_i + x_i' \left(B + (1-X)s_K - s_i\right) + (1-X)U_{Ki} - x_i' U_{iK}, \tag{2.60}$$

where

$$X = \sum_{k=1}^{K-1} x_k', \quad B = \sum_{k=1}^{K-1} s_k x_k', \quad \text{and} \quad A_i = \sum_{k \neq i}^{K-1} (x_k' U_{ki} - x_i' U_{ik}). \tag{2.61}$$

The $K$-allele marginal probability distribution, from which the diffusion approximation will be reached, is given by

$$P(\boldsymbol{n}, t+1) = \sum_{\boldsymbol{n'}} P(\boldsymbol{n}|\boldsymbol{n}') P(\boldsymbol{n'}, t). \tag{2.62}$$

### 2.3.1 The diffusion approximation for $N(t)$ constant

Starting from equation (2.62) one can argue as was done for two alleles and $N(t)$ constant, and end up with

$$P(\boldsymbol{x}, t+1) = \int_\sigma P(\boldsymbol{x}|\boldsymbol{x}') P(\boldsymbol{x'}, t) \mathrm{d}\boldsymbol{x}'. \tag{2.63}$$

The interval $\sigma$ is defined as $[0,1]^{K-1}$. Let $Q(\boldsymbol{x})$ be a scalar valued test function with compact support on $\sigma$ for which

$$\partial^\alpha Q(\boldsymbol{0}) = \partial^\alpha Q(\boldsymbol{1}) = 0, \tag{2.64}$$

for all non-negative integer elements in $\alpha = (\alpha_1, \alpha_2, \ldots, \alpha_{K-1})$. Here $\partial^\alpha$ is a multi-index notation for the partial derivative. See the appendix for details on this notation. Multiplying both sides of equation (2.63) with the $K-1$-dimensional Taylor expansion of $Q(\boldsymbol{x})$, integrating over $\sigma$, and then rearranging terms, one finds

$$\begin{aligned}
\int_\sigma f(\boldsymbol{x}, t+1) Q(\boldsymbol{x}) \mathrm{d}\boldsymbol{x} = &\int_\sigma f(\boldsymbol{x}', t) \left[ \int_\sigma f(\boldsymbol{x}|\boldsymbol{x}') \mathrm{d}\boldsymbol{x} \right] Q(\boldsymbol{x}') \mathrm{d}\boldsymbol{x}' \\
&+ \sum_{i=1}^{K-1} \int_\sigma f(\boldsymbol{x}', t) \left[ \int_\sigma \Delta x_i f(\boldsymbol{x}'|\boldsymbol{x}') \mathrm{d}\boldsymbol{x} \right] \frac{\partial Q(\boldsymbol{x}')}{\partial x_i} \mathrm{d}\boldsymbol{x}' \\
&+ \frac{1}{2} \sum_{i,j=1}^{K-1} \int_\sigma f(\boldsymbol{x}', t) \left[ \int_\sigma \Delta x_i \Delta x_j f(\boldsymbol{x}|\boldsymbol{x}') \mathrm{d}\boldsymbol{x} \right] \frac{\partial^2 Q(\boldsymbol{x}')}{\partial x_i x_j} \mathrm{d}\boldsymbol{x}' + \mathcal{O}(\epsilon^2).
\end{aligned}$$

In the appendix, this is shown to be equal to

$$\int_\sigma Q(\boldsymbol{x})f(\boldsymbol{x},t+1)\mathrm{d}\boldsymbol{x} = \int_\sigma f(\boldsymbol{x}',t)Q(\boldsymbol{x}')\mathrm{d}\boldsymbol{x}' + \sum_{i=1}^{K-1}\int_\sigma f(\boldsymbol{x}',t)\left[\int_{\sigma_1}\Delta x_i' f(x_i|\boldsymbol{x}')\mathrm{d}x_i\right]\frac{\partial Q(\boldsymbol{x}')}{\partial x_i}\mathrm{d}\boldsymbol{x}' + \frac{1}{2}\sum_{i,j=1}^{K-1}\int_\sigma f(\boldsymbol{x}',t)\left[\iint_{\sigma_1\sigma_1}\Delta x_i\Delta x_j f(x_i,x_j|\boldsymbol{x}')\mathrm{d}x_i\mathrm{d}x_j\right]\frac{\partial^2 Q(\boldsymbol{x}')}{\partial x_i x_j}\mathrm{d}\boldsymbol{x}' + \mathcal{O}(\epsilon^2). \tag{2.65}$$

The terms within the squared brackets correspond to $\mathrm{E}[\Delta x_i|\boldsymbol{x}']$ and $\mathrm{E}[\Delta x_i\Delta x_j|\boldsymbol{x}']$ respectively, whose values are found in the appendix. Inserting these values gives

$$= \int_\sigma f(\boldsymbol{x}',t)Q(\boldsymbol{x}')\mathrm{d}\boldsymbol{x}' - \sum_{i=1}^{K-1}\int_\sigma f(\boldsymbol{x}',t)\epsilon\mathcal{M}_i\frac{\partial Q(\boldsymbol{x}')}{\partial x_i}\mathrm{d}\boldsymbol{x}' + \frac{1}{2N}\sum_{i,j=1}^{K-1}\int_\sigma f(\boldsymbol{x}',t)C(x_i',x_j')\frac{\partial^2 Q(\boldsymbol{x}')}{\partial x_i x_j}\mathrm{d}\boldsymbol{x}' + \mathcal{O}(\epsilon^2), \tag{2.66}$$

where

$$C(x_i',x_j') = \begin{cases} -x_i'x_j' & \text{if } i\neq j \\ x_i'(1-x_i') & \text{if } i=j. \end{cases}$$

Interpreting the integrals in equation (2.66) as distributions, and interpreting the differentiated $Q$ functions as differentiation in the distributional sense, one ends up with

$$(f^{t+1},Q) = (f^t,Q) - \sum_{i=1}^{K-1}(\epsilon\frac{\partial f^t\mathcal{M}_i}{\partial x_i},Q) + \sum_{i,j=1}^{K-1}\frac{\epsilon}{2N}(\frac{\partial^2 f^t C_{i,j}}{\partial x_i x_k},Q) + \mathcal{O}(\epsilon^2). \tag{2.67}$$

The variable $\boldsymbol{x}'$ has changed name to $\boldsymbol{x}$ for convenience. Thus, the following difference equation is satisfied in the distributional sense

$$f(\boldsymbol{x},t+1)-f(\boldsymbol{x},t) = -\epsilon\sum_{i=1}^{K-1}\frac{\partial}{\partial x_i}f(\boldsymbol{x},t)\mathcal{M}_i(\boldsymbol{x},t)+\frac{\epsilon}{2N}\sum_{i,j=1}^{K-1}\frac{\partial^2}{\partial x_i x_j}f(\boldsymbol{x},t)C(x_i,x_j)+\mathcal{O}(\epsilon^2). \tag{2.68}$$

The transformation of time given by

$$\tau = \frac{1}{N} \tag{2.69}$$

is introduced, the difference equation is divided by $\Delta\tau = 1/N$, and the limit $N \to \infty$ is taken to give

$$\frac{\partial}{\partial t} f(\boldsymbol{x}, t) = -\sum_{i=1}^{K-1} \frac{\partial}{\partial x_i}[f(\boldsymbol{x}, t)\mathcal{M}_i(\boldsymbol{x}, t)] + \frac{1}{2}\sum_{i,j=1}^{K-1} \frac{\partial^2}{\partial x_i x_j}[f(\boldsymbol{x}, t)C(x_i, x_j)] \tag{2.70}$$

to $\mathcal{O}(\epsilon^2)$ after allowing for abuse of notation. This is the diffusion equation for population genetics for $K$ alleles and constant population size. The term $C(x_i, x_j)$ comes from the covariance of the frequencies $x_i$ and $x_j$, and $\mathcal{M}_i$ has been scaled by multiplication with $N$.

### 2.3.2 The diffusion approximation for $N(t)$ variable

By combining the derivation procedures of the Fokker-Planck equations for $K$ alleles with constant population size and two alleles with variable population size, it is shown in the appendix that the Fokker-Planck equation for $K$ alleles and non-constant population size is given by

$$\frac{\partial f(\boldsymbol{x}, t)}{\partial t} = -\sum_{i=1}^{K-1} \frac{\partial}{\partial x_i}[f(\boldsymbol{x}, t)\mathcal{M}_i(\boldsymbol{x}, t)] + \frac{1}{2}\sum_{i,j=1}^{K-1} \frac{\partial^2}{\partial x_i x_j}[f(\boldsymbol{x}, t)C(x_i, x_j)] \tag{2.71}$$

to $\mathcal{O}(\epsilon^2)$. The only difference between this equation and the one obtained for $K$ alleles and constant population size is that the model parameters in $\mathcal{M}_i(\boldsymbol{x}, t)$ are scaled by multiplication with $N(t+1)$ instead of $N$.

### 2.3.3 The mean allele frequency equation

The mean allele frequency equation for $K$-alleles and $N(t)$ constant is derived in the same way as for 2 alleles. By performing the same initial steps as were taken for the case with 2 alleles, but for higher dimensions, one ends up with the $K$-allele form of equation (2.49) for the mean of $\boldsymbol{x}$:

$$\langle \boldsymbol{x} \rangle(t+1) = \langle \boldsymbol{x} \rangle(t) + \epsilon \langle \mathcal{M}(\boldsymbol{x}, t) \rangle + \mathcal{O}(\epsilon^2). \tag{2.72}$$

Isolating the $i$th element of this equation gives:

$$\langle x_i \rangle(t+1) = \langle x_i \rangle(t) + \epsilon \langle \mathcal{M}_i(\boldsymbol{x}, t) \rangle + \mathcal{O}(\epsilon^2). \tag{2.73}$$

The expansion of $\mathcal{M}_i(\boldsymbol{x}, t)$ about $\langle \boldsymbol{x} \rangle$ is given by

$$\mathcal{M}_i(\boldsymbol{x}, t) = \mathcal{M}_i(\langle \boldsymbol{x} \rangle, t) + D\mathcal{M}_i(\langle \boldsymbol{x} \rangle, t)^T(\boldsymbol{x} - \langle \boldsymbol{x} \rangle) + \frac{1}{2}(\boldsymbol{x} - \langle \boldsymbol{x} \rangle)^T D^2 \mathcal{M}_i(\langle \boldsymbol{x} \rangle, t)^T(\boldsymbol{x} - \langle \boldsymbol{x} \rangle), \tag{2.74}$$

where $D\mathcal{M}_i(\langle \boldsymbol{x} \rangle, t)$ is the gradient and $D^2\mathcal{M}_i(\langle \boldsymbol{x} \rangle, t)$ is the Hessian of $\mathcal{M}_i(\langle \boldsymbol{x} \rangle, t)$. Performing the differentiations and taking the mean on both sides of the ensuing equation gives

$$\langle \mathcal{M}_i(\boldsymbol{x}, t) \rangle = \mathcal{M}_i(\langle \boldsymbol{x} \rangle, t) + \frac{1}{2} \sum_{j=1}^{K-1} (s_j - s_K) \mathrm{Cov}(x_j, x_i), \tag{2.75}$$

Inserting this value into equation (2.72) and proceeding as for the derivation for 2 alleles, one finds

$$\frac{\mathrm{d}\langle x_i \rangle}{\mathrm{d}t} = \mathcal{M}_i(\langle \boldsymbol{x} \rangle, t) + \frac{1}{2} \sum_{j=1}^{K-1} (S_j - S_K) \mathrm{Cov}(x_j, x_i) \tag{2.76}$$

to $\mathcal{O}(\epsilon^2)$. These equations describe the mean allele frequencies for alleles 1 through $K-1$. In the applications it will be argued that the terms with covariance are small enough to be neglected.

# 3 The Moran model

The Moran model was introduced in the late 1950s by P Moran [5]. As the Wright-Fisher model, it is a model of the change of allele frequency in a randomly mating population. The difference between the two is that the Moran model has overlapping generations. At each time step in the Moran model, one individual is randomly chosen to die and another individual is chosen to be born with a probability equal to $\phi(x, t)$. The stochastic process of the Moran model is as such a birth-death process. Since one individual dies and one is born at the same time, the Moran model is at heart a model with constant population size $N$. The time step in the Moran model is defined as one generation [6], even though only one individual reproduces. Hence the definition of generation is different in the Moran and Wright-Fisher models. In the Moran model one generation is one reproduction event, and in the Wright-Fisher model one generation is $N$ simultaneous reproduction events.

The Moran model will first be given a discrete stochastic description. From the discrete description a continuous allele frequency approximation will then be sought. To obtain the continuous model description, the Moran model will be specified in terms of a Master equation which in turn will be approximated to a Fokker-Planck equation.

The methods for the derivations of the Master and Fokker-Planck equations for the Moran model are adopted from [6]. In that paper, Blythe and McKane develop the Fokker-Planck approximation for the Moran model without selection, that is for $s = 0$. In this work, their result is extended by deriving the Fokker-Planck equation with $s \neq 0$. An added complexity in the derivation for $s \neq 0$ is that the approximation of $\phi(x, t)$ to $\mathcal{O}(\epsilon^2)$ must be introduced.

## 3.1 The Moran model for two alleles

Assume a population of size $N$, where at time $t$ there are $n$ individuals with allele A and $N-n$ individuals with allele B. Whereas in the Wright-Fisher model the allele frequency could theoretically go from $1/N$ to 1 in one generation, in the Moran model the number of alleles can at each time step only decrease by one, stay the same, or increase by one. The number of A will increase by one if a B is chosen to die and an A is born; stay the same if an A (B) is chosen to die and an A (B) is born; increase by one if a B is chosen to die and an A is born. The probability that an A dies at each event is simply given by its frequency, $x$, and thus the probability that a B dies is given by $1-x$. The probability that an A is born is given by $\phi(x,t)$ and the probability that a B is born is given by $1-\phi(x,t)$. The equation for $\phi(x,t)$ is obtained using the same reasoning as was used to find $\phi(x,t)$ in the Wright-Fisher model in section 2.2. The three elements of the transition matrix of the Moran model are thus given given by:

$$P(n'+1|n') = (1-x')\phi(x',t) \tag{3.1}$$

$$P(n'|n') = x'\phi(x,t) + (1-x')(1-\phi(x',t)) \tag{3.2}$$

$$P(n'-1|n') = x'(1-\phi(x',t)). \tag{3.3}$$

Since in each generation there are only three possible steps of action, it is possible to express the Moran model by a Master equation in an easy-to-follow manner. The Moran model is a Markov model, so it is defined by the marginal distribution

$$P(n,t+1) = \sum_{n'} P(n|n')P(n',t), \tag{3.4}$$

where $P(n|n')$ is found from (3.1). Subtracting $P(n,t)$ on both sides of (3.4) gives

$$P(n,t+1) - P(n,t) = \sum_{n'} P(n|n')P(n',t) - P(n,t) \tag{3.5}$$

$$= \sum_{n'\neq n} P(n|n')P(n',t) - \sum_{n'\neq n} P(n'|n)P(n,t), \tag{3.6}$$

since $\sum_{n'} P(n'|n) = 1$. The terms with $n = n'$ cancel and are therefore omitted.

It is now desired to change to a continuous time variable. In order to do that, two further assumptions about the Moran process must be made. First assume that time is divided into discrete parts, $\Delta t$, such that at most one reproduction event occurs within $\Delta t$. Dividing the previous equation with $\Delta t$ one thus finds

$$\frac{P(n,t+\Delta t) - P(n,t)}{\Delta t} = \sum_{n'\neq n} \frac{P(n|n')}{\Delta t} P(n',t) - \sum_{n'\neq n} \frac{P(n'|n)}{\Delta t} P(n,t). \tag{3.7}$$

Then assume that reproduction events take place at a *unit* rate, such that *on average* one reproduction event occurs every generation. The rate at which reproduction is assumed to take place is taken to be $T(n|n')$. This rate is related to the original transition probability by

$$P(n|n') = T(n|n')\Delta t + \mathcal{O}((\Delta t)^2), \tag{3.8}$$

where the last term represents the probability that two or more reproduction events occur during $\Delta t$. Inserting (3.8) into (3.7) and taking the limit $\Delta t \to 0$ gives

$$\frac{\partial P(n,t)}{\partial t} = \sum_{n' \neq n} T(n|n')P(n',t) - \sum_{n' \neq n} T(n'|n)P(n,t), \tag{3.9}$$

which is the Master equation for the Moran model. For large values of $N$, it will be shown that the Master equation can be approximated by a Fokker-Planck equation. To get there, the approximation for $\phi(x,t)$, the probability that an individual with allele A is born, as given by

$$\phi(x,t) = x + \epsilon M(x,t) + \mathcal{O}(\epsilon^2) \tag{3.10}$$

must be used. This approximation ensures that the following Fokker-Planck equation is $\mathcal{O}(\epsilon^2)$. However, since in the following an expansion about the parameter $N^{-1}$ to $\mathcal{O}(N^{-3})$ will be used, the $\mathcal{O}(\epsilon^2)$ term will be replaced by $\mathcal{O}(N^{-3})$. This does not cause any problems since $1/N = \mathcal{O}(\epsilon)$. As well, the model parameters $\mu, \nu$, and $s$ in $M(x,t)$ have to be scaled as follows:

$$U = \frac{N\mu}{2}, \quad V = \frac{N\nu}{2}, \quad S = \frac{Ns}{2}. \tag{3.11}$$

Using these parameters (3.10) turns into

$$\phi(x,t) = x + \frac{\epsilon}{N}\hat{M}(x,t) + \mathcal{O}(\epsilon^2), \tag{3.12}$$

where $\hat{M}(x,t) = (N/2)M(x,t)$. The difference between the scalings in (3.11) and the ones for the Wright-Fisher model as found in (2.32), is the factor 2. It will ensure that the resulting Fokker-Planck equation is of identical form to the one of the Wright-Fisher model.

The strategy for obtaining the Fokker-Planck equation is to first approximate the transition rates $T(n|n')$ to $\mathcal{O}(N^{-3})$ and Taylor expand the probabilities $P(n',t)$ about $P(x,t)$. These approximations are then inserted into the Master equation and multiplied to $\mathcal{O}(N^{-3})$. Finally, allowing $N \to \infty$ results in a Fokker-Planck equation identical to the one of the Wright-Fisher model, only with different scaling of the parameters and time. The full details of the following derivation is given in the appendix.

Fully expressed, the Master equation (3.9) is given by:

$$\frac{\partial P(n,t)}{\partial t} = T(n|n-1)P(n-1,t)+T(n|n+1)P(n+1,t)-[T(n-1|n)+T(n+1|n)]\,P(n,t).$$

The four transition rates are given by

$$\begin{aligned}
T(n|n-1) &= (1-\frac{n-1}{N})\phi(\frac{n-1}{N},t)\\
T(n|n+1) &= (\frac{n+1}{N})(1-\phi(\frac{n+1}{N},t))\\
T(n-1|n) &= \frac{1}{N}(1-\phi(x,t))\\
T(n+1|n) &= (1-\frac{1}{N})\phi(x,t),
\end{aligned}$$

which are sorted to $\mathcal{O}(N^{-3})$ using the approximation for $\phi(x,t)$ as in (3.10). The allele number $n$ is now assumed to be a continuous variable. The probability functions $P(n,t)$ with $n$ continuous are thus assumed constructed from the functions with discrete $n$ by interpolation such that they retain the same properties of the original functions. The Taylor-expanded probability functions with continuous $n$ are given by

$$\begin{aligned}
P(n-1,t) &= f(x,t) - \frac{1}{N}\frac{\partial f(x,t)}{\partial x} + \frac{1}{2N^2}\frac{\partial^2 f(x,t)}{\partial x^2} + \mathcal{O}(N^{-3}),\\
P(n+1,t) &= f(x,t) + \frac{1}{N}\frac{\partial f(x,t)}{\partial x} + \frac{1}{2N^2}\frac{\partial^2 f(x,t)}{\partial x^2} + \mathcal{O}(N^{-3}),
\end{aligned}$$

where a change of variable from $n$ to $x = n/N$ has been introduced and $f(x,t)$ is the name for $P(x,t)$ after the change of variable. Inserting these approximations into the Master equation and sorting to $\mathcal{O}(N^{-3})$, one obtains

$$\begin{aligned}
\frac{\partial f(x,t)}{\partial t} = \frac{2}{N^2}\Bigg[&S(2x-1)+(U+V)-2)f(x,t)\\
&+(1-2x-\epsilon M(x,t))P_x(x,t)+\frac{1}{2}x(1-x)P_{xx}(x,t)\Bigg]+\mathcal{O}(N^{-3}).
\end{aligned}$$

This equation can be simplified to

$$\frac{\partial f(x,t)}{\partial t} = \frac{2}{N^2}\left[-\frac{\partial}{\partial x}[f(x,t)M(x,t)]+\frac{1}{2}\frac{\partial^2}{\partial x^2}[f(x,t)D(x)]\right]+\mathcal{O}(N^{-3}).$$

Making the variable change on $t$ as given by $\tau = 2t/N^2$ and letting $N \to \infty$ gives

$$\frac{\partial f(x,t)}{\partial t} = -\frac{\partial}{\partial x}[f(x,t)M(x,t)]+\frac{1}{2}\frac{\partial^2}{\partial x^2}[f(x,t)D(x)],$$

which is the same Fokker-Planck equation as for the Wright-Fisher model. As before, abuse of notation has been used on $M(x,t)$ and $f(x,t)$. $M(x,t)$ and $D(x)$ are defined as for the Wright-Fisher model, except that the model parameters in $M(x,t)$ are scaled by multiplication with $N/2$ in stead of $N$. This equation describes the time evolution of the allele probability distribution with overlapping generations to $\mathcal{O}(\epsilon^2)$. The scaling of $t$ shows that with the Fokker-Planck description of the models, one generation in the Wright-Fisher model is $N/2$ as long as in the Moran model.

## 3.2 The Moran model for $K$ alleles

The arguments in section 3.1 can be generalized to include $K$ alleles. By doing this, the Master equation for $K$ alleles is found to be

$$\frac{\partial P(\boldsymbol{n},t)}{\partial t} = \sum_{\boldsymbol{n}'\neq\boldsymbol{n}} T(\boldsymbol{n}|\boldsymbol{n}')P(\boldsymbol{n}',t) - \sum_{\boldsymbol{n}'\neq\boldsymbol{n}} T(\boldsymbol{n}'|\boldsymbol{n})P(\boldsymbol{n},t), \tag{3.13}$$

where as in the Wright-Fisher model $\boldsymbol{n}$ is the vector of allele numbers.

It is shown in the appendix that by performing the same steps as outlined in the previous section, one finds the same Fokker-Planck equation as was found for the K-allele Wright-Fisher model:

$$\frac{\partial f(\boldsymbol{x},t)}{\partial t} = -\sum_{i=1}^{K-1}\frac{\partial}{\partial x_i}\mathcal{M}_i(\boldsymbol{x},t)f(\boldsymbol{x},t) + \frac{1}{2}\sum_{i,j=1}^{K-1}\frac{\partial^2}{\partial x_i x_j}C(x_i x_j)f(\boldsymbol{x},t). \tag{3.14}$$

Again time is scaled as $\tau = 2t/N^2$ and the mutation rates and selection coefficients in $\mathcal{M}_i(\boldsymbol{x},t)$ have been multiplied by $N/2$.

## 3.3 The mean allele frequency equation

Just as for the Wright-Fisher model, equations for the mean allele frequency in the Moran model can be derived. Since the derivation of the mean allele frequency equations for the Moran model is similar to the one for the Wright-Fisher equations, the case with $K$ alleles is considered directly without first performing the 2-allele derivation.

Again the identity $\mathrm{E}[\boldsymbol{n}] = \mathrm{E}[\mathrm{E}[\boldsymbol{n}|\boldsymbol{n}']]$ will be used to find $\langle\boldsymbol{x}\rangle(t+1)$. The expression for $\mathrm{E}[\boldsymbol{n}|\boldsymbol{n}']$ is given by

$$\langle\boldsymbol{n}|\boldsymbol{n}'\rangle = \sum_{\boldsymbol{n}} \boldsymbol{n}P(\boldsymbol{n}|\boldsymbol{n}'). \tag{3.15}$$

Each $n_i \in \boldsymbol{n}$ in this sum can either increase by one, decrease by one or stay the same in one reproduction event. Using the notation of (E.1) in the

appendix to simplify the expression of $P(\boldsymbol{n}|\boldsymbol{n'})$, this sum is expressed as

$$\langle \boldsymbol{n}|\boldsymbol{n'}\rangle = \boldsymbol{n'}P(\boldsymbol{n'}|\boldsymbol{n'}) + \sum_{i=1}^{K-1} \begin{pmatrix} n'_1 \\ \vdots \\ n'_i + 1 \\ \vdots \\ n'_{K-1} \end{pmatrix} P(n'_i + 1|\boldsymbol{n'}) + \begin{pmatrix} n'_1 \\ \vdots \\ n'_i - 1 \\ \vdots \\ n'_{K-1} \end{pmatrix} P(n'_i - 1|\boldsymbol{n'}),$$

which can be written as

$$= \boldsymbol{n'}\Big(P(n'_1 + 1|\boldsymbol{n'}) + \cdots + P(n'_{K-1} - 1|\boldsymbol{n'}) + P(\boldsymbol{n'}|\boldsymbol{n'})\Big) + \begin{pmatrix} P(n'_1 + 1|\boldsymbol{n'}) - P(n'_1 - 1|\boldsymbol{n'}) \\ \vdots \\ P(n'_i + 1|\boldsymbol{n'}) - P(n'_i - 1|\boldsymbol{n'}) \\ \vdots \\ P(n'_{K-1} + 1|\boldsymbol{n'}) - P(n'_{K-1} - 1|\boldsymbol{n'}) \end{pmatrix}.$$

The sum of probabilities in the first line of this equation is equal to 1, since it represents all possible state-transitions during reproduction. By isolating element $i$ in the previous vector equation, one finds

$$\langle n_i|\boldsymbol{n'}\rangle = n'_i + P(n'_i + 1|\boldsymbol{n'}) - P(n'_i + 1|\boldsymbol{n'}). \tag{3.16}$$

Inserting for

$$P(n'_i{+}1|\boldsymbol{n'}) = (1{-}x_i)\phi_i(\boldsymbol{x}, t) \quad \text{and} \quad P(n'_i{+}1|\boldsymbol{n'}) = x_i(1{-}\phi_i(\boldsymbol{x}, t)), \tag{3.17}$$

one finds

$$\langle n_i|\boldsymbol{n'}\rangle = n'_i + \epsilon\mathcal{M}_i(\boldsymbol{x}, t) + \mathcal{O}(\epsilon^2). \tag{3.18}$$

Applying the identity $\mathrm{E}[\boldsymbol{n}] = \mathrm{E}[\mathrm{E}[\boldsymbol{n}|\boldsymbol{n'}]]$ to equation (3.18) gives

$$\langle n_i\rangle(t + 1) = \langle n_i\rangle(t) + \epsilon\langle\mathcal{M}_i(\boldsymbol{x}, t)\rangle + \mathcal{O}(\epsilon^2). \tag{3.19}$$

As before, one seeks the function $\mathcal{M}(\boldsymbol{x}, t)$ in this expression in terms of $\langle\boldsymbol{x}\rangle$. Expanding $\mathcal{M}_i(\boldsymbol{x}, t)$ about $\langle\boldsymbol{x}\rangle$ gives

$$\mathcal{M}_i(\boldsymbol{x}, t) = \mathcal{M}_i(\langle\boldsymbol{x}\rangle, t) + D\mathcal{M}_i(\langle\boldsymbol{x}\rangle, t)^T(\boldsymbol{x}{-}\langle\boldsymbol{x}\rangle) + \frac{1}{2}(\boldsymbol{x}{-}\langle\boldsymbol{x}\rangle)^T D^2\mathcal{M}_i(\langle\boldsymbol{x}\rangle, t)^T(\boldsymbol{x}{-}\langle\boldsymbol{x}\rangle), \tag{3.20}$$

where $D\mathcal{M}_i(\langle\boldsymbol{x}\rangle, t)$ is the gradient and $D^2\mathcal{M}_i(\langle\boldsymbol{x}\rangle, t)$ is the Hessian of $\mathcal{M}_i(\langle\boldsymbol{x}\rangle, t)$. By taking the mean on both sides of this equation, one finds

$$\langle\mathcal{M}_i(\boldsymbol{x}, t)\rangle = \mathcal{M}_i(\langle\boldsymbol{x}\rangle, t) + \frac{\epsilon}{2}\sum_{j=1}^{K-1}(s_j - s_K)\mathrm{Cov}(x_i, x_j). \tag{3.21}$$

Inserting this expression into (3.19) gives

$$\langle n_i\rangle(t+1)-\langle n_i\rangle(t) = \epsilon\mathcal{M}_i(\langle \boldsymbol{x}\rangle, t)+\frac{\epsilon}{2}\sum_{j=1}^{K-1}(s_j-s_K)\mathrm{Cov}(x_i,x_j)+\mathcal{O}(\epsilon^2). \quad (3.22)$$

Dividing this equation by $N$ gives

$$\langle x_i\rangle(t+1)-\langle x_i\rangle(t) = \frac{1}{N}\left[\epsilon\mathcal{M}_i(\langle \boldsymbol{x}\rangle, t) + \frac{\epsilon}{2}\sum_{j=1}^{K-1}(s_j - s_K)\mathrm{Cov}(x_i,x_j) + \mathcal{O}(\epsilon^2)\right]. \quad (3.23)$$

What can be seen from this equation is that in the time scale of the Moran model the change of allele frequency for each reproduction event is $\mathcal{O}(\frac{\epsilon}{N})$. In the Wright-Fisher model, terms of $\mathcal{O}(\epsilon/N)$ where neglected as $\mathcal{O}(\epsilon^2)$ and left out of the equation. Anticipating the scaling of $t$, this term is allowed to remain for now.

Scaling the time parameter as $\tau = \frac{2t}{N^2}$ and dividing (3.23) by $\Delta\tau = \frac{2}{N^2}$ gives

$$\frac{\langle x_i\rangle(\tau+\Delta\tau) - \langle x_i\rangle(\tau)}{\Delta\tau} = \epsilon\hat{\mathcal{M}}_i(\langle \boldsymbol{x}\rangle, \tau) + \frac{\epsilon}{2}\sum_{j\neq i}^{K-1}(S_j - S_K)\mathrm{Cov}(x_i,x_j) + \mathcal{O}(\epsilon^2). \quad (3.24)$$

Letting $N\to\infty$ gives

$$\frac{\partial\langle x_i\rangle}{\partial t} = \epsilon\mathcal{M}_i(\langle \boldsymbol{x}\rangle, t) + \frac{1}{2}\sum_{j\neq i}^{K-1}(S_j - S_K)\mathrm{Cov}(x_i,x_j), \quad (3.25)$$

to $\mathcal{O}(\epsilon^2)$ where the usual abuse of notation has been used. This is the mean allele frequency differential equation for the Moran model. The parameters in $M(x,t)$ are multiplied by $N/2$, as for the Fokker-Planck approximation of the Moran model.

# Part II
# Numerical Methods

In this part, numerical methods to solve the population genetics models of the previous part will be introduced and discussed.

## 4 The Fokker-Planck equation for two alleles

The Wright-Fisher and Moran models have in part 1 been approximated by the same Fokker-Planck equation (3.14) when the population size $N(t)$

is constant for the Wright-Fisher model. What tells them apart after the diffusion approximation is how the time step is interpreted and that the model parameters in the Moran model are scaled slightly differently. In this section the solution method of the Fokker-Planck equation for two alleles and $N(t)$ constant will be discussed. Throughout it will be the generations and parameters of the Wright-Fisher model that will be used.

## 4.1 Boundary conditions and fixation probabilities

The range of the variable $x$ in the Fokker-Planck equation is [0,1]. The boundaries of this range correspond to two interesting phenomena in allelic evolution; when $x$ takes the values of 0 and 1 allele A is respectively lost or fixed in the population. Traditionally, the Fokker-Planck equation for population genetics has been assumed to be valid only in the range (0,1), as the appearance of singularities at the boundaries made analytical investigations difficult. Citing Kimura: "This is a partial differential equation with singularities at the boundaries, so that no arbitrary conditions can be imposed there" [7]. The singularities appear where in the discrete Wright-Fisher and Moran models probabilities would sum to 1. The continuous approximation has turned these values into a Dirac delta functions. Previously, one has resorted to the backward Kolmogorov equation to investigate loss and fixation of alleles, see e.g. [8]. One could argue that this is not a satisfactory result for a mathematical theory of population genetics: a full description of the probability distribution for $x$ is given by the discrete models, so one should hope that this full description is recaptured in the diffusion approximations. Quite recently, McKane and Waxman proposed a way by which the full essence of the original discrete model is maintained in the diffusion approximation [1]. Their result is that the singularities observed at the boundaries should be associated with the probabilities for fixation and loss.

The fundamental principle in the paper by McKane and Waxman is that the probability that the allele frequency is in the range [0,1] should at all times be unity. In other words, probability should be conserved:

$$\int_0^1 f(x,t)\mathrm{d}x = 1 \tag{4.1}$$

for all values of $t$. The diffusion equation can be written as

$$\frac{\partial f(x,t)}{\partial t} = -\frac{\partial j(x,t)}{\partial x}, \tag{4.2}$$

where

$$j(x,t) = M(x,t)f(x,t) - \frac{1}{2}\frac{\partial}{\partial x}[D(x)f(x,t)]. \tag{4.3}$$

The function $j(x,t)$ is thus the probability density flux. Integrating (4.2) over $x \in [0,1]$ and applying (4.1) gives

$$j(0,t) = j(1,t). \tag{4.4}$$

This facilitates boundary conditions on the probability flux. By equation (4.1), no probability can be lost to the area outside [0,1]. It is therefore reasonable to impose the boundary conditions

$$j(0,t) = 0 \tag{4.5}$$

$$j(1,t) = 1. \tag{4.6}$$

Using these boundary conditions, it is shown in the paper by McKane and Waxman that the general form of the solution to the diffusion equation is

$$f(x,t) = \Pi_0(t)\delta(x) + \Pi_1(t)\delta(1-x) + \bar{f}(x,t). \tag{4.7}$$

In this equation $\Pi_0(t)$ and $\Pi_1(t)$ are the probabilities that $x$ has become 0 and 1 by time $t$ respectively (in other words, that allele A has become lost and fixed), $\delta(x)$ and $\delta(1-x)$ are Dirac delta functions, and $\bar{f}(x,t)$ is the solution of (4.2) for $x \in (0,1)$.

It is shown in the same paper that $\Pi_0(t)$ will appear in the solution only if there is no mutation from allele B to allele A and $\Pi_1(t)$ will appear only if there is no mutation from allele A to allele B. If there is no mutation either way both $\Pi_0(t)$ and $\Pi_1(t)$ will appear. This is reasonable, since an allele can strictly never be fixed in a population if there is a possibility that it mutates into a different allele. However, since mutations are always present in nature there can never be any true, mathematical fixation of alleles in a natural population. It is nevertheless interesting to investigate what might be called quasi-fixation of alleles in the diffusion approximation, since this is what is actually observed in nature. Quoting Kimura: "In other words, after a sufficient number of generations almost all populations will be in such a situation that the gene is either almost fixed in the population or almost lost from it. To distinguish this from the fixation or loss in the case of small effective population number, the terms 'quasi-fixation' and 'quasi-loss' are proposed." [7]. Although Kimura defined these terms for the quasi-loss and quasi-fixation he observed with a stochastic selection coefficient, they will be used in this work as interpretations of the accumulation of probability close to the boundaries in the solution of the Fokker-Planck equation.

The numerical solution method of the diffusion equation that will be introduced in the next section will facilitate this identification. The values $f(0,t)$ and $f(1,t)$ from the numerical solutions will be associated with the probabilities for respectively loss and fixation of allele A. In the absence of mutations, these values will be absolute fixation and loss probabilities, while in the presence of mutations these values will be called quasi-fixation and quasi-loss probabilities.

## 4.2 Numerical solution scheme

The numerical solution scheme will be developed from the equation

$$\frac{\partial f(x,t)}{\partial t} = -\frac{\partial j(x,t)}{\partial x}, \tag{4.8}$$

where

$$j(x,t) = f(x,t)M(x,t) - \frac{1}{2}\frac{\partial}{\partial x}[f(x,t)D(x)]. \tag{4.9}$$

The notation

$$f(x_k, t^n) \doteq f_k^n \tag{4.10}$$

is used to denote the numerical approximation, where $k$ and $n$ are indices that discretize values for $x$ and $t$. Central difference approximation is used for the right hand side of (4.8):

$$\frac{\partial}{\partial x}F(x_k, t^n) \doteq \frac{1}{\Delta x}(F^n_{k+\frac{1}{2}} - F^n_{k-\frac{1}{2}}), \tag{4.11}$$

to $\mathcal{O}((\Delta x)^2)$. Note that here $\Delta x$ is the step size of the discretization of $x$, and not the difference between the allele A frequency in generation $t$ and $t+1$ as in part I. Using the forward difference approximation for the left hand side of (4.8) gives

$$\frac{\partial}{\partial t}f(x_k, t^n) \doteq \frac{1}{\Delta t}(f_k^{n+1} - f_k^n), \tag{4.12}$$

to $\mathcal{O}(\Delta t)$. Combining these two expressions, the time advancement is given by

$$f_k^{n+1} = \frac{\Delta t}{\Delta x}(F^n_{k+\frac{1}{2}} - F^n_{k-\frac{1}{2}}) + f_k^n, \tag{4.13}$$

where the total error is given by a term $\mathcal{O}(\Delta t) + \mathcal{O}((\Delta x)^2)$. To identify the half-step terms on the right hand side of this equation, central difference is used on the right hand side of (4.9):

$$F_k^n = (fM)_k^n - \frac{1}{2\Delta x}((fD)^n_{k+\frac{1}{2}} - (fD)^n_{k-\frac{1}{2}}), \tag{4.14}$$

to $\mathcal{O}((\Delta x^2))$. Using this equation one can identify

$$\begin{aligned} F^n_{k+\frac{1}{2}} &= (fM)^n_{k+\frac{1}{2}} - \frac{1}{2\Delta x}((fD)^n_{k+1} - (fD)^n_k) \\ F^n_{k-\frac{1}{2}} &= (fM)^n_{k-\frac{1}{2}} - \frac{1}{2\Delta x}((fD)^n_k - (fD)^n_{k-1}). \end{aligned} \tag{4.15}$$

Using the identity

$$g(x + \frac{\Delta x,}{2}, t) = \frac{g(x+\Delta x, t) + g(x,t)}{2} + O((\Delta x^2)), \tag{4.16}$$

one finds

$$(fM)^n_{k+\frac{1}{2}} = \frac{1}{2}[(fM)^n_{k+1} - (fM)^n_k] + O((\Delta x^2))$$
$$(fM)^n_{k-\frac{1}{2}} = \frac{1}{2}[(fM)^n_k - (fM)^n_{k-1}] + O((\Delta x^2)). \tag{4.17}$$

Inserting (4.17) into (4.15) and then the ensuing result into (4.13), one obtains

$$f^{n+1}_k = \frac{\Delta t}{2\Delta x}\left(\left[(fM)^n_{k+1} - 2(fM)^n_k + (fM)^n_{k-1}\right] - \frac{1}{\Delta x}\left[(fD)^n_{k-1} - 2(fD)^n_k + (fD)^n_{k-1}\right]\right) - f^k_n \tag{4.18}$$

to $\mathcal{O}(\Delta t) + \mathcal{O}((\Delta x^2))$. This is the solution scheme for the Fokker-Planck equation (4.9). The boundary values (4.5) impose the conditions

$$f^{n+1}_0 = \frac{\Delta t}{\Delta x}(F^n_{\frac{1}{2}} - 0) - f^n_0$$
$$f^{n+1}_N = \frac{\Delta t}{\Delta x}(0 - F^n_{N-\frac{1}{2}}) - f^n_N, \tag{4.19}$$

where the index $n = N$ identifies $x = 1$. These conditions do not use that the flux is zero in F(0,t) and F(1,t), as were the boundary conditions. Rather it is assumed that the flux that zero in the half-steps outside the range of $x$: $F(-0.5, t) = F(1.05, t) = 0$. In spite of not using the exact boundary conditions, this configuration still solves the equation well.

Let $p$ be the initial allele A frequency. Since $f(x, t)$ is a distribution, the initial condition can be given in terms of a delta function:

$$f(x, 0) = \delta(x - p). \tag{4.20}$$

In the discrete, numerical version this is translated to

$$f^0_k = p. \tag{4.21}$$

The above solution scheme together with the boundary values and initial condition provides a numerical solution algorithm for the diffusion equation for population genetics.

## 4.3 Sample solutions

Some solutions of the Fokker-Planck equation are presented in this section to show how the different parameter regimes of $\mu$, $\nu$ and $s$ affect the probability distribution for the allele frequency. The numbers given for the parameters in the figures are multiplied by $N$; when the figure shows $\mu$, the number is $\mu N$. Calculations were run as long as it took for the distributions to attain equilibrium.

Figure 1 shows how the probability distribution changes for various mutation rates. Observe the change in shape as $\mu N$ and $\nu N$ pass below the value 0.5. For rates above this value the allele could not be lost or fixed. For rates below this value mutation is no longer strong enough to resist the random effects of diffusion that make it likely that the allele will either quasi-disappear from the population or become quasi-fixed. Since fixation probabilities will be central when investigating the applications of the Fokker-Planck equation, the regime will be then be $\mu N, \nu N < 0.5$.

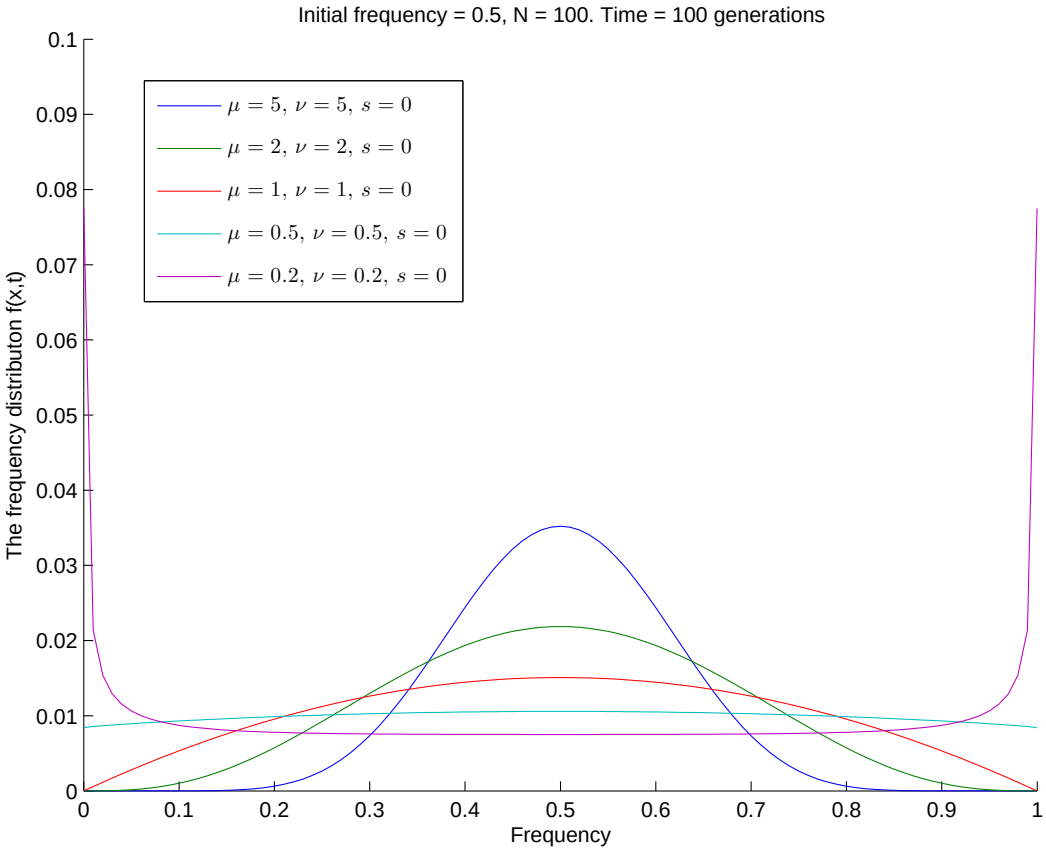


Figure 1: The solutions of the diffusion equation are given for equal mutation rates and zero selection to emphasis the change that occurs for decreasing mutation rates. At $\mu N,\ \nu N = 0.5$ the distribution undergoes a crucial change of shape.

Figure 2 shows the effect selection has on the shape of the probability distribution. Recall that the selection coefficient gives allele B a fitness disadvantage when positive. Its effect on the distribution is intuitively to shift the peak of the distribution in the direction of increased frequency of allele A.

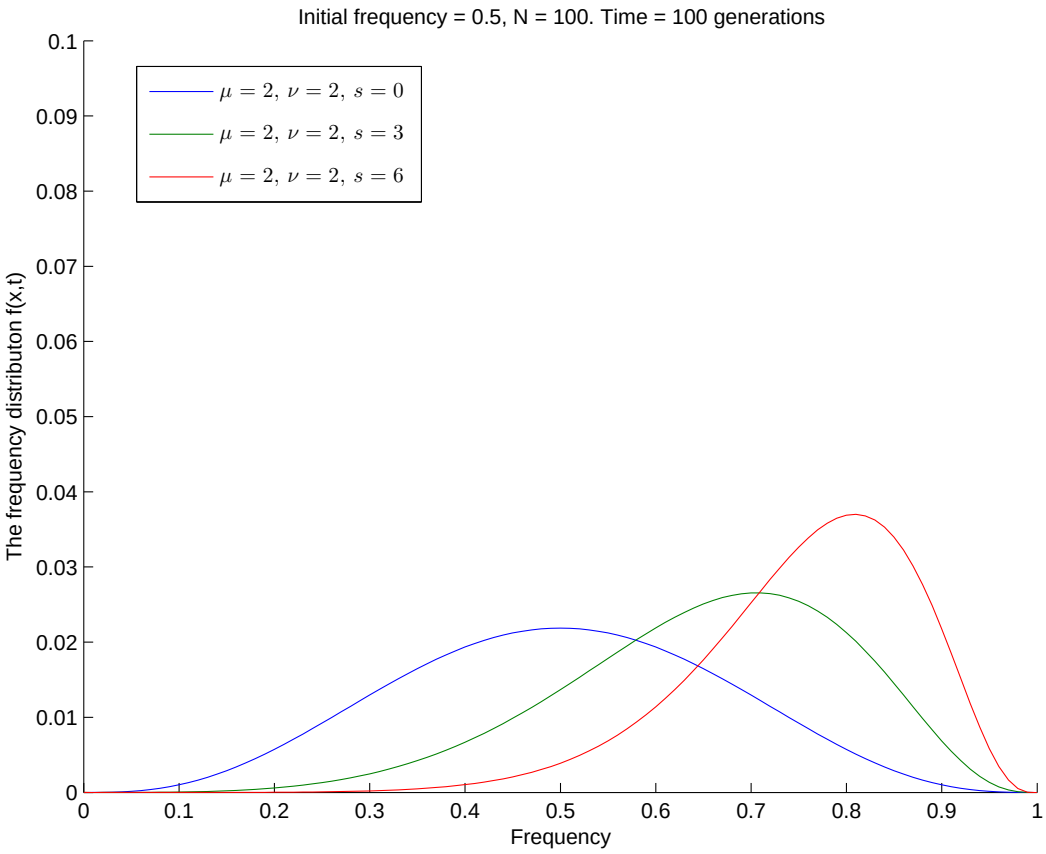


Figure 2: The solutions of the diffusion equation for increasing values of $s$. The larger the selection coefficient, the further the peak of the frequency distribution is shifted in the direction of increasing $x$.

Dirac delta functions will appear at the boundaries when either $\mu N$ or $\nu N$ are 0. These are, in a mathematical sense, true fixation probabilities [1]. Figure 3 shows solutions for decreasing rate $\nu N$ until it reaches zero. When this occurs, a delta function at $x = 1$ appears in a numerical sense. In other words, all the values close to the boundary are very close to zero, while the boundary will have some value that is interpreted as the true fixation probability.

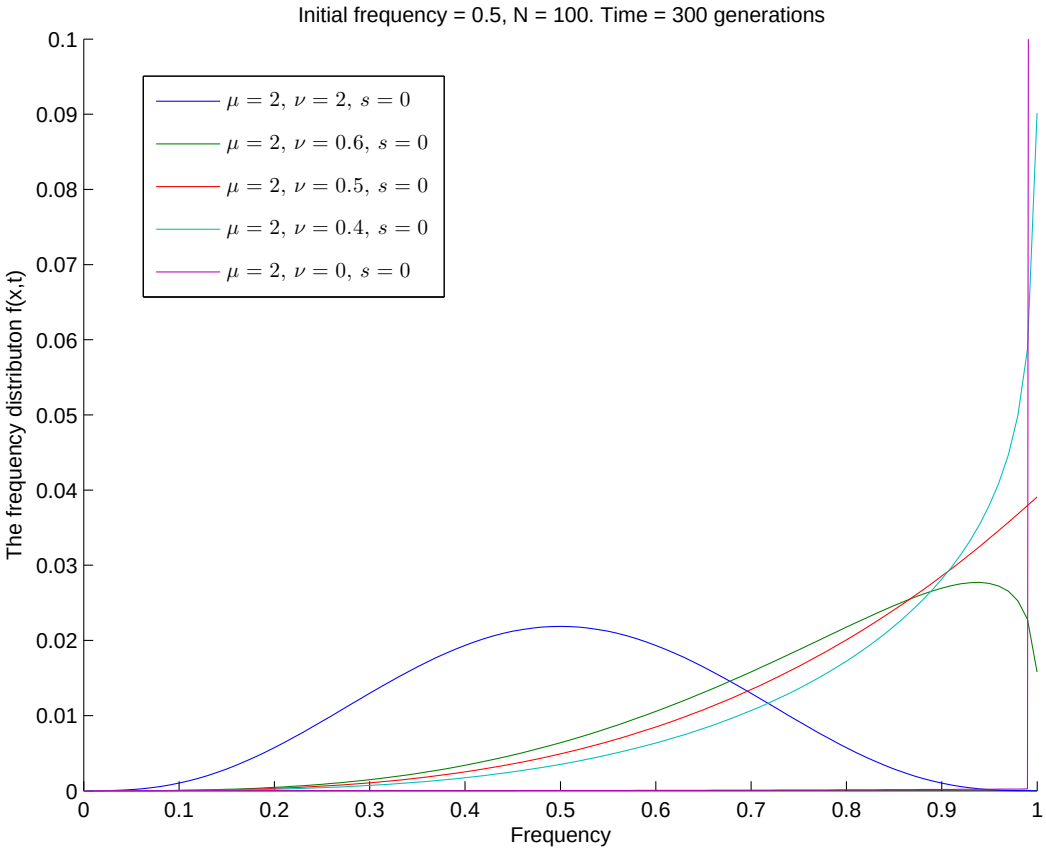


Figure 3: Decreasing the mutation rate $\nu N$ until it reaches zero. Observe the drastic change that occurs around $\nu N = 0.5$.

## 5 The mean field equation for four alleles

The mean field equations will be solved for $K = 4$ alleles, which results in three differential equations describing the allele frequency of alleles $x_1, x_2$

and $x_3$:

$$\frac{\partial\langle x_1\rangle}{\partial t} = M_1(\langle \boldsymbol{x}\rangle, t) + \frac{1}{2}\sum_{j=1}(S_j - S_4)\mathrm{Cov}(x_j, x_1) \tag{5.1}$$

$$\frac{\partial\langle x_2\rangle}{\partial t} = M_2(\langle \boldsymbol{x}\rangle, t) + \frac{1}{2}\sum_{j=1}(S_j - S_4)\mathrm{Cov}(x_j, x_2) \tag{5.2}$$

$$\frac{\partial\langle x_3\rangle}{\partial t} = M_3(\langle \boldsymbol{x}\rangle, t) + \frac{1}{2}\sum_{j=1}(S_j - S_4)\mathrm{Cov}(x_j, x_3). \tag{5.3}$$

$$\tag{5.4}$$

These equations will be solved under the parameter regime $1/s << N << 1/\mu$, where $s$ and $\mu$ represent all selection coefficients and mutation rates. In terms of $S$ and $U$ this is the same as $S >> 1$ and $U << 1$. Under these conditions Rouzine et al. show in [9] that for $K = 2$ alleles the steady state variance of $x$ is given by

$$\mathrm{Var}(x) = \frac{U}{2S^2}, \tag{5.5}$$

when $1/s\ln(s/\mu) << N << 1/\mu$, where in their work $\mu = \nu$. For these values of $N$ the variance is small enough that it can be omitted from the equations above. This coincides with the deterministic limits of the diffusion equation, obtained when $N \to \infty$. In this limit the two allele mean frequency equation is given by

$$\frac{\partial\langle x\rangle}{\partial t} = M(\langle x\rangle, t), \tag{5.6}$$

as shown in [9]. Since Rouzine et al. only consider the case of two alleles, there is no approximate value for the covariance between allele frequencies. However, it is assumed here that for $N$ large enough both, the covariance terms in (5.1) can be neglected. Thus the system that will be solved is written in vector notation as

$$\frac{\partial\langle \boldsymbol{x}\rangle}{\partial t} = \boldsymbol{M}(\langle \boldsymbol{x}\rangle, t), \tag{5.7}$$

where $\boldsymbol{x} = (x_1, x_2, x_3)^{\mathrm{T}}$ and $\boldsymbol{M}(\langle \boldsymbol{x}\rangle, t) = (M_1(\langle \boldsymbol{x}\rangle, t), M_2(\langle \boldsymbol{x}\rangle, t), M_3(\langle \boldsymbol{x}\rangle, t))^{\mathrm{T}}$. A standard Runge-Kutta method of fourth order will be used to solve this equation numerically.

The strength of the mean field equations is that systems with more than two alleles can be investigated numerically. This allows for a broader scope of applications. Their weakness is that the parameter regime as discussed above must be upheld. As such, notions of fixation probability will not be applied to the mean field equation; however, one still hopes to obtain some relevant information from their solutions.

## 5.1 Sample solutions

The four allele mean field equations will be now be solved for a simple example, that shows how one can use the mean field equations to predict an evolutionary process.

Assume that the model organisms lives in an environment that fluctuates between two conditions, called condition 1 and condition 2. It is assumed that the organism is a bacteria, but no assumptions about the characteristics of the alleles are specified. The different alleles are however assumed to give different benefits to their host organisms in the different environmental conditions. Table 1 lists the selection coefficients of the alleles in the two conditions. Note that these selection coefficients are given as $S = sN$. From this table one can for example read that allele 1 is detrimental to the organism in condition 1 but beneficial in condition 2. Assume further that

Table 1: Selection coefficients

| Allele | Condition 1 | Condition 2 |
|---|---|---|
| 1 | 4 | -2 |
| 2 | 0 | 2 |
| 3 | 3 | -1 |
| 4 | -3 | 5 |

there exist two distinct populations, separated in space with no interaction between them. Call them population $\alpha$ and population $\beta$. Not only are they separated in space, but they are also separated in how often they are subjected to the two different conditions of the environment. Assume that species $\alpha$ is in condition 1 10% of the time and in condition 2 90% of the time. Assume that species $\beta$ is in both conditions equally often. This could for example correspond to two different population of the same species living in two different climates.

The number of generations within which the conditions vary is set to 80. Different bacteria can have wildly different cell division times, so one can in general not define a typical time frame for a bacterial generation. Assuming that this species of bacteria is replicating itself fairly rapidly in both environments, 80 generations in the Wright-Fisher sense can correspond to a time frame of between two weeks and two months. However, the results of the simulations for the mean field equations are stable with respect to variations in this number. It is for the most time the time spent in either condition that determines the outcome, as will be seen for a specific example in part III.

The mutation rates for the simulations are $U_{ij} = 0.01$ for $i, j = 1, \ldots, 4$, the initial frequencies of the alleles are 0.25.

The populations were sampled for 700 generations. Figure 4 shows the fate of population $\alpha$. This population has a polarized allele distribution, with allele 1 dominating other three. Figure 5 shows the fate of population $\beta$. In this population there is no dominant allele, and the population contains a mix of the different alleles.

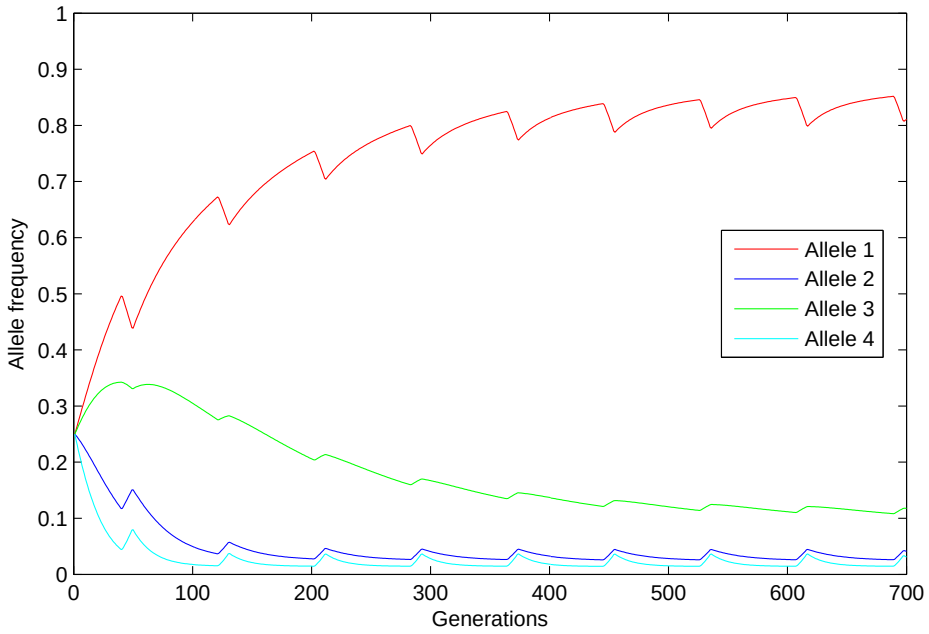


Figure 4: The mean frequencies of the four alleles for population $\alpha$. Allele 1 stands out as the dominant allele.

This example can be used to explain speciation, which is the process by which new species arise. Assume that over time, allele 1 becomes fixed in population $\alpha$. Assuming that different alleles at other loci become fixed in the two populations as well as a result of spending time in two different environments, the DNA of the two populations might over time become different enough that it makes sense to classify them as two different species.

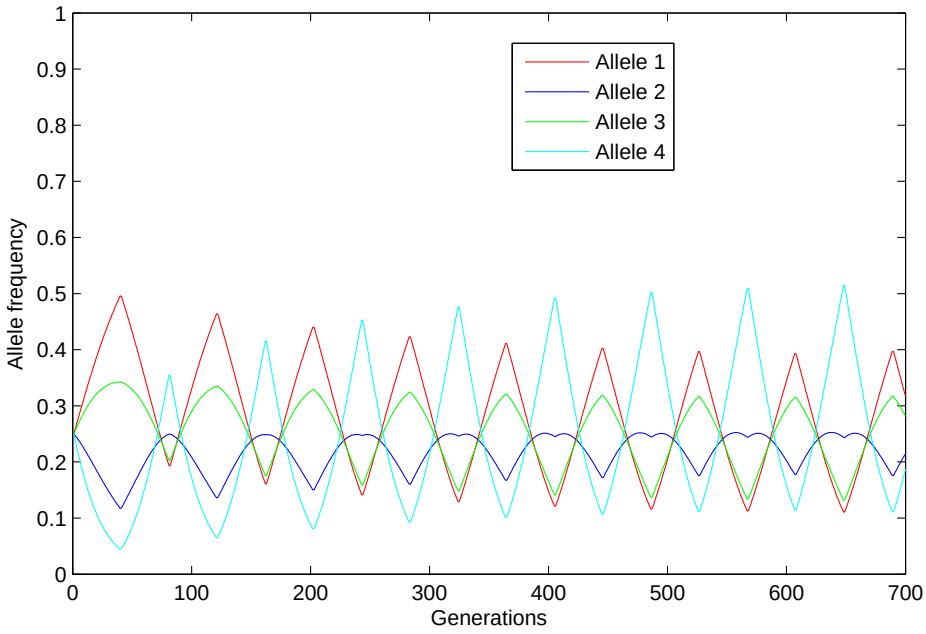


Figure 5: The mean frequencies of the four alleles for population $\beta$. There is no dominant allele, leading to a mixed population.

# Part III
# Model Applications

This part will begin with a presentation of the areas and mechanisms within molecular biology that motivate the solutions of the population genetics models in this work. This will be followed by a presentation and a discussion of the obtained solutions. Finally, applications of the models to cancer biology will be discussed.

## 6 Biological setting

Bacteria, which are single celled organisms, use carbohydrate as energy to power the chemical processes that enable them to survive and reproduce. Three carbohydrates in particular will be considered in the model applications, namely *glucose*, *lactose* and *arabinose*. These three sugars are available to the bacterium *E. coli* in its natural environment in mammalian intestines. Through a series of chemical reactions, *E. coli* can degrade carbohydrates into other molecules which it will then use for fuel and building materials. These reactions are catalyzed by proteins manufactured by the cell itself. The proteins involved in carbohydrate degradation are called **degradation proteins**.

The cell produces proteins from its protein blueprints: the genes. In a bacterium, this process happens as follows. A molecule called RNA polymerase will 'read' a gene and produce an RNA molecule, called mRNA, that contains the same information that is carried by the gene. This process is called **gene transcription**, since part of the chromosome DNA molecule is 'transcribed' to an RNA molecule. The information contained in the mRNA molecule can then be used by the cell to construct a protein. The cell constructs the protein by linking together a sequence of amino acids in accordance with the information contained in the mRNA molecule. The process of making a protein from an mRNA molecule is called **translation**, since information in the alphabet of RNA nucleotides is 'translated' into the language of amino acids. Thus, information is carried from the gene the through an intermediate mRNA molecule to the final protein product. The sum of the processes of transcription and translation is called **gene expression**.

Different carbohydrates are be available to *E. coli* at different times, and different degradation proteins are needed to degrade the different carbohydrates. At any given time, how does the cell know what degradation proteins to produce? A naive way of making use of carbohydrates would be for the cell to produce all kinds of degradation proteins at a constant rate. In this case, when carbohydrates enter the cell, regardless of their type,

degeneration proteins stand ready to metabolize them. Unfortunately, this approach is detrimental to the cell. The benefit of having energy instantly available does not weigh up for the energy needed for constant production of the degradation proteins.

While there are some genes that are being transcribed almost constantly (called house-keeping genes), most genes are **regulated**. A regulated gene is a gene that is expressed only when its gene product is needed. More specifically, and to be further explained below: *what is regulated is the access of RNA polymerase to the gene's DNA*. As a response to an environment with many different carbohydrates, the genes for the degradation proteins are regulated. For a good and concise introduction to the theory of gene regulation see [10].

## 6.1 Regulating genes with activators and repressors

The way gene regulation works is by signals. Production of the degradation proteins may start in the presence of a signal, or more specifically a signaling molecule, often called an inducer. Since different carbohydrates have different inducers, the cell knows which specific degradation proteins to produce when a carbohydrate is available. Though in the presence of a specific inducer, before producing the degradation proteins, the cell will first consider if degrading this carbohydrate is worth the metabolic effort. In other words, it will consider if the energy spent metabolizing the carbohydrate is worth the energy received from this process. There could be several reason why a cell would choose not to react on the presence of a carbohydrate's inducer. For instance other, more easily obtainable energy sources (usually other carbohydrates) might be available. If the cell chooses to react upon the carbohydrate's inducer, however, it will allow RNA polymerase increased access to the relevant DNA, and thus initiate production of degradation proteins. This 'considering' of the cell can be achieved with gene regulation.

There are several ways by which the bacterial cell can allow RNA polymerase increased access to DNA. What most of them have in common is that they involve regulators called **transcription factors**. Transcription factors are proteins that can bind to DNA, often close to the genes they regulate, and interact with RNA polymerase by either blocking or facilitating its access to DNA. Transcription factors can be categorized as **activators** or **repressors**, depending on if they increase or decrease the activity of RNA polymerase on DNA. Figure 6 shows the difference between the two modes of regulation in the presence of a yellow transcription mediating inducer molecule. While looking at this figure, consider also that inducers can play the opposite role. The inducer could bind instead of lift the repressor, and lift instead of bind the activator. Later, an example will be encountered where the inducer plays this role for the activator for a specific regulatory

mode in *E. coli* that is involved in metabolizing *lactose*.

**Repressor regulation.** The place where the repressor binds DNA is called the **repressor binding site**. When the repressor sits on DNA it **blocks** the access of RNA polymerase to DNA, by decreasing the chemical binding strength between the polymerase and DNA.

An example of a gene system[3] regulated by a repressor is the genes for degrading the *lactose* carbohydrate in *E. coli*. When *lactose* is not present in the cell, the repressor will be bound to DNA, blocking RNA polymerase's access to DNA, as in Figure 6 a). When *lactose* is present, however, the *lactose* inducer molecule interacts with the repressor in such a way that the repressor falls of DNA, thus allowing access to RNA polymerase, as in Figure 6 b).

**Activator regulation.** The action of the activator is opposite that of the repressor: when an activator is bound to the **activator binding site** it **facilitates** the access of RNA polymerase to DNA, by increasing the binding affinity of the polymerase. Put simple, the activator 'pulls' RNA polymerase down to the DNA.

An example of a gene system regulated by an activator is the genes for degrading *arabinose* in *E. coli*. When *arabinose* is not present in the cell, the activator is not able to bind DNA, as in Figure 6 c). This in turn keeps RNA polymerase from binding and subsequently initiating transcription. However, when *arabinose* is present, the *arabinose* inducer molecule interacts with the activator so that it can bind DNA, as in Figure 6 d). This in turn allows RNA polymerase to bind DNA via the activator and thus initiate transcription of the gene.

**Two or more regulators.** A single repressor or activator offers basic control of RNA polymerase's access to DNA. For even tighter regulation, two or more transcription factors can be used in combination.

An example of a gene system with two regulators is in fact the *lactose* degradation system considered above. What was not mentioned then is that an additional activator is almost constantly bound to DNA near the *lactose* system. The activator only falls off when the carbohydrate *glucose* is present in the cell. In the absence of *glucose*, the system is hence controlled by the repressor alone. However, when *glucose* is present a signal molecule causes the activator to unbind, efficiently halting gene transcription. This is the above mentioned case when the inducer plays the opposite role than in Figure 6. *Glucose* is not available to *E. coli* most of the time, so to say that the *lactose* system is regulated by a repressor is a good approximation. This extra mechanism has evolved because *E. coli* will metabolize *glucose* before any other carbohydrate, since *glucose* is the most efficient energy source available to *E. coli*.

The example just discussed shows how a gene's regulatory mechanism

[3] A system is a set of genes regulated by the same transcription factor(s).

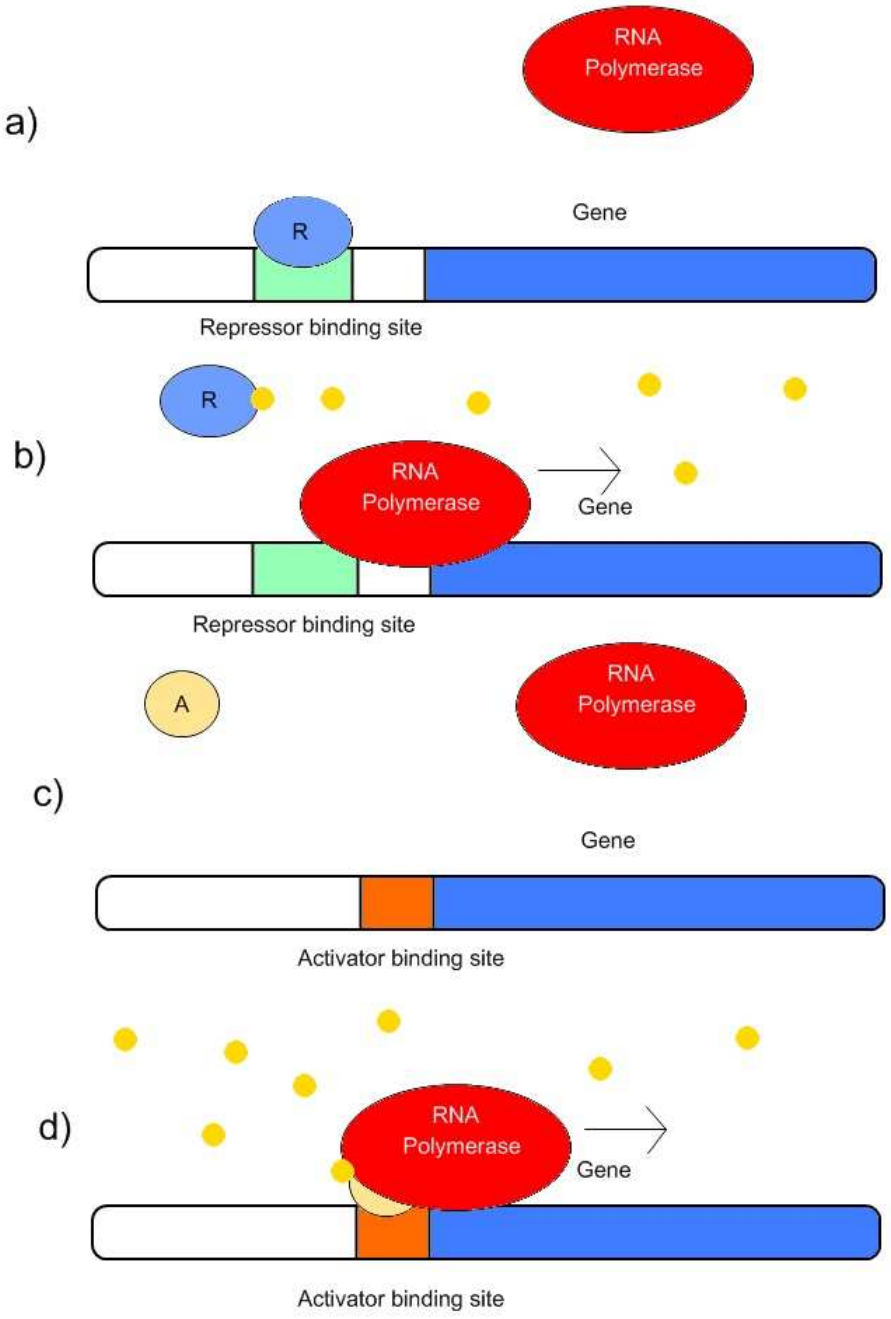


Figure 6: An inducer initiates gene transcription using two different regulatory mechanisms. In (a) the repressor blocks RNA polymerase's access to DNA, while in (b) inducer signals have caused the repressor to unbind its binding site, leaving DNA polymerase free to transcribe the gene. In (c) RNA polymerase does not bind DNA because the activator is not bound, while in (d) inducer signals have caused the activator to bind so that DNA polymerase might initiate transcription.

acts as a **logic gate**. For IF *glucose* AND *lactose* → use only *glucose*. IF NOT *glucose* AND *lactose* → use *lactose*. The more AND, IF and OR considerations the cell needs to make before initiation gene transcription, the more commands will be needed in the logic gate. In terms of biology: more transcription factors will be needed in the regulatory mechanism.

### 6.2 Predicting the outcome of evolution

Considering that many activators and repressors can work together to regulate gene transcription, and keeping in mind that each activator or repressor can be either activated or deactivated by an inducer, and that there exist even more regulatory functions than those considered here, it is evident that transcriptional regulation can be very complex indeed. Citing systems biologist Uri Alon: "(...) cells evolved to survive, and not for scientists to understand" [11]. From this complexity, evolution has chosen for each gene system a particular regulatory configuration of repressors and activators. Thus the question rises: for a given gene system, is there any way to predict the regulatory system as an outcome of evolution? This is one of the questions Alon tries to answer in his book "An Introduction to Systems Biology: Design Principles of Biological Circuits" [12]. The biological examples and much of the discussion in part III of this work is based on chapter 11 in this book. To answer the question if the outcome of evolution can be predicted, the models derived in part I of this work will be solved motivated by the Savageau demand rule and the rule of minimal error load to postulate outcomes of evolution of regulatory systems.

## 7 Defining and testing the Savageau demand rule

If gene expression in the presence of an inducer is desired, as in Figure 6, both regulation by activator and by repressor can achieve this, as can be seen in that Figure. Why, then, has evolution chosen different regulatory modes for the aforementioned *arabinose* and *lactose* degradation systems? This question was raised by Savageau [13], which he answered by formulating what is called the **Savageau demand rule**, which for the purposes of this work can be stated as: 'The mode of gene regulation is correlated with the demand for gene expression in the organism's natural environment. High demand genes are regulated by activators and low demand genes are regulated by repressors'.

The demand of a gene is a way of quantifying how often the function carried out by the gene product is needed in the cell. In the mathematical models, the demand of a gene will correspond to the frequency of time that the *signal* that prompts RNA polymerase to transcribe the gene is present in the cell.

## 7.1 The demand rule explained by stability against mutations

The argument by stability against mutations was the argument for the demand rule given by Savageau [12][13]. The observation behind the argument is that most mutations to an organism's DNA are harmful. In the context of gene regulation, a harmful mutation is in this argument taken to be one that causes loss of regulation. By **loss of regulation** it is meant that the regulating transcription factor (the activator or the repressor) *is no longer able to bind DNA* [12]. See Figure 7 for examples of loss of regulation in an activator and a repressor.

### 7.1.1 An example with a high demand gene

To understand the stability argument, assume as an example a gene that is in **high demand**. Assume also that gene transcription should be started in the presence of an inducer. Since the gene is in high demand, the inducer molecules will be almost constantly present in the cell.

First, assume that this gene is **regulated by an activator**. Loss of regulation in a mutant thus corresponds to the activator in that mutant becoming unable to bind DNA. If the activator can not bind DNA, RNA polymerase can in turn not bind either, see Figure 7 a), so that gene transcription can not be initiated. Since the gene is in high demand its protein product must be important. Loss of regulation causes a drastic reduction in the production of this protein, so the mutated cell will likely have a marked reduction in fitness compared to other, unharmed cells in the population. Therefore, mutants that have lost their regulatory function are *selected against strongly* and quickly removed from the population. Since mutants that have lost their regulatory function are quickly removed from the population, the regulation by activator is a *stable regulatory mode* in a high demand gene.

Next, assume that this gene is **regulated by a repressor**. Loss of regulation now results in constitutive (i.e. constant) gene expression, see Figure 7 b). Since the gene is in high demand its gene product is needed almost all the time. A mutant with lost regulation is producing the gene product all the time, so it does not have its fitness greatly reduced. Such a mutant might survive in the population unpunished. By genetic drift its descendants might eventually take over the population, causing the regulatory mode to be lost for all cells, and thus reducing the fitness of the entire population. When this population at some later point will encounter difficult living conditions or competition for resources with another population, it is more likely to die out because of its collectively reduced fitness. Thus, regulation by repressor is *not a stable regulatory mode* for a high demand gene.

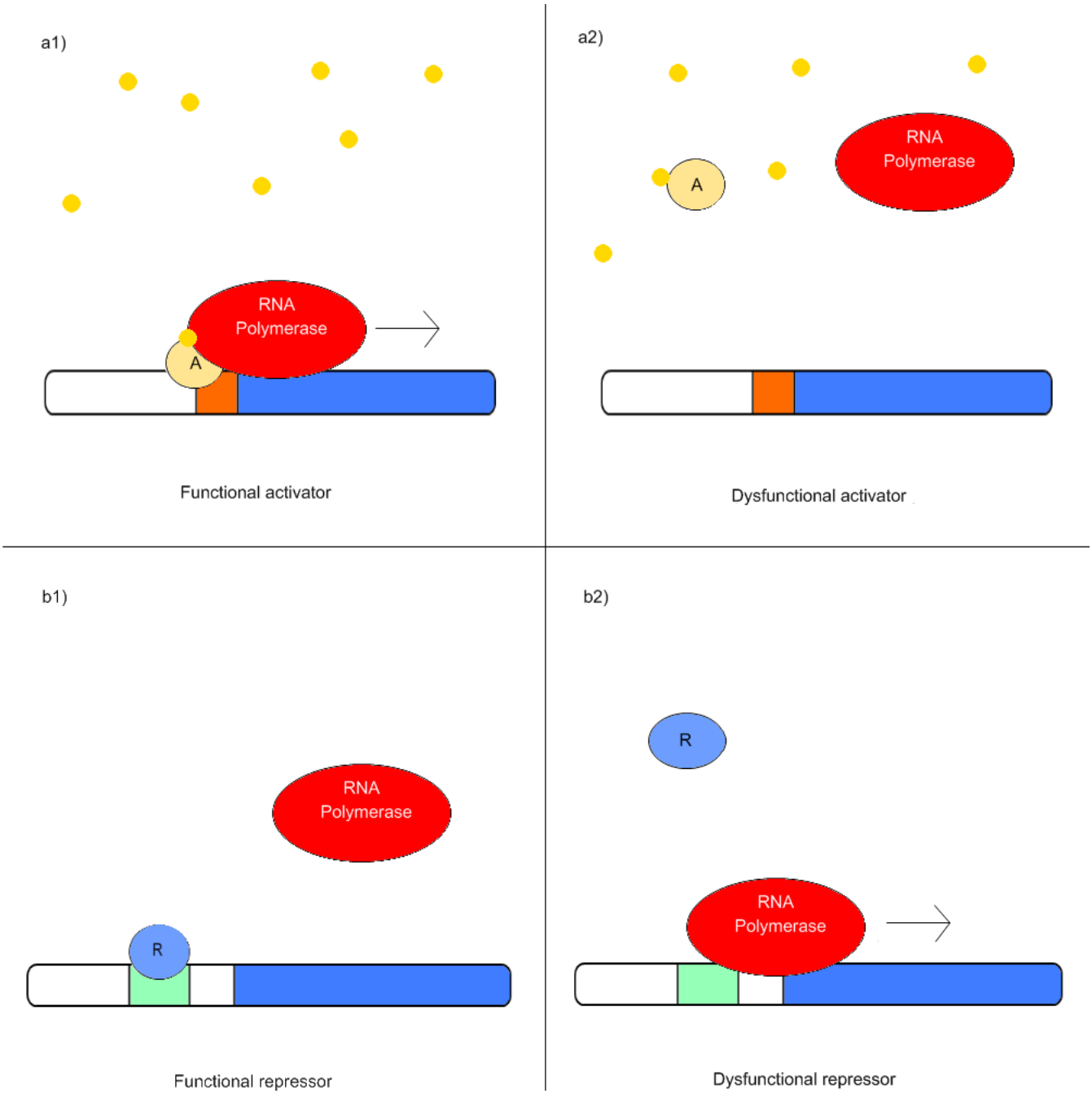


Figure 7: Comparing functional to dysfunctional regulators. (1a) The activator is unable to bind DNA, halting expression. (2b) The repressor is unable to bind DNA, leading to constitutive expression.

By having considered a high demand gene regulated by a repressor and an activator, it is concluded that regulation by activator is the more stable regulatory mode in high demand genes. For a gene in low demand, the argument is symmetrical, but will be sketched for clarity.

### 7.1.2 An example with a low demand gene

Consider a gene that is in low demand whose transcription should be started in the presence of an inducer.

First assume that this gene is regulated by an activator. For a mutant where regulation is lost, gene expression will stop all together. Since the gene was in low demand, one may assume that the gene product is not of critical or immediate importance to the cell, and therefore the fitness of the mutant is not greatly reduced. Since the fitness of the mutant is not greatly reduced, its descendants might take over the population by genetic drift, causing a fitness reduction for the entire population.

If on the other hand the gene is regulated by a repressor, loss of regulation in a mutant results in constitutive expression. The metabolic burden suffered by a mutant from constantly producing mostly unneeded proteins is assumed to be quite large, so that such a mutant will be at a strong selective disadvantage, and thus quickly removed from the population. It is therefore concluded that regulation by repressor is the more stable regulatory mode in low demand genes.

### 7.1.3 Conclusions and limitations of the stability argument

The stability argument concludes that regulation by activators is the most stable mode for high demand genes and that regulation by repressors is the most stable mode for low demand genes. This speaks in favor of the demand rule. Indeed, "[The] demand rule appears to be in agreement with 100 gene systems (...) from *E. coli* and other organisms (...)" [12]. Two examples are the *arabinose* and *lactose* degradation systems considered earlier. *Arabinose* is frequently available to *E.coli* and its system is regulated by an activator, and *lactose* is rarely available to *E.coli* and its system is regulated by a repressor.

The question then rises: does this apply to the cases with two or more regulators? Say that a gene should be regulated by two regulators. Will this gene be regulated by two activators if demand is high and by repressors if demand is low? As will be seen, this is not the case.

The stability argument has some limitations. First, it does not take into account that one regulatory mode can mutate into another, so that both modes may coexist in a population. If both regulatory modes coexisted in a population, one would have to consider if one of the modes has an inherent fitness advantage over the other. Only by assuming that there is no such

inherent fitness advantage to either regulatory mode is it plausible to test populations regulated by repressors and populations regulated by activators separately, as was done in the stability argument.

It is however known that relatively few mutations can cause a repressor to turn into an activator and vice versa [10]. If by mutations both modes are present in the population and one has a fitness advantage over the other, the advantageous mode might take over the population by natural selection, thus overriding the effect from stability against mutations. According to Alon "Mutant-selection arguments are valid only if there is no intrinsic fitness advantage to one of the two modes of control. If such intrinsic differences exist, they would dominate over the differential effect of mutations." [12]. In his book, Alon then introduces another argument, which originates from a paper that he co-authored [14] and which is based on the concept of minimal error load. However, before the error load argument is introduced, the Fokker-Planck equation will be used to show that the stability argument can be put to the test numerically and shown to be correct for the case of a single repressor or activator.

## 7.2 Testing the demand rule for a single regulator

The stability argument predicts that the regulatory mode chosen by evolution is the one that is most stable against detrimental mutations. The stability of the regulatory mode will be tested by assuming that a mutant with loss of regulation has just been introduced to a population, and the probability that the allele of the introduced mutant eventually becomes fixed in the population is used as a measure of the stability of the established regulatory mode.

The organisms of the population are assumed to be to *E. coli*, but the outcome of the test can be applied to haploid, asexual organisms in general. Allele A will correspond to the functional regulatory mode, initially present in every cell. Allele B will correspond to the introduced allele that causes loss of regulation. Allele A should be interpreted as the collection of DNA which ensures functional regulation. This is the DNA of the transcription factor binding site and the DNA of the gene that is the blueprint for transcription factor protein itself. Allele B should be interpreted as any mutant of allele A that causes loss of regulation. To measure the stability of allele A to withstand the introduced mutant, the probability that the introduced allele B will eventually become quasi-fixated in the population is used. From here on, when discussing quasi-fixation of alleles in the presence of mutations, the term 'quasi' will be dropped for convenience.

The evolution of the two regulatory modes in the population will be tested for both a high demand gene and a low demand gene. This gives four possible test scenarios (repressor low and high, and activator low and high), but because of the symmetry of the stability argument only two scenarios

need to be investigated. For simplicity, a high demand gene is defined as a gene where the inducing signal is present 90% of the time and a low demand gene is defined as a gene where the inducing signal is present 10% of the time. The number of generations within which demand varies is set to 80, the same as for the mean field simulations in part II. Applying these values for demand to the discussion in 7.1 results in the two selection reduction vectors for allele B as given in Figure 8 and Figure 9. These plots show during which generations allele B is selected against (when $S > 0$) and during which generations allele B is subject to genetic drift (when $S = 0$).

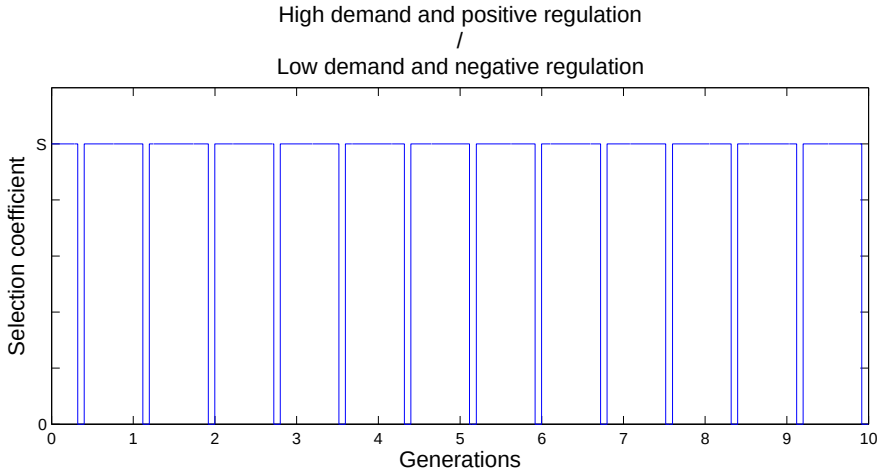


Figure 8: In these two scenarios, allele B is selected against most of the time. Generations are in units of $10^3$.

As can be seen from Figure 8 and Figure 9, the four scenarios offered by two regulatory modes and two types of demand are reduced to just two different selection vectors, and thus to just two numerical simulations. This is because it has been assumed that the fitness reduction from stopping production of a high demand gene product is the same as the fitness reduction from over expressing a low demand gene product. For real regulatory systems this can not be assumed to be true, but it is assumed here for the sake of simplicity.

1000 generations were simulated for $N = 100$ individuals for values of $S$ from 0 to 10 for a high demand gene under both repressor and activator regulation. A value of $S = 4$ corresponds to allele B having a fitness reduction of 0.04 per cent compared to allele A. Step lengths of $\Delta x = 1/100$ for the frequency and $\Delta t = 1/100^2$ for time were used. The initial value of allele A

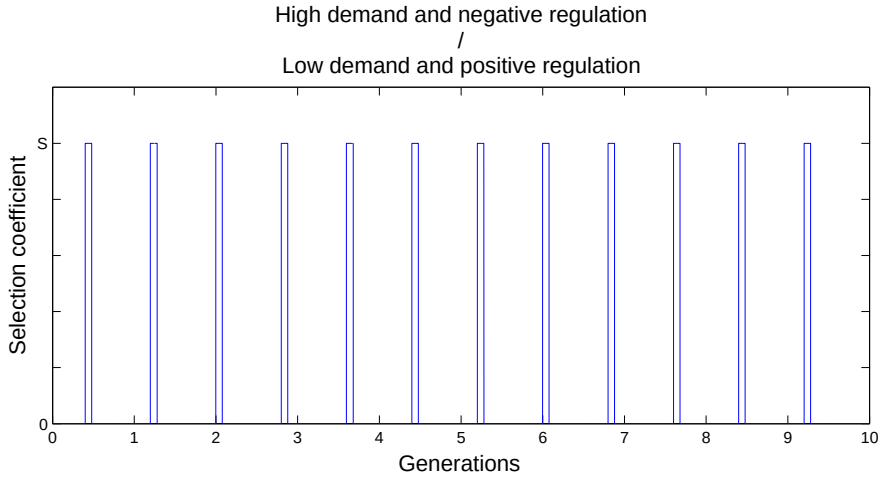


Figure 9: In these two scenarios, allele B is subject to genetic drift most of the time. Generations are in units of $10^3$.

was $f(0,t) = 0.99$. The mutation rates[4] are $N\mu = N10^{-7}$ and $n\nu = N10^{-6}$. The mutation rates values are so small however that they have little impact on the calculations.

For each simulation the probability that allele B had fixated in the population after the 1000 generations was stored. A plot of the final fixation probabilities is given in Figure 10. The fixation probabilities are calculated from the value of the numerical probability distribution at $f(0,t)$ where $t = 1000$. It is clear from Figure 10 that if a high demand gene is regulated by an activator, the chance of a dysfunctional mutant taking over the population is much smaller than if the gene is regulated by a repressor. This result validates the predictions of the demand rule. The results for a low demand gene are not shown since they mirror the results in Figure 10. They verify however that a low demand gene is more stable against mutations when regulated by a repressor, in compliance with the demand rule.

It can be asked if the use of fluctuating selection values in stead of a mean value for $s(t)$ has an effect on the fixation probabilities. In order to investigate this, the solutions obtained by solving the Fokker-Planck equation were compared to an exact formula for the fixation probability of allele B, which is valid when there are no mutations and $s(t)$ is constant. This

[4]It is assumed to be 10 times more likely to mutate from functional to dysfunctional regulator than the other way around.

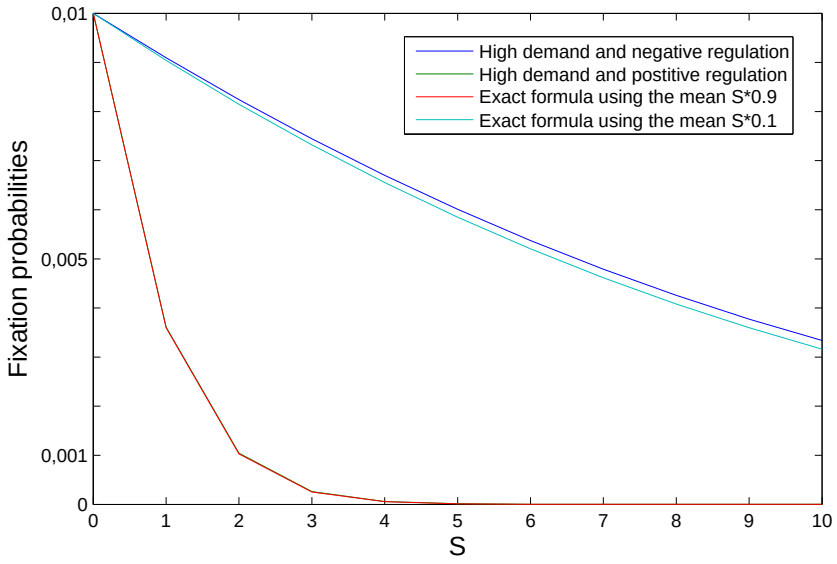


Figure 10: Fixation probabilities of the mutated allele B. Allele B has a much higher probability to become fixed when the high demand gene is regulated by a repressor. The selection values on the horizontal axis correspond to $S$ as shown in Figures 8 and 9. The solutions from the exact formula show that fluctuating selection values do not greatly affect the fixation probabilities for this example.

formula is given by (see [4] p. 119 and [8])

$$\Pi_0 = \frac{e^{-2S(N-1)/N} - e^{-2S}}{1 - e^{-2S}}. \tag{7.1}$$

By evaluating this formula for $N = 100$ and $S = \bar{S}$ found from the integral of $S(t)$ in Figures 8 and 9, fixation probabilities for allele B almost identical to those from the solution of the Fokker-Planck equation were found. The discrepancies as seen in Figure 10 are not significant over numerical sources of error. The conclusion is that fluctuating selection values do not affect the conclusions of the stability argument for the demand rule.

# 8 Defining and testing the rule of minimal error load

The argument behind the rule of minimal error load is that regulatory modes suffer fitness reductions - called error loads - when its transcription factors are not bound to DNA. Two sources of these errors will be explained: non-specific binding and residual binding.

A transcription factor is either bound to its binding site or diffusing in the cell. What causes a transcription factor, or any other protein for that matter, to bind DNA are forces between the protein molecule and the DNA molecule. A protein's binding affinity for a length of DNA is a function of the sequence of nucleotides along that length. What identifies the transcription factor's binding site is then a specific sequence of nucleotides that the transcription factor binds strongly. However, every protein has a probability to bind to every suitable length of DNA. A transcription factor may therefore bind other sites along DNA than its intended binding site. This is called **non-specific binding**. Thus, when a regulatory mode's binding site is vacant, other transcription factors than those associated with the regulatory mode may bind to this binding site, interfering with transcription. This interference may result in either increased or decreased transcriptional activity, each of which reduces the fitness of the cell [12].

**Residual binding** is when the transcription factor for a regulatory mode binds its own binding site when it is not supposed to. An signal molecule typically reduces the binding affinity of its transcription factor by only one to two orders of magnitude compared to when the inducer is absent [12]. Thus, even if the signals says that the transcription factor should be free of DNA, it can still bind, but with a reduced affinity.

Both non-specific binding and residual binding cause a fitness reduction for the cell when a binding site is unbound its regulatory protein. **The error load rule** can be stated as: 'evolution chooses the regulatory mode for each system so that the mode's transcription factors are bound most of the time.'

## 8.1 The argument by minimal error load for one regulator

For subsequent comparison with the Fokker-Planck and mean field models, an account of Alon's error load argument, using his notation, will now be given. Alon identifies what he calls the total error load for systems regulated by repressors and activator by $E_R$ and $E_A$, respectively. These values are given by the fraction of time the regulatory mode has an unbound transcription factor times the selection disadvantage suffered during this time:

$$E_R = ps_1 \tag{8.1}$$

$$E_A = (1-p)s_2, \tag{8.2}$$

where $p$ is the demand of the system, and the two $s$'es are the fitness reductions suffered when the transcription factor is not bound to its binding site.

A brief note: in this work's language, $E_S$ and $E_R$ can be identified as the integrals of the functions in Figure 8 and Figure 9 when $s_1 = s_2 = S$. From this identification one can use the selection estimates from the error load argument in the Fokker-Planck equation.

By assuming that evolution seeks the minimal error load for each system, Alon finds the parameter values for $p$, $s_1$ and $s_2$ where evolution would choose regulation by repressor by evaluating

$$E_R < E_A, \tag{8.3}$$

which gives that repressors are advantageous whenever

$$p < 1/(1 + \frac{s_1}{s2}). \tag{8.4}$$

This corresponds to repressor regulation for small values of $p$ and activator regulation for large values of $p$, which is in agreement with the argument by stability against mutations.

For an introduced mutant, being advantageous is not enough to achieve fixation. Alon states that only if the difference $E_A - E_R$ exceeds some threshold $s_{min}$ will a mutant that regulates by repressor achieve fixation in a population where regulation by activator is the default.

The model that Alon puts forth is good as it gives much insight with little mathematics. However, the error load argument is basically just a way of identifying fitness differences caused by different regulatory modes. The argument that evolution will choose regulation by repressor when $E_R < E_A$ and $E_A - E_R > s_{min}$, is just a way of saying that evolution will choose the most fit organism, and that a mutant organism can take over a population if its selective advantage is strong enough.

There are two immediate advantages gained from analysing the error load argument with the Fokker-Planck equation instead of with Alon's inequalities. The first is that exact fixation probabilities for each parameter

regime can be found. The fixation probabilities are a way of saying *when* the selective advantage is strong enough that one regulatory mode will overtake the other. The other is that by taking into account the effect of mutations, one can rule out the effect of fixation whenever $\mu N > 0.5$, as can be seen from Figure 1 in part II. Although $S = 0$ in that figure, fixation will not occur for any $\mathcal{O}(\epsilon)$-permitted value of $S = sN$ in the $\mu N > 0.5$ regime.

However, the Fokker-Planck equation will not be used in this work to investigate the error load argument for one transcription factor. In stead, the mean field equations will be used to study the the error load argument for a gene regulated by two regulators.

## 8.2 The argument by minimal error load for two regulators

By arguing as for the case with one regulator, Alon produces error load inequalities for two simultaneous regulators. While the error load argument supported the argument by stability against mutations in the case of a single repressor or activator, the inequalities he obtains show that the matter is more complicated when more advanced regulatory mechanisms are considered. This will be seen from simulations in the next section, where the case with two simultaneous regulators controlled by two inducers will be tested using the mean field equations derived in part I. With two simultaneous regulators, there are four possible regulatory modes: AA, AR, RA and RR, where R symbolizes regulation by a repressor and A symbolizes regulation by an activator. The order of the letters identify what regulatory mode the two inducers in the cell act on. This will be further explained below.

## 8.3 Testing the rule of minimal error load for two simultaneous regulators

Assume that the transcriptional activity of a gene Z in a cell is governed by two transcription factors that in turn are controlled by two signal molecules called L and G. There are many ways by which two signals and two transcription factors can regulate gene transcription; see [12] ch. 2 for a brief discussion. The combination that will be considered numerically in this section is as follows: transcription of gene Z should be initiated in the presence of signal L but should be shut down in the presence of signal G, regardless if L is present or not. This is an example considered by Alon in [12], and it can be shown to correspond to the gene regulatory mechanism for the *lactose* degradation proteins in *E. coli*.

A notation for when either signal is present or absent in the cell will be needed. If 0 and 1 denote the absence and presence of a signal respectively, (G,L) = (0,1) indicates that G is absent and L is present in the cell, and will be referred to as **state** (0,1). Table 2 shows how the regulatory mechanism should react in the different signaling states.

Table 2: *Desired response of the regulatory mechanism. The first entry in the parenthesis corresponds to G and the second entry corresponds to L. 1 denotes presence and 0 denotes absence of either signal.*

| ON | OFF |
|---|---|
| (0,1) | (1,1) |
| | (1,0) |
| | (0,0) |

The desired response of the regulatory mechanism to the two signals L and G as given in Table 2 can be realized with either of the four possible regulatory modes: $A_GA_L$, $A_GR_L$, $R_GA_L$ or $R_GR_L$. As mentioned above, the order of the letters show what inducer acts on what transcription factor; here indicated by a subscript.

The task at hand now is to identify when the transcription factors are bound and unbound for each of the four regulatory modes. When that identification is made, errors can be attributed to each mode whenever it has unbound transcription factors. These errors can then be used for the selection coefficients $S_1$, $S_2$, $S_3$ and $S_4$ in the mean field equations.

### 8.3.1 Analysing regulatory mode $A_GA_L$ in detail

Table 3 summarizes the analysis of regulatory mode $A_GA_L$.

Table 3: *Analysis of the $A_GA_L$ regulatory mode. $p_{ij}$ is the fraction of time the system spends in state (i,j) and $S_A$ and $S_R$ are the selection disadvantages suffered when respectively an activator and a repressor are unbound their binding sites.*

| (G,L) | Transcription mode | # of bound transcription factors | Fitness reduction |
|---|---|---|---|
| (0,1) | ON | 2 | 0 |
| (1,1) | OFF | 1 | $p_{11}S_A$ |
| (1,0) | OFF | 0 | $2p_{10}S_A$ |
| (0,0) | OFF | 1 | $p_{00}S_A$ |

In the first row the state is (0,1), so the transcription mode is ON. The way by which transcription can be ON in a mechanism regulated by two activators, is if both activators are bound to their binding sites. Since both transcription factors are bound, the error load argument states that while in this state, the regulatory mode $A_GA_L$ suffers no reduction of fitness.

In the second row the state is in (1,1), so the transcription mode is OFF. By considering the row, it is understood that the activator controlled by G is bound only for $G = 0$. Since $G = 1$ in this state, the conclusion is that the activator controlled by G is not bound in state (1,1). Similarly, it is seem from the first row that the activator controlled by L is bound only when $L = 1$. Since $L = 1$ in this state, the conclusion is that the activator controlled by L is bound in this state (1,1). Hence, only the activator controlled by L is bound to DNA in state (1,1). The regulatory mode $A_G A_L$ thus suffers a fitness reduction $S_A$ when in state (1,1).

The remaining two rows are treated in the same way, and so are the modes AR, RA and RR. The fitness reductions of each mode in each state are given in Table 4. Note that the row for mode AA corresponds to the last column in Table 3. The rows for the other modes correspond to the last columns of their corresponding tables (not shown). From Table 3 the selection coefficients for each regulatory mode is found.

Table 4: *Selection coefficients for the regulatory modes in each of the different signal states.*

| Mode | (0,1) | (0,0) | (1,0) | (1,1) |
|---|---|---|---|---|
| AA | 0 | $p_{00}S_A$ | $2p_{10}S_A$ | $p_{11}S_A$ |
| AR | $p_{01}S_R$ | 0 | $p_{10}S_A$ | $p_{11}(S_A + S_R)$ |
| RR | $2p_{01}S_R$ | $p_{00}S_R$ | 0 | $p_{11}S_R$ |
| RA | $p_{01}S_R$ | $p_{00}(S_R + S_A)$ | $p_{10}S_A$ | 0 |

### 8.3.2 Why the demand rule is no longer applicable

For this system, demand is defined as the fraction of time the system spends in state (0,1). The reason why demand is no longer a predictor of regulatory mode is simply because there are now four possible states, instead of just two as for the case with one regulator. While the fraction of time spent in the ON state (0,1) determines demand, there are now three OFF states. The time spent in each OFF state contributes differently to the fitness reduction of the regulatory mode, as can be seen from Table 4. For the case with only one regulator, there was only one OFF state, so there could be no variation in fitness reduction.

The specific experimental setting for testing the error load argument will now be defined. Assume that a gene Z is in low demand, which again is taken to mean that the cell spends 10% of its time in state (0,1), so that $p_{01} = 0.1$. The times spent in the remaining states can in principle be distributed freely, subject to $p_{00} + p_{10} + p_{01} + p_{11} = 1$. Two distributions of these values will be considered in the simulations.

In the first simulation, the cell spends equally much time in the remaining states. This means that $p_{00} = p_{10} = p_{11} = 0.3$. Using these values, the evolution of the allele frequencies in the population is given in Figure 11. In this figure the regulatory mode RA stands out as dominant most of the

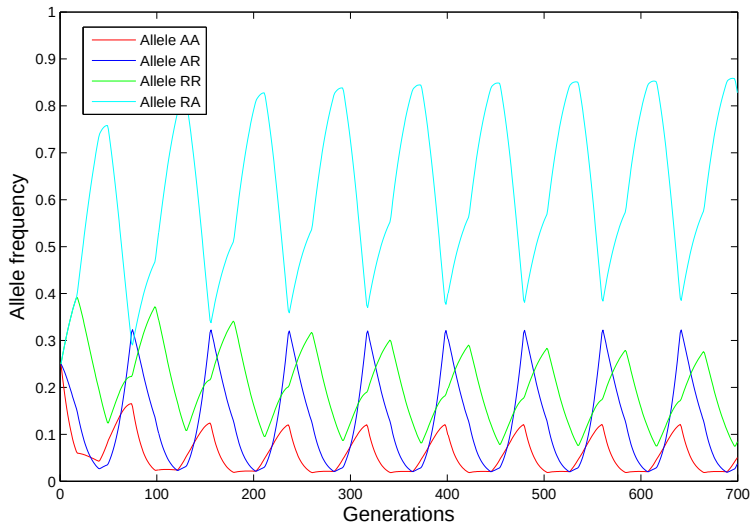


Figure 11: The evolution of allele frequencies in a population that is in the ON state (0,1) 10% of the time and in the states (1,0), (0,0), and (1,1) 30% of the time. Regulatory mode RA dominates the population most of the time.

time.

In the second simulation, the cell spends time in the remaining states according to the following values: $p_{00} = 0.8$, $p_{10} = 0.1$, and $p_{11} = 0$. This corresponds to a situation where both signals occur rarely in the cell, and where they never occur together. The evolution of the population using these values is given in Figure 12. In this figure it is clear that regulatory mode AR is dominant. It is not difficult to imagine that due to stochastic effects it could become fixed in the population with a high probability.

In both simulations, the model parameters are given by $S_R = S_A = 6$ and $U_{ij} = 0.1$ for all $i$ and $j$. Further, all four alleles are initially present at a frequency of 0.25. Although this is not a likely real-world scenario, the main point of this simulation is to show that the demand rule cannot be used for regulatory modes with more than one transcription factor. Comparing the Figures for the two simulations, one sees that even though in both cases the gene Z is in low demand, the error load argument predicts two different evolutionary scenarios. This demonstrates that the question of regulatory mode is no longer a question of demand. Indeed, by going through differ-

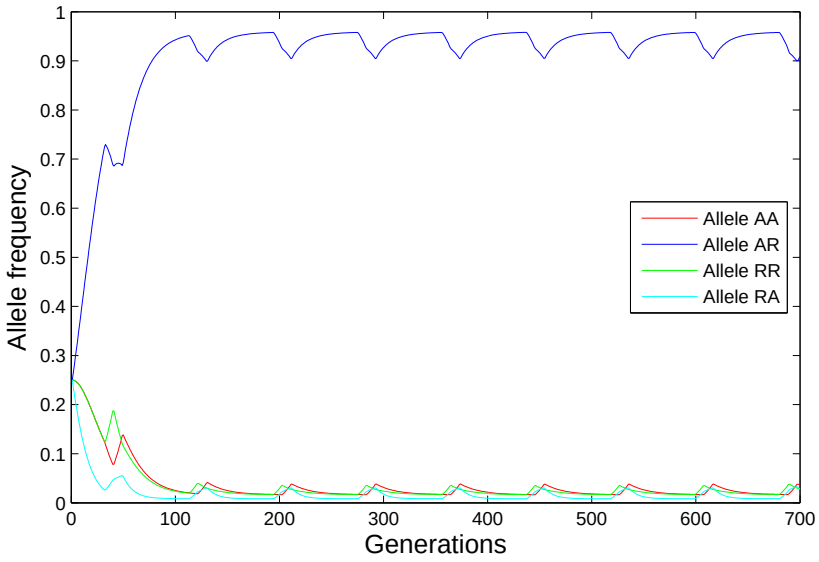


Figure 12: The evolution of allele frequencies in a population that is in the ON state (0,1) 10% of the time and in states (1,0), (0,0) and (1,1) 10%, 80% and 0% of the time, respectively. Regulatory mode AR dominates the population.

ent distributions for the values $p_{ij}$, any of the four modes can be made to dominate the population.

The values for $p_{01}$, $p_{00} = 0.8$, $p_{10} = 0.1$, and $p_{11} = 0$ in the last simulation are inspired by the *lactose* degradation system. Gene Z can symbolize the gene system that codes for the *lactose* degradation proteins, and G and L are the *glucose* and *lactose* inducer molecules respectively. The reason why the $p_{ij}$ are distributed like this for the *lactose* system, is because when *glucose* is present in the cell the access of *lactose* and other carbohydrates to the cell is blocked [12]. Further, both *glucose* and *lactose* are rare in *E. coli's* environment, which explains why the cell spends most of its time in state (0,0). As predicted by the simulation, the *lactose* system in *E. coli* is indeed regulated in this fashion: the *glucose* inducer controlls the activator and the *lactose* inducer controlls the repressor.

# 9 Applications to other areas in biology

The population genetics models have in this work been applied to the evolution of gene regulatory mechanisms in a population. The way that gene regulation could be described by the population genetics models, was that different gene regulatory modes could be identified as different alleles and

selection coefficients could be associated with each allele through the arguments of stability against mutations and error load. In the same way, any characteristic in an organism that can be realistically identified with a gene that has alleles of varying fitness can be modeled in the same way.

A field where population genetics models are in frequent use is the modeling of the evolution of cancer. Put simply, cancer is a collective term for cells that grow abnormally fast. In a population of normal cells, the initial, fast growing cancer cell is produced from the normal cells by mutations. The mutation rate at which that happens can be use in population genetics models. The different growth rates between cancer cells and normal cells provide a way of quantifying the cancer cells' selectional advantage. Assume that normal cells have growth rate $r_1$ and cancer cells have growth rate $r_2$ (the growth rate of cells can be measured in cell divisions per hour). One can then identify the normal cells' selective disadvantage by $s = r_1/r_2$ [15]. With both mutation rates and selection coefficients readily available, cancers are well suited for the population genetics models found in part I.

A study that uses the Moran model to study the evolution of so called cancer stem cells is one by Dingli *et al.* [16]. It is argued that since the number of stem cells in certain compartments of the body are kept constant by intercellular communication, the constant population size Moran model is well suited to describe the evolution of stem cells. Stem cells divide slowly, and may either divide into two new stem cells, one stem cell and one differentiated cell, or two differentiated cells. The stem cells alternate between these three types of division to both keep the number of stem cells constant as well as to be able to renew the surrounding tissue with fresh cells. A differentiated cell is a specialized cell, the specialization of which depends on the tissue that the stem cell resides in [16].

It is hypothesized that a mutant stem cell that divides into two new stem cells more often than it should, might be a source of cancer. Dingli *et al.* use the Moran model to show that such a mutant will take over the stem cell pool rapidly. Based on this, the conclusion is that research is needed to identify genes in stem cells that play a part in deciding which of the three types of division the stem cell chooses, to better be able to understand tumor growth and possibly identify novel targets for cancer therapy.

A study by Nowak *et al.* uses the Moran model to study cancer in the colon [15]. This study highlights some of the limitations of the Moran and Wright-Fisher models. "The Moran process describes evolutionary dynamics in a well mixed population of cells. All cells are in equivalent positions (...). There are no spatial effects." [15]. The authors then construct a stochastic process, called the linear process, that does take into account the architecture of so called colonic crypts. Using fixation probabilities, they show that the linear process predicts a much lower number of cancers than the Moran model. Since the linear process is assumed to better describe the colonic crypts, they argue that "(...) patterns of cell division in tissues

of multicellular organisms have evolved to delay the onset of cancer." [15], using the Moran model as a kind of evolutionary neutral reference.

# Part IV
# Concluding remarks

## 10 Summary

In this work, two models for population genetics were developed, namely the Wright-Fisher model and the Moran model. The Wright-Fisher model had its foundation in stochastic sampling from a binomial distribution, while the Moran model was a stochastic birth-and-death model. It was shown that the continuous approximations of both models yielded the same Fokker-Planck equation. As well, equations that describe the change of the mean allele frequency over time were deduced for both models.

Two results of interest can be noted from derivations of the models. The first is that an extension to the continuous Wright-Fisher model was suggested by introducing variable population size. This extension not part of text book material on the subject. The second is that a Fokker-Planck approximation of the Moran model with selection was derived. This is an extension of the result obtained in [6] where $s = 0$.

A solution method for the Fokker-Planck equation with constant population size by finite differences was introduced, and boundary values of this equation and the concept of fixation were discussed. Both the Fokker-Planck equation and the mean field equations were applied to simulate an evolutionary process where different gene regulatory mechanisms were 'competing' against each other. The results of these simulations were used to argue that the Savageu demand rule is correct for the case for regulation with one transcription factor, but does in general not hold when the number of transcription factors involved in regulation exceeds 1. Additionally, fluctuating selection values compared to mean values did not affect the conclusions of the mutational argument for the demand rule for the example considered. Finally, applications of population genetics models to the field of cancer biology were discussed.

## 11 Discussion

### 11.1 The Wright-Fisher model for $N(t)$ variable

As mentioned, the continuous Wright-Fisher model, or Kolmogorov forward equation, with a variable population size is not text-book material on the subject. One reason for this could be that while the Kolmogorov forward

equation describes the allele probability distribution, it is the Kolmogorov backward equation that has been used to find formulas for interesting results such as fixation probabilities and the mean time until fixation. Indeed, to investigate the effect of logistic growth and decline of the population size, Kimura and Otha used the Kolmogorov backward equation [17]. However, McKane and Waxman have have recently identified the singularities appearing at the boundaries of the Kolmogorov forward equation as fixation probabilities. As such, a similar investigation to theirs for $N(t)$ variable could allow workers to identify singularities at the boundaries as fixation probabilities for the Kolmogorov forward equation with variable population size. This would abolish the need for the backward equation in an investigation of fluctuating population size.

## 11.2 The validity of the simulation results

The results of the numerical simulations have been used to argue what gene regulatory mechanism in *E. coli* evolution favors in different environments. Of course, these conclusions are valid only if the models themselves can be thought to accurately describe the bacteria in their natural setting: the mammalian intestines. No such justification has been given in this work, and thus the presented results should be considered with some sceptisism. For example, in the sample solution of the Fokker-Planck equation, the population size was assumed to be $N = 100$, which is far from an accurate estimate of the number of *E.coli* cells in the gut. The actual number is $\mathcal{O}(10^{10})$. Using that number would however not give an accurate model description either, since the total population is likely divided into subpopulations, each of which may be describable as a population on its own, with possibly some migration of cells between populations. However, in principle a justification of this model application can be made. In order to do so, a thorough literature search on *E. coli* must be made, or the modeler must engage in collaboration with biologists. Then hopefully one can identify plausible values of $N$; decide if one should take into account fluctuating mutation rates; determine how often the different carbohydrates are available to *E. coli*; and decide if other evolutionary mechanisms, such as migration between populations, should be included in the model. This however is outside the scope of this work.

## 11.3 Applications of the rule of minimal error load

It has been suggested in this work that the rule of minimizing error load should be used to predict the regulatory mode of genes or gene systems in bacteria. This should be tested by comparing the predictions of this rule with more real world examples. If found to be correct, this argument gives novel insight into the best way by which a cell can regulate its genes.

This information is not just of pure theoretical use. Synthetic biology is an emerging field within biotechnology [18] that aims to create designer organisms. To see how far this field has come, a team recently synthesized an entire bacteria's genome from scratch [19] by using other bacteria and yeast as molecular machines. If one should create a synthetic bacteria, then one would inevitably have to decide how to regulate that bacteria's genes. As has been seen in this work, there are many ways by which gene regulation may be realized in the cell. If the error load argument holds, researches can use this information to design the gene regulatory mode for their synthetic bacteria for the best possible result.

# 12 Suggestions for further studies

The argument by stability against mutations and error load have been presented as two different arguments. However, if one solved the Fokker-Planck equation for 3 alleles, one for repressor regulation, one for activator regulation and one for lost regulation, the stability argument and the error load argument can be combined in one model. The statement quoted by Alon in part 3: "Mutant-selection arguments are valid only if there is no intrinsic fitness advantage to one of the two modes of control. If such intrinsic differences exist, they would dominate over the differential effect of mutations.", is of course a simplification of the matter. Using a model with three alleles, one could investigate exactly *when* intrinsic fitness differences dominate over stability against mutations.

Another interesting prospect is to solve the Fokker-Planck equation for variable population size. In order to do this, effects at the boundary must first be rigorously investigated. This investigation is outside the scope of this work. There are several interesting population size scenarios one could consider. One would be to investigate increasing and decreasing logistic functions for $N(t)$ to verify the result obtained by Kimura and Ohta, namely that a selectively advantageous allele has an has a "higher chance of avoiding extinction in an increasing population than in a declining population" [17]. Another fascinating study would be to investigate the effect of so called population bottlenecks, which is when a population's size decreases rapidly in a short period of time, sometimes followed by a period of rapid growth. It is a well known result in evolutionary theory (as can be shown using other population genetics models) that in populations of small sizes genetic drift is a stronger force than what it is in large populations, where selection becomes a more dominant force.

# Appendices

## A The diffusion approximation for two alleles and $N(t)$ constant

### A.1 Approximating $\phi(x', t)$

Multiplying $\mu$, $\nu$, and $s$ by the bookkeeping parameter $\epsilon = 1$, (2.8) one finds

$$\begin{aligned}\phi(x', t) &= \frac{\epsilon\mu + (1 - \epsilon\nu - \epsilon\mu)x'}{1 - \epsilon s(1 - \epsilon\mu) + \epsilon s(1 - \epsilon\nu - \epsilon\mu)x'} \\ &= \frac{\epsilon\mu + (1 - \epsilon\nu - \epsilon\mu)x'}{1 - \epsilon s(1 - x') + \mathcal{O}(\epsilon^2)} \\ &= (\epsilon\mu + (1 - \epsilon\nu - \epsilon\mu)x')[1 + \epsilon s(1 - x') + \mathcal{O}(\epsilon^2)] \\ &= x' + \epsilon M(x', t) + \mathcal{O}(\epsilon^2)\end{aligned} \tag{A.1}$$

where

$$M(x', t) = \mu + s(t)(1 - x')x' - (\nu + \mu)x'.$$

### A.2 Moments of $\Delta x|x'$

The moments of $\Delta x|x'$ are found using that $n$, the number of A alleles in the next generation, is binomially distributed when $n'$, the number of A alleles in the previous generation, is known. To shorten expressions, $\phi(t, x')$ is written as $\phi$ and $M(x', t)$ as $M$. Further $N$ is considered so large that $1/N < \epsilon$.

$$\begin{aligned}\mathrm{E}[\Delta x|x'] &= \mathrm{E}[x - x'|x'] \\ &= \mathrm{E}[x|x'] - x' \\ &= \frac{1}{N}\mathrm{E}[n|n'] - x' \\ &= \phi - x' \\ &= \epsilon M + \mathcal{O}(\epsilon^2).\end{aligned}$$

$$\begin{aligned}\mathrm{E}[\Delta x^2|x'] &= \mathrm{Var}(\Delta x) + \mathrm{E}[\Delta x|x']^2 \\ &= \mathrm{Var}(\Delta x) + \mathcal{O}(\epsilon^2).\end{aligned}$$

$$\begin{aligned}
\mathrm{Var}(\Delta x|x') &= \mathrm{Var}(x - x'|x') \\
&= \frac{1}{N^2}\mathrm{Var}(n|n') \\
&= \frac{\phi(1-\phi)}{N} \\
&= \frac{x'(1-x') + \epsilon M(1-2x')}{N} + \mathcal{O}(\epsilon^2) \\
&= \frac{\epsilon x'(1-x')}{N} + \mathcal{O}(\epsilon^2),
\end{aligned}$$

since $\epsilon/N < \mathcal{O}(\epsilon^2)$ and $1/N = \mathcal{O}(\epsilon)$. To find the higher moments of $\Delta x|x'$, the moment generating function for the binomial distribution, given by

$$\Psi_m(\phi, t) = (1 - \phi - \phi e^t)^m, \tag{A.2}$$

was differentiated for $m = 3$ to find

$$\mathrm{E}[x^3|x'] = \frac{(N-1)(N-2)\phi^3 + 3(N-1)\phi^2 + \phi}{N^2}.$$

Next, from $\phi(x', t) = x' + \epsilon M + \mathcal{O}(\epsilon^2)$ the following identities were found:

$$\begin{aligned}
\phi^2(x', t) &= (x')^2 + 2\epsilon x' M + \mathcal{O}(\epsilon^2) \\
\phi^3(x', t) &= (x')^3 + 3\epsilon (x')^2 M + \mathcal{O}(\epsilon^2).
\end{aligned}$$

Thus

$$\begin{aligned}
\mathrm{E}[\Delta x^3|x'] &= E[x^3|x'] - 3x'E[x^2|x'] + 3(x')^2 E[x|x'] - (x')^3 \\
&= (1 - \frac{1}{N})(1 - \frac{2}{N})\phi^3 + 3(1 - \frac{1}{N})\frac{\phi^2}{N} + \frac{\phi}{N^2} - 3x'\left[(1 - \frac{1}{N})\phi^2 + \frac{\phi}{N}\right] \\
&\quad + 3(x')^2\phi - (x')^3 \\
&= (1 - \frac{3}{N})\phi^3 + \frac{3\phi^2}{N} - 3x'\left[(1 - \frac{1}{N})\phi^2 + \frac{\phi}{N} - x'\phi\right] - (x')^3 + \mathcal{O}(\epsilon^2) \\
&= (x')^3 + 3(x')^2\epsilon M - \frac{3(x')^3}{N} + \frac{3(x')^2}{N} \\
&\quad - 3x'\left[(x')^2 + 2x'\epsilon M - \frac{(x')^2}{N} + \frac{x'}{N} - (x')^2 - x'\epsilon M\right] \\
&\quad - (x')^3 + \mathcal{O}(\epsilon^2) \\
&= \mathcal{O}(\epsilon^2).
\end{aligned} \tag{A.3}$$

## A.3 The Fokker-Planck equation

The derivation in this section is partly adopted from [4].

Rewrite

$$P(n, t+1) = \sum_{n'=0}^{N} P(n|n')P(n', t) \tag{A.4}$$

as

$$P\left(N\frac{x}{N}, t+1\right) = \sum_{n'=0}^{N} P\left(N\frac{x}{N}|N\frac{x'}{N}\right) P\left(N\frac{x'}{N}, t\right). \tag{A.5}$$

This equation can in turn be written as

$$\bar{P}(x_n, t+1) = \sum_{n'=0}^{N} \bar{P}(x_n|x_{n'})\bar{P}(x_{n'}, t). \tag{A.6}$$

Now scale the equation by $N$ so that

$$N\bar{P}(x_n, t+1) = \sum_{n'=0}^{N} N\bar{P}(x_n|x_{n'})N\bar{P}(x_{n'}, t)\delta x, \tag{A.7}$$

where $\delta x = 1/N$. Rewrite this as

$$f(x_n, t+1) = \sum_{n'=0}^{N} f(x_n|x'_n)f(x'_n, t)\delta x. \tag{A.8}$$

Letting $N \to \infty$ this can be approximated by

$$f(x, t+1) = \int_S f(x|x')f(x', t)\mathrm{d}x', \tag{A.9}$$

where $S \in [0, 1]$. $n$ is now considered to be a continuous variable and $f(x, t)$ is taken to be a continuous probability distribution with the same moments as $P(x_n, t)$. Define the space $\mathcal{D}$ of functions $Q(x) \in C^\infty[0, 1]$ with compact support on $S$ and boundary values

$$Q^{(k)}(0) = Q^{(k)}(1) = 0 \tag{A.10}$$

for all non-negative integers $k$. Because of the boundary values, $\mathcal{D}$ is a linear test space. Multiplying both sides of (A.9) by $Q(x)$ and integrating over $x$, one obtains

$$\int_S f(x, t+1)Q(x)\mathrm{d}x = \iint_S f(x|x')f(x', t)Q(x)\mathrm{d}x'\mathrm{d}x. \tag{A.11}$$

The left hand side of this equation is now associated with the distribution $f^{t+1}$:

$$(f^{t+1}, Q) = \int_S f(x, t+1)Q(x)\mathrm{d}x \tag{A.12}$$

A Taylor expansion of $Q(x)$ about $x's$ gives

$$Q(x) = Q(x') + \frac{\partial Q(x')}{\partial x}\Delta x + \frac{1}{2}\frac{\partial^2 Q(x')}{\partial x^2}(\Delta x)^2 + O((\Delta x)^3),$$

where $\Delta x = x - x'$. $Q(x')$ and its derivatives are in $\mathcal{D}$ since $x' \in S$. Inserting the expansion for $Q(x)$ in the right-hand side of (A.11) gives

$$= \int_S f(x',t)\left[\int_S f(x|x')\mathrm{d}x\right] Q(x')\mathrm{d}x' + \int_S f(x',t)\left[\int_S \Delta x f(x|x')\mathrm{d}x\right]\frac{\partial Q}{\partial x}(x')\mathrm{d}x' + \frac{1}{2}\int_S f(x',t)\left[\int_S (\Delta x)^2 f(x|x')\mathrm{d}x\right]\frac{\partial^2 Q}{\partial x}(x')\mathrm{d}x' + \mathcal{O}(\epsilon^2).$$

The first omitted term in this equation is an integral with integrand $\mathrm{E}[(\Delta x)^3|x']$, which is $\mathcal{O}(\epsilon^2)$. Higher order terms correspond to integrals of higher order moments, which are $\mathcal{O}(\epsilon^2)$ as well. The terms within the squared brackets correspond to 1, $\mathrm{E}[\Delta x|x']$ and $\mathrm{E}[(\Delta x)^2|x']$ respectively, which are inserted to give

$$= \int_S f(x',t)Q(x')\mathrm{d}x' + \epsilon\int_S f(x',t)M\frac{\partial Q}{\partial x}(x')\mathrm{d}x' + \frac{\epsilon}{2N}\int_S (x')f(x',t)D\frac{\partial^2 Q}{\partial x}(x')\mathrm{d}x' + \mathcal{O}(\epsilon^2),$$

These terms are now identified as distributions. The variable $x'$ changes name to $x$ since the disctinction between the two is no longer needed. The terms with differentiated test functions are interpreted as differentiations in the distributional sense. Thus

$$= (f^t, Q) - (\epsilon\frac{\partial f^t M}{\partial x}, Q) + (\frac{\epsilon}{2N}\frac{\partial^2 f^t D}{\partial x^2}Q(x)) + \mathcal{O}(\epsilon^2). \tag{A.13}$$

Inserting this expression into (A.11) and rearranging terms one obtains

$$(f^{t+1} - f^t, Q(x)) = (H, Q) + \mathcal{O}(\epsilon^2), \tag{A.14}$$

where

$$H(x,t) = -\epsilon\frac{\partial}{\partial x}[f(x,t)M(x,t)] + \frac{\epsilon}{2N}\frac{\partial^2}{\partial x^2}[f(x,t)D(x)].$$

Thus the following difference equation is satisfied in the distributional sense:

$$f(x,t+1) - f(x,t) = -\epsilon\frac{\partial}{\partial x}[f(x,t)M(x,t)] + \frac{\epsilon}{2N}\frac{\partial^2}{\partial x^2}[f(x,t)D(x)] + \mathcal{O}(\epsilon^2). \tag{A.15}$$

To go from difference to differential equation, the following change of coordinates on the time variable is introduced:

$$\tau = \frac{t}{N}. \tag{A.16}$$

Thus

$$f(x, t + \Delta t) = f(x, N\tau + \Delta\tau) = \hat{f}(x, \tau + \Delta\tau), \tag{A.17}$$

where $\Delta t = 1$ and $\Delta\tau = \frac{1}{N}$. Dividing (A.15) by $\Delta\tau$ one obtains

$$\begin{aligned}\frac{\hat{f}(x, \tau + \Delta\tau) - \hat{f}(x, \tau)}{\Delta\tau} = &-\frac{\partial}{\partial x}[\hat{f}(x, \tau)\hat{M}(x, \tau)] \\ &+ \frac{\epsilon}{2)}\frac{\partial^2}{\partial x^2}(\hat{f}(x, \tau)D(x)) + \mathcal{O}(\epsilon^2).\end{aligned} \tag{A.18}$$

Here $\hat{M}(x, \tau)$ is $NM(x, \tau)$. This is taken to be a scaling of the model parameters, so that

$$\hat{M}(x, \tau) = U + S(1 - x)x - (V + U)x, \tag{A.19}$$

where $U$, $V$ and $S$ are the model parameters $\mu$, $\nu$ and $s$ multiplied with $N$. By letting $N \to \infty$ the above difference equation becomes a differential equation,

$$\frac{\partial f(x, t)}{\partial t} = -\frac{\partial}{\partial x}[f(x, t)M(x, t)] + \frac{1}{2}\frac{\partial^2}{\partial x^2}[f(x, t)D(x)], \tag{A.20}$$

to $\mathcal{O}(\epsilon^2)$. Abuse of notation has been used in this equation by writing $\hat{f}(x, \tau) = f(x, t)$ and $\hat{M}(x, t) = M(x, t)$.

# B The diffusion approximation for two alleles and $N(t)$ variable

## B.1 The moments of $\Delta x|x'$

It is assumed that $\frac{1}{N(t)} < \epsilon$ for all $t$. The moments of $\Delta x|x'$ are then given by:

$$\begin{aligned}\mathrm{E}[\Delta x|x'] &= \mathrm{E}[x - x'|x'] \\ &= \frac{1}{N(t+1)}N(t+1)\phi - x' \\ &= \phi - x' \\ &= \epsilon M + \mathcal{O}(\epsilon^2).\end{aligned}$$

$$\begin{aligned}\mathrm{E}[\Delta x^2|x'] &= \mathrm{Var}(\Delta x) + \mathrm{E}[\Delta x|x']^2 \\ &= \mathrm{Var}(\Delta x) + \mathcal{O}(\epsilon^2)\end{aligned}$$

$$\begin{aligned}\mathrm{Var}(\Delta x|x') &= \mathrm{Var}(x - x'|x') \\ &= \frac{1}{N^2(t+1)}\mathrm{Var}(n|n') \\ &= \frac{\epsilon x'(1-x')}{N(t+1)} + \mathcal{O}(\epsilon^2).\end{aligned}$$

$$\mathrm{E}[\Delta x^3|x'] = \mathcal{O}(\epsilon^2) \tag{B.1}$$

as for the case when $N(t)$ is constant.

## B.2 Obtaining the The Fokker-Planck equation

Rewrite

$$P(n, t+1) = \sum_{n'=0}^{N(t)} P(n|n')P(n', t) \tag{B.2}$$

as

$$P\left(N(t+1)\frac{n}{N(t+1)}, t+1\right) = \sum_{n'=0}^{N(t)} P\left(N(t+1)\frac{n}{N(t+1)}|N(t)\frac{n'}{N(t)}\right) P\left(N(t)\frac{n'}{N(t)}, t\right). \tag{B.3}$$

This equation can in turn be written as

$$\bar{P}(x_n, t+1) = \sum_{n'=0}^{N(t)} \bar{P}(x_n|x_{n'})\bar{P}(x_{n'}, t). \tag{B.4}$$

Multiply this equation from the left with $1/N(t+1)$ and multiply and divide by $N(t)$ on the right hand side to give

$$N(t+1)\bar{P}(x_n, t+1) = \sum_{n'=0}^{N(t)} N(t+1)\bar{P}(x_n|x_{n'})N(t)\bar{P}(x_{n'}, t)\delta x, \tag{B.5}$$

where $\delta x = 1/N(t)$. Rewrite this as

$$f(x_n, t+1) = \sum_{n'=0}^{N(t)} f(x_n|x_{n'})f(x_{n'}, t)\delta x, \tag{B.6}$$

where $f(x_n, t+1) = N(t+1)P(x_n, t+1)$, $f(x_n|x'_n) = N(t+1)P(x_n|x_{n'}$ and $f(x_{n'}, t) = N(t)P(x_{n'}, t)$. Letting $N(t) \to \infty$ equation (B.6) can be approximated by

$$f(x, t+1) = \int_S f(x|x')f(x', t)\mathrm{d}x', \tag{B.7}$$

where $S \in [0,1]$. $n$ is now considered to be a continuous variable and $f(x,t)$ is taken to be a continuous probability distribution with the same moments as $P(x_n,t)$. Define the space $\mathcal{D}$ of functions $Q(x) \in C^\infty[0,1]$ with compact support on $S$ and boundary values

$$Q^{(k)}(0) = Q^{(k)}(1) = 0 \tag{B.8}$$

for non-negative integers $k$. $\mathcal{D}$ is as such a linear test space. Multiplying both sides of (B.7) by $Q(x)$ and integrating over $x$, one obtains

$$\int_S f(x,t+1)Q(x)\mathrm{d}x = \iint_S f(x|x')f(x',t)Q(x)\mathrm{d}x'\mathrm{d}x. \tag{B.9}$$

The left hand side of this equation is now associated with the distribution $f^{t+1}$:

$$(f^{t+1},Q) = \int_S f(x,t+1)Q(x)\mathrm{d}x \tag{B.10}$$

A Taylor expansion of $Q(x)$ about the value $x'$ gives

$$Q(x) = Q(x') + \frac{\partial Q(x')}{\partial x}\Delta x + \frac{1}{2}\frac{\partial^2 Q(x')}{\partial x^2}(\Delta x)^2 + O((\Delta x)^3),$$

where $\Delta x = x - x'$. $Q(x')$ and its derivatives are in $\mathcal{D}$ since $x' \in S$. Proceeding similarly from here on as was done for the case with $N(t)$ constant, one arrives at the equation

$$(f^{t+1} - f^t, Q) = (HN^t, Q) + \mathcal{O}(\epsilon^2), \tag{B.11}$$

where,

$$H(x,t) = \frac{\partial}{\partial x}\left[f(x,t)\epsilon M(x,t)\right] + \frac{\epsilon}{2N(t+1)}\frac{\partial^2}{\partial x^2}\left[f(x,t)D(x)\right]. \tag{B.12}$$

Thus, the following difference equation is valid in a distributional sense

$$f(x,t+1) - f(x,t) = H(x,t) + \mathcal{O}(\epsilon^2). \tag{B.13}$$

Define a new time variable $\tau$ as

$$\tau = \int_0^{t+1} \frac{\mathrm{d}s}{N(s)}. \tag{B.14}$$

Since this is a 1-1, strictly increasing function one can define $t = g^{-1}(\tau)$ from $\tau = g(t)$. Hence one can rewrite $f(x,t)$ as follows:

$$f(x,t) = f(x,g^{-1}(\tau)) = \hat{f}(x,\tau) \tag{B.15}$$

and

$$f(x,t+\Delta t) = \hat{f}(x,\tau+\Delta\tau). \tag{B.16}$$

The generation time step in the new variable is given by

$$\Delta\tau = \int_{t+1}^{t+2} \frac{\mathrm{d}s}{N(s)}. \tag{B.17}$$

Since $N(t)$ is constant in the interval $[t+1, t+2)$:

$$\Delta\tau = \frac{1}{N(t+1)} \tag{B.18}$$

Introducing $\tau$ to equation (B.13) and dividing by $\Delta\tau$ one obtains

$$\frac{f(x,\tau+\Delta\tau) - f(x,\tau)}{\Delta\tau} = \hat{H}(x,\tau) + \mathcal{O}(\epsilon^2), \tag{B.19}$$

where $\hat{H}(x,\tau) = N(t+1)H(x,\tau)$. Letting $N(t+1) \to \infty$ for all $t$, one finds, by allowing abuse of notation,

$$\frac{\partial f(x,t)}{\partial t} = H(x,t), \tag{B.20}$$

to $\mathcal{O}(\epsilon^2)$. $H(x,t)$ is given by

$$H(x,t) = \frac{\partial}{\partial x}\left[f(x,t)\epsilon M(x,t)\right] + \frac{1}{2}\frac{\partial^2}{\partial x^2}\left[f(x,t)D(x)\right]. \tag{B.21}$$

# C The diffusion approximation for $K$ alleles and $N(t)$ constant

## C.1 Obtaining and approximating the $\phi$ function for $K$ alleles

The $K$-allele probability function $\phi(\boldsymbol{x}, t)$ can be found by following the same procedure as for the 2-allele $\phi(x,t)$ in the main text.

First considering only the effect of mutations. Assume that each allele can mutate into any of the others. Let $U_{ik}$ be the probability of mutation from allele $i$ to allele $k$. Thus the probability that a child is sampled with allele $i$ is so far given by

$$P_i = x_i'(1 - \sum_{k\neq i} U_{ik}) + \sum_{k\neq i} x_k' U_{ki}. \tag{C.1}$$

Then introduce selection by defining $\phi_i(\boldsymbol{x}', t)$ as

$$\phi_i(\boldsymbol{x}, t) = \frac{\omega_i P_i}{Z}, \tag{C.2}$$

where $\omega_i = (1 - s_i)$ and $Z = \sum \omega_k P_k$. Introducing the model parameters as $\mathcal{O}(\epsilon)$ and proceeding as in (A.1), one finally obtains

$$\phi_i(\boldsymbol{x}', t) = x_i' + \epsilon\mathcal{M}(\boldsymbol{x}', t) + \mathcal{O}(\epsilon^2), \tag{C.3}$$

where

$$\mathcal{M}_i(\boldsymbol{x}', t) = A_i + x_i'\,(B + (1-X)s_K - s_i) + (1-X)U_{Ki} - x_i' U_{iK}. \quad \text{(C.4)}$$

In this equation

$$X = \sum_{k=1}^{K-1} x_k', \quad B = \sum_{k=1}^{K-1} s_k x_k', \quad \text{and} \quad A_i = \sum_{k\neq i}^{K-1} (x_k' U_{ki} - x_i' U_{ik}) \quad \text{(C.5)}$$

The terms with subscript $K$ in (C.4) appear from the implicit writing of $x_K'$ as $1 - X$.

## C.2 Moments of $\Delta\boldsymbol{x}|\boldsymbol{x'}$

The moments of $\Delta\boldsymbol{x}|\boldsymbol{x'}$ are obtained from the properties of the multinomial distribution (2.58) using the approximation (C.3). One thus finds

$$\mathrm{E}[\Delta x_i|\boldsymbol{x'}] = \epsilon\mathcal{M}_i + \mathcal{O}(\epsilon^2).$$

$$\begin{aligned}\mathrm{E}[\Delta x_i \Delta x_j|\boldsymbol{x'}] &= \mathrm{Cov}(\Delta x_i, \Delta x_j|\boldsymbol{x'}) + \mathrm{E}[\Delta x_i|\boldsymbol{x'}]\mathrm{E}[\Delta x_j|\boldsymbol{x'}] \\ &= \mathrm{Cov}(\Delta x_i, \Delta x_j|\boldsymbol{x'}) + \mathcal{O}(\epsilon^2).\end{aligned}$$

$$\mathrm{Cov}(\Delta x_i, \Delta x_j|\boldsymbol{x'}) = \begin{cases} \frac{-\epsilon x_i' x_j'}{N} + \mathcal{O}(\epsilon^2) & \text{if } i \neq j \\ \frac{\epsilon x_i'(1-x_i')}{N} + \mathcal{O}(\epsilon^2) & \text{if } i = j. \end{cases}$$

After a lengthy calculation along the same lines of (A.3) one finds

$$E[\Delta x_i \Delta x_j \Delta x_l|\boldsymbol{x'}] = \mathcal{O}(\epsilon^2) \quad \text{(C.6)}$$

for $i, j, l \in [1, 2 \ldots, K-1]$.

## C.3 Obtaining the Fokker-Planck equation

The marginal distribution for the number of alleles in generation $t+1$ is given by

$$P(\boldsymbol{n}, t+1) = \sum_{|\boldsymbol{n'}|+n_K=N} P(\boldsymbol{n}|\boldsymbol{n'})P(\boldsymbol{n'}, t), \quad \text{(C.7)}$$

where $|\boldsymbol{n'}| = n_1' + n_2' + \cdots + n_{K-1}$. One can argue as for the case with two alleles, and end up with

$$f(\boldsymbol{x}, t+1) = \int_\sigma f(\boldsymbol{x}|\boldsymbol{x'}) f(\boldsymbol{x'}, t)\mathrm{d}\boldsymbol{x'}. \quad \text{(C.8)}$$

The interval $\sigma$ is defined as $[0,1]^{K-1}$. In general define $\sigma_i = [0,1]^i$. Let $Q(\boldsymbol{x})$ be a scalar valued test function with compact support on $\sigma$ for which

$$\partial^\alpha Q(\mathbf{0}) = \partial^\alpha Q(\mathbf{1}) = 0, \quad \text{(C.9)}$$

for all non-negative integer values of $\alpha_i$ in $\alpha$, where $\alpha = (\alpha_1, \alpha_2, \ldots, \alpha_i, \ldots \alpha_{K-1})$. The multi-index notation is given by $\partial^\alpha = \partial_1^{\alpha_1}\partial_2^{\alpha_2}\ldots\partial_n^{\alpha_n}$ where $\partial_i^{\alpha_i} = \partial^{\alpha_i}/\partial x_i^{\alpha_i}$. $\mathbf{0}$ and $\mathbf{1}$ are the $K-1$ dimensional zero vector and a $K-1$ dimensional vectors of 1s. Multiplying both sides of equation (C.8) with the $K-1$-dimensional Taylor expansion of $Q(\boldsymbol{x})$ and and then integrating over $\sigma$, one finds

$$\begin{aligned}\int_\sigma f(\boldsymbol{x}, t+1)Q(\boldsymbol{x})\mathrm{d}\boldsymbol{x} &= \int_\sigma f(\boldsymbol{x}', t)\left[\int_\sigma f(\boldsymbol{x}|\boldsymbol{x}')\mathrm{d}\boldsymbol{x}\right]Q(\boldsymbol{x}')\mathrm{d}\boldsymbol{x}' \\ &+ \sum_{i=1}^{K-1}\int_\sigma f(\boldsymbol{x}', t)\left[\int_\sigma \Delta x_i f(\boldsymbol{x}'|\boldsymbol{x}')\mathrm{d}\boldsymbol{x}\right]\frac{\partial Q(\boldsymbol{x}')}{\partial x_i}\mathrm{d}\boldsymbol{x}' \\ &+ \frac{1}{2}\sum_{i,j=1}^{K-1}\int_\sigma f(\boldsymbol{x}', t)\left[\int_\sigma \Delta x_i\Delta x_j f(\boldsymbol{x}|\boldsymbol{x}')\mathrm{d}\boldsymbol{x}\right]\frac{\partial^2 Q(\boldsymbol{x}')}{\partial x_i x_j}\mathrm{d}\boldsymbol{x}' + \mathcal{O}(\epsilon^2),\end{aligned}$$

after changing the order of intergration. Left out terms correspond to $\mathcal{O}(\mathrm{E}[\Delta x_i \Delta x_j \Delta x_l])$ and the orders of higher moments, all of which are $\mathcal{O}(\epsilon^2)$. The notation $\partial x_i x_j$ is shorthand writing for $\partial x_i \partial x_j$. The left hand side of this equation is identified with the distribution $f^{t+1}$:

$$(f^{K-1}, Q) = \int_\sigma Q(\boldsymbol{x})f(\boldsymbol{x}, t+1)\mathrm{d}\boldsymbol{x}. \tag{C.10}$$

In the second and third term on the right hand side, the order of integration is changed once more to obtain

$$\begin{aligned}&= \int_\sigma f(\boldsymbol{x}', t)Q(\boldsymbol{x}')\mathrm{d}\boldsymbol{x}' \\ &\quad + \sum_{i=1}^{K-1}\int_\sigma f(\boldsymbol{x}', t)\left[\int_{\sigma_1}\Delta x_i\left(\int_{\sigma_{K-2}} f(\boldsymbol{x}|\boldsymbol{x}')\mathrm{d}\boldsymbol{x}\right)\mathrm{d}x_i\right]\frac{\partial Q(\boldsymbol{x}')}{\partial x_i}\mathrm{d}\boldsymbol{x}' \\ &+\frac{1}{2}\sum_{i,j=1}^{K-1}\int_\sigma f(\boldsymbol{x}', t)\left[\iint_{\sigma_1\sigma_1}\Delta x_i\Delta x_j\left(\int_{\sigma_{K-3}} f(\boldsymbol{x}|\boldsymbol{x}')\mathrm{d}\boldsymbol{x}\right)\mathrm{d}x_i\mathrm{d}x_j\right]\frac{\partial^2 Q(\boldsymbol{x}')}{\partial x_i x_j}\mathrm{d}\boldsymbol{x}' + \mathcal{O}(\epsilon^2).\end{aligned}$$

Recognizing that the innermost integrals are marginal distributions for $f(x_i|\boldsymbol{x}')$ and $f(x_i, x_j|\boldsymbol{x}')$ the previous equation becomes

$$\begin{aligned}\int_\sigma Q(\boldsymbol{x})f(\boldsymbol{x}, t+1)\mathrm{d}\boldsymbol{x} &= \int_\sigma f(\boldsymbol{x}', t)Q(\boldsymbol{x}')\mathrm{d}\boldsymbol{x}' \\ &+ \sum_{i=1}^{K-1}\int_\sigma f(\boldsymbol{x}', t)\left[\int_{\sigma_1}\Delta x_i' f(x_i|\boldsymbol{x}')\mathrm{d}x_i\right]\frac{\partial Q(\boldsymbol{x}')}{\partial x_i}\mathrm{d}\boldsymbol{x}' \\ &+\frac{1}{2}\sum_{i,j=1}^{K-1}\int_\sigma f(\boldsymbol{x}', t)\left[\iint_{\sigma_1\sigma_1}\Delta x_i\Delta x_j f(x_i, x_j|\boldsymbol{x}')\mathrm{d}x_i\mathrm{d}x_j\right]\frac{\partial^2 Q(\boldsymbol{x}')}{\partial x_i x_j}\mathrm{d}\boldsymbol{x}' + \mathcal{O}(\epsilon^2).\end{aligned} \tag{C.11}$$

The terms within the squared brackets correspond to $\mathrm{E}[\Delta x_i|\boldsymbol{x}']$ and $\mathrm{E}[\Delta x_i \Delta x_j|\boldsymbol{x}']$ respectively. Inserting these values gives

$$= \int_\sigma f(\boldsymbol{x}',t)Q(\boldsymbol{x}')\mathrm{d}\boldsymbol{x}' - \sum_{i=1}^{K-1}\int_\sigma f(\boldsymbol{x}',t)\epsilon\mathcal{M}_i\frac{\partial Q(\boldsymbol{x}')}{\partial x_i}\mathrm{d}\boldsymbol{x}' + \frac{\epsilon}{2N}\sum_{i,j=1}^{K-1}\int_\sigma f(\boldsymbol{x}',t)C(x_i',x_j')\frac{\partial^2 Q(\boldsymbol{x}')}{\partial x_i x_j}\mathrm{d}\boldsymbol{x}' + \mathcal{O}(\epsilon^2),$$

where

$$C(x_i',x_j') = \begin{cases} -x_i'x_j' & \text{if } i \neq j \\ x_i'(1-x_i') & \text{if } i = j \end{cases}$$

Interpreting the integrals as distributions, and interpreting the differentiated $Q$ functions as differentiation in the distributional sense, one ends up with

$$(f^{t+1},Q) = (f^t,Q) - \sum_{i=1}^{K-1}(\epsilon\frac{\partial f^t\mathcal{M}_i}{\partial x_i},Q) + \sum_{i,j=1}^{K-1}\frac{\epsilon}{2N}(\frac{\partial^2 f^t C_{i,j}}{\partial x_i x_k},Q) + \mathcal{O}(\epsilon^2). \tag{C.12}$$

The variable $\boldsymbol{x}'$ has changed name to $\boldsymbol{x}$. Thus, the following difference equation is satisfied in the distributional sense

$$f(\boldsymbol{x},t+1) - f(\boldsymbol{x},t) = -\sum_{i=1}^{K-1}\epsilon\frac{\partial}{\partial x_i}f(\boldsymbol{x},t)\mathcal{M}_i(\boldsymbol{x},t) + \frac{\epsilon}{2N}\sum_{i,j=1}^{K-1}\frac{\partial^2}{\partial x_i x_j}f(\boldsymbol{x},t)C(x_i,x_j) + \mathcal{O}(\epsilon^2). \tag{C.13}$$

This equation is divided by $\Delta\tau = 1/N$ obtained from the introduced time transformation

$$\tau = \frac{1}{N} \tag{C.14}$$

and the limit $N \to \infty$ is taken to give

$$\frac{\partial}{\partial t}f(\boldsymbol{x},t) = -\sum_{i=1}^{K-1}\frac{\partial}{\partial x_i}[f(\boldsymbol{x},t)\mathcal{M}_i(\boldsymbol{x},t)] + \frac{1}{2}\sum_{i,j=1}^{K-1}\frac{\partial^2}{\partial x_i x_j}[f(\boldsymbol{x},t)C(x_i,x_j)] \tag{C.15}$$

to $\mathcal{O}(\epsilon^2)$, where $\mathcal{M}_i$ has been scaled by multiplication with $N$.

# D The diffusion approximation for $K$ alleles and $N(t)$ variable

## D.1 The moments of $\Delta\boldsymbol{x}|\boldsymbol{x}'$

The moments of $\Delta\boldsymbol{x}|\boldsymbol{x}'$ are given by:

$$\mathrm{E}[\Delta x_i|\boldsymbol{x}'] = \epsilon\mathcal{M}_i + \mathcal{O}(\epsilon^2).$$

$$\begin{aligned}\mathrm{E}[\Delta x_i \Delta x_j|\boldsymbol{x'}] &= \mathrm{Cov}(\Delta x_i, \Delta x_j|\boldsymbol{x'}) + \mathrm{E}[\Delta x_i|\boldsymbol{x'}]\mathrm{E}[\Delta x_j|\boldsymbol{x'}] \\ &= \mathrm{Cov}(\Delta x_i, \Delta x_j|\boldsymbol{x'}) + \mathcal{O}(\epsilon^2).\end{aligned}$$

$$\mathrm{Cov}(\Delta x_i, \Delta x_j|\boldsymbol{x'}) = \begin{cases} \frac{-\epsilon x'_i x'_j}{N(t+1)} + \mathcal{O}(\epsilon^2) & \text{if } i \neq j \\ \frac{\epsilon x'_i(1-x'_i)}{N(t+1)} + \mathcal{O}(\epsilon^2) & \text{if } i = j. \end{cases}$$

$$E[\Delta x_i \Delta x_j \Delta x_l|\boldsymbol{x'}] = \mathcal{O}(\epsilon^2) \tag{D.1}$$

for $i, j, l \in [1, 2 \ldots, K-1]$.

## D.2 Obtaining the Fokker-Planck equation

Starting from

$$P(\boldsymbol{n}, t+1) = \sum_{|\boldsymbol{n'}|+n_K=N(t)} P(\boldsymbol{n}|\boldsymbol{n'})P(\boldsymbol{n'}, t). \tag{D.2}$$

and proceeding as for 2 alleles with variable $N(t)$, one finds

$$P(\boldsymbol{x}, t+1) = \int_\sigma P(\boldsymbol{x}|\boldsymbol{x'})P(\boldsymbol{x'}, t)\mathrm{d}\boldsymbol{x'}. \tag{D.3}$$

From this equation one can proceed as for the $K$-allele case when $N(t)$ is constant. One ends up with the following difference equation satisfied in the distributional sense:

$$f(\boldsymbol{x}, t+1) - f(\boldsymbol{x}, t) = -\sum_{i=1}^{K-1} \epsilon \frac{\partial}{\partial x_i} f(\boldsymbol{x}, t)\mathcal{M}_i(\boldsymbol{x}, t) + \frac{\epsilon}{2N(t+1)} \sum_{i,j=1}^{K-1} \frac{\partial^2}{\partial x_i x_j} f(\boldsymbol{x}, t)C(x_i, x_j) + \mathcal{O}(\epsilon^2), \tag{D.4}$$

where $\mathcal{M}_i(\boldsymbol{x}, t)$ and $C(x_i, x_j)$ are defined as for the case when $N(t)$ is constant. This equation is the divided by $\Delta\tau = 1/N(t+1)$ obtained from the transformation

$$\tau = \int_0^{t+1} \frac{1}{N(s)} \mathrm{d}s \tag{D.5}$$

and the limit $N \to \infty$ is taken to give

$$\frac{\partial}{\partial t} f(\boldsymbol{x}, t) = -\sum_{i=1}^{K-1} \frac{\partial}{\partial x_i}[f(\boldsymbol{x}, t)\mathcal{M}_i(\boldsymbol{x}, t)] + \frac{1}{2} \sum_{i,j=1}^{K-1} \frac{\partial^2}{\partial x_i x_j}[f(\boldsymbol{x}, t)C(x_i, x_j)] \tag{D.6}$$

to $\mathcal{O}(\epsilon^2)$, where $\mathcal{M}_i$ has been scaled by multiplication with $N(t+1)$.

# E The Fokker-Planck approximation of the Master equation from the Moran model

The Master equation is given as a function of the allele number while the Fokker-Planck equation is given in terms of the frequency. The variable change from between these two occurs as the probability functions and transition rates of the Master equation are expanded to $\mathcal{O}(N^{-3})$. First the Fokker-Planck equation for two alleles will be derived followed by a generaliztion to $K$ alleles. For simplicity of notation $N(t+1)$ and $\mathcal{M}(\boldsymbol{x},t)$, and in the $K$-allele case $P(\boldsymbol{x},t)$, are written as $N$ and $\mathcal{M}$, and $P$. Further, the notation

$$(n_1,\ldots,n_i-1,\ldots,n_j+1,\ldots,n_{K-1}) = (n_i-1;n_j+1) \tag{E.1}$$

is used to simplify the expressions for $T(\boldsymbol{n}|\boldsymbol{n}')$, $P(\boldsymbol{n}',t)$ and $\phi_i(\boldsymbol{x},t)$ in the $K$ allele derivation. As stated in the main text the correctness of this derivation to $\mathcal{O}(\epsilon^2)$ is implicit through the approximation of $\phi(\boldsymbol{x},t)$.

## E.1 The Fokker-Planck approximation for 2 alleles

The Master equation for two alleles is given by

$$\frac{\partial P(n,t)}{\partial t} = \sum_{n'\neq n} T(n|n')P(n',t) - \sum_{n'\neq n} T(n'|n)P(n,t). \tag{E.2}$$

In the first sum, $\sum_{n'\neq n} T(n|n')P(n',t)$, the approximated transition rates are

$$\begin{aligned}
T(n|n-1) &= (1-\frac{n-1}{N})\phi(\frac{n-1}{N},t) \\
&= (1-x+\frac{1}{N})(x-\frac{1}{N}+\frac{1}{N}M^-) \\
&= x(1-x)+\frac{1}{N}(2x+2M(1-x)-1)+\frac{1}{N^2}(2G(1-x)+2M-1)+\mathcal{O}(N^{-3}) \\
T(n|n+1) &= (\frac{n+1}{N})(1-\phi(\frac{n+1}{N},t)) \\
&= (x+\frac{1}{N})(1-x-\frac{1}{N}-\frac{1}{N}M^+) \\
&= x(1-x)+\frac{1}{N}(1-2x-2xM)+\frac{1}{N^2}(2Gx-2M-1)+\mathcal{O}(N^{-3}),
\end{aligned}$$

where

$$\begin{aligned}
M^- &= M+\frac{G}{N}+\mathcal{O}(N^{-2}), \\
M^+ &= M-\frac{G}{N}+\mathcal{O}(N^{-2}),
\end{aligned}$$

in which $G = S(2x-1) + (U+V)$. $M^-$ and $M^+$ were obtained from evaluating $M(x - \frac{1}{N}, t)$ and $M(x + \frac{1}{N}, t)$ respectively using the scaled parameters

$$U = \frac{N\mu}{2}, \quad V = \frac{N\nu}{2}, \quad S = \frac{Ns}{2}.$$

The Taylor expansions of $P(n', t)$ are given by

$$P(n-1,t) = P(x,t) - \frac{1}{N}P_x(x,t) + \frac{1}{2N^2}P_{xx}(x,t) + \mathcal{O}(N^{-3}),$$
$$P(n+1,t) = P(x,t) + \frac{1}{N}P_x(x,t) + \frac{1}{2N^2}P_{xx}(x,t) + \mathcal{O}(N^{-3}).$$

After some algebra the first sum in the Master equation thus becomes

$$\sum_{n' \neq n} T(n|n')P(n',t) = 2x(1-x) + \frac{2}{N}M(1-2x)P(x,t)$$
$$+\frac{1}{N^2}\left[2(G-2)P(x,t) + 2(1-2x-M)P_x(x,t) + x(1-x)P_{xx}(x,t)\right] + \mathcal{O}(N^{-3}).$$

From the transition rates

$$T(n-1|n) = x(1-\phi(\frac{n}{N}, t)) \tag{E.3}$$
$$T(n+1|n) = (1-x)\phi(\frac{n}{N}, t) \tag{E.4}$$

the second sum is more straightforward to find:

$$\sum_{n' \neq n} T(n'|n)P(n,t) = \left[2x(1-x) + \frac{2}{N}M(1-2x)\right]P(x,t).$$

Subtracting the last sum from the first, the Master equation is

$$\frac{\partial P(x,t)}{\partial t} = \frac{2}{N^2}\left[(G-2)P(x,t) + (1-2x-M)P_x(x,t) + \frac{1}{2}x(1-x)P_{xx}(x,t)\right] + \mathcal{O}(N^{-3}),$$

which can be simplified to

$$\frac{\partial f(x,t)}{\partial t} = \frac{2}{N^2}\left[-\frac{\partial}{\partial x}(f(x,t)M(x,t)) + \frac{1}{2}\frac{\partial^2}{\partial x^2}(D(x)f(x,t))\right] + \mathcal{O}(N^{-3}).$$

Introducing $\tau = 2t/N^2$ and letting $N \to \infty$ gives the same Fokker-Planck equation as for the Wright-Fisher model. The correctness of this equation is given to $\mathcal{O}(\epsilon^2)$.

## E.2 The Fokker-Planck approximation for $K$ alleles

The Master equation for $K$ alleles is given by:

$$\frac{\partial P(\boldsymbol{n},t)}{\partial t} = \sum_{\boldsymbol{n}'\neq\boldsymbol{n}} T(\boldsymbol{n}|\boldsymbol{n}')P(\boldsymbol{n}',t) - \sum_{\boldsymbol{n}'\neq\boldsymbol{n}} T(\boldsymbol{n}'|\boldsymbol{n})P(\boldsymbol{n},t) \tag{E.5}$$

Since it is quite comprehensive, the derivation of the $K$-allele model will be more detailed than for the 2-allele model.

There are six distinct transition rates that go into the Master equation, up to summation over $i$ and $j$. The following three transition rates

- $T(\boldsymbol{n}|n_i - 1)$
- $T(\boldsymbol{n}|n_i + 1)$
- $T(\boldsymbol{n}|n_i - 1; n_j + 1)$

go into the first sum, $\sum_{\boldsymbol{n}'\neq\boldsymbol{n}} T(\boldsymbol{n}|\boldsymbol{n}')P(\boldsymbol{n}',t)$. The last three

- $T(n_i - 1|\boldsymbol{n})$
- $T(n_i + 1|\boldsymbol{n})$
- $T(n_i - 1; n_j + 1|\boldsymbol{n})$

go into the last sum, $\sum_{\boldsymbol{n}'\neq\boldsymbol{n}} T(\boldsymbol{n}'|\boldsymbol{n})P(\boldsymbol{n},t)$. To clearly see the task ahead, here is the Master equation with all terms included:

$$\begin{aligned}
&= \sum_{i=1}^{K-1} \Bigg( T(\boldsymbol{n}|n_i - 1)P(n_i - 1,t) + T(\boldsymbol{n}|n_i + 1)P(n_i + 1,t) \\
&\qquad\qquad + \sum_{j\neq i} T(\boldsymbol{n}|n_i - 1; n_j + 1)P(n_i - 1; n_j + 1),t) \Bigg) \\
&- \sum_{i=1}^{K-1} \left( T(n_i - 1|\boldsymbol{n}) + T(n_i + 1|\boldsymbol{n}) + \sum_{j\neq i} T(n_i - 1; n_j + 1|\boldsymbol{n}) \right) P(\boldsymbol{n},t)
\end{aligned} \tag{E.6}$$

In the following section the terms that condition on $\boldsymbol{n}$ will be expanded to $\mathcal{O}(N^{-3})$.

### E.2.1 $T(n_i - 1|\boldsymbol{n})$, $T(n_i + 1|\boldsymbol{n})$ and $T(n_i - 1; n_j + 1|\boldsymbol{n})$

Since the number of alleles in the parent generation is given by $\boldsymbol{n}$ the transition rates in the last sum of the Master equation are easier to compute.

One finds

$$\sum_{\boldsymbol{n}'\neq\boldsymbol{n}} T(\boldsymbol{n}'|\boldsymbol{n})P(\boldsymbol{n},t) = \left( \sum_{i=1}^{K-1} x_i(1-\sum_{k=1}^{K-1}\phi_k(\boldsymbol{x},t))+(1-X)\phi_i(\boldsymbol{x},t))+\sum_{j\neq i} x_i\phi_j(\boldsymbol{x},t)) \right) P(\boldsymbol{x},t). \tag{E.7}$$

After inserting for $\phi$ and simplifying this becomes

$$= \left( \sum_{i=1}^{K-1} \left[ 2x_i(1-X) + \frac{2}{N}(\mathcal{M}_i(1-X) - x_i\mathcal{M}) + \sum_{j\neq i} x_i(x_j + \frac{2}{N}\mathcal{M}_j) \right] \right) P(\boldsymbol{x},t), \tag{E.8}$$

where

$$\mathcal{M} = \sum_{k=1}^{K-1} \mathcal{M}_i, \quad X = \sum_{i=1}^{K-1} x_i \tag{E.9}$$

### E.2.2 $T(\boldsymbol{n}|n_i - 1; n_j + 1)$

This is the rate by which an individual with allele $i$ is born and an individual with allele $j$ dies, given that in the parent population there were $n_i - 1$ and $n_j + 1$ individuals with allele $i$ and $j$ respectively. Thus the transition rate becomes

$$\sum_{i=1}^{K-1}\sum_{j\neq i} T(\boldsymbol{n}|n_i - 1; n_j + 1) = \sum_{i=1}^{K-1}\sum_{j\neq i}(x_j + \frac{1}{N})\phi_i(x_i - \frac{1}{N}; x_j + \frac{1}{N}, t). \tag{E.10}$$

The expression for $\phi_i(x_i - \frac{1}{N}; x_j + \frac{1}{N}, t)$ is found after some calculation to be

$$\phi_i(x_i - \frac{1}{N}; x_j + \frac{1}{N}, t) = \mathcal{M}_i + \frac{1}{N}\mathcal{G}_i + \mathcal{O}(N^{-2}), \tag{E.11}$$

where

$$\mathcal{G}_i = U_i + U_{ij} + U_{iK} + s_i(1 - x_i) + x_i s_j - (1 - X)s_K - B. \tag{E.12}$$

Hence (E.10) becomes

$$= \sum_{i=1}^{K-1}\sum_{j\neq i} x_i x_j + \frac{1}{N}(x_i - x_j + 2x_j\mathcal{M}_i) + \frac{1}{N^2}(2x_j G_i + 2\mathcal{M}_i - 1) + \mathcal{O}(N^{-3}). \tag{E.13}$$

The Taylor expansion of $P(n_i - 1; n_j + 1, t)$ about $P(\boldsymbol{n}, t)$, after changing to variable $\boldsymbol{x}$, is given by

$$P(n_i - 1; n_j + 1, t) = P + \frac{1}{N}(P_j - P_i) + \frac{1}{2N^2}(P_{jj} + P_{ii} - 2P_{ij}) + \mathcal{O}(N^{-3}). \tag{E.14}$$

Multiplying this equation with (E.13) one finds

$$= \sum_{i=1}^{K-1} \sum_{j \neq i} P x_i x_j + \frac{1}{N} \left[ P(x_i - x_j + 2x_j \mathcal{M}_i) + (P_j - P_i) x_j x_i \right]$$
$$+ \frac{1}{N^2} \left[ P(2x_j G_i + 2M_i - 1) + (P_j - P_i)(x_i - x_j + 2x_j \mathcal{M}_i) + \frac{1}{2}(P_{jj} + P_{ii} - 2P_{ij}) \right] + \mathcal{O}(N^{-3}). \tag{E.15}$$

After subtracting common terms found in (E.8) and observing that by symmetry

$$\sum_{i=1}^{K-1} \sum_{j \neq i} P(x_i - x_j) + (P_j - P_i) x_j x_i = 0 + 0 = 0 \tag{E.16}$$

one is left with

$$= \sum_{i=1}^{K-1} \sum_{j \neq i} \frac{1}{N^2} \Big[ P(2x_j G_i + 2M_i - 1) + (P_j - P_i)(x_i - x_j + 2x_j \mathcal{M}_i) + \frac{1}{2}(P_{jj} + P_{ii} - 2P_{ij}) \Big] + \mathcal{O}(N^{-3}). \tag{E.17}$$

The sum over $j$ must be evaluated since this equation will be added terms where the sum is only over $i$. The final result is

$$= \frac{2}{N^2} \sum_{i=1}^{K-1} P \Big[ \hat{\mathcal{G}}_i (X - x_i) + \sum_{j \neq i} x_j U_{ji} + x_i \sum_{j \neq i} x_j (s_j - S_K) - (X - x_i) U_{Ki} + (K-2)\mathcal{M}_i - \frac{K}{2} + 1 \Big] + P_i \Big[ X + x_i \mathcal{M} - (K-1) x_i - \mathcal{M} X \Big] + \frac{1}{2}(P_{ii} x_i (X - x_i) - \sum_{j \neq i} P_{ij} x_j x_i + \mathcal{O}(N^{-3}) \tag{E.18}$$

### E.2.3 $T(\boldsymbol{n}|n_i - 1)$ and $T(\boldsymbol{n}|n_i + 1)$

First consider $T(\boldsymbol{n}|n_i - 1)$, the rate by which allele $K$ dies and allele $i$ is born, given that in the parent population there were $n_i - 1$ individuals with allele $i$. The transition rate is given by

$$\sum_{i=1}^{K-1} T(\boldsymbol{n}|n_i - 1) = \sum_{i=1}^{K-1} (1 - x_i + \frac{1}{N}) \phi_i (x_i - \frac{1}{N}, t). \tag{E.19}$$

The expression for $\phi_i(x_i - \frac{1}{N}, t)$ is given by

$$\phi_i (x_i - \frac{1}{N}, t) = x_i - \frac{1}{N} + \frac{1}{N} \mathcal{M}_i + \frac{1}{N^2} \hat{\mathcal{G}}_i + O(N^{-3}) \tag{E.20}$$

where $\hat{\mathcal{G}}_i$ is given by

$$\hat{\mathcal{G}}_i = U_i + U_{ik} + U_{ki} + s_i(1-x_i) + x_i s_K - B - (1-X)s_K. \tag{E.21}$$

Hence the expression (E.19) is given by

$$\begin{aligned} &= \sum_{i=1}^{K-1} x_i(1-X) + \frac{1}{N}\left[x_i + (2\mathcal{M}_i - 1)(1-X)\right] \\ &\qquad + \frac{1}{N^2}\left[2\mathcal{M}_i - 1 + 2\hat{\mathcal{G}}_i(1-X)\right] + \mathcal{O}(N^{-3}), \end{aligned} \tag{E.22}$$

The Taylor expansion of $P(n_i - 1, t)$ is given by

$$P(n_i - 1, t) = P - \frac{1}{N}P_i + \frac{1}{2N^2}P_{ii} + \mathcal{O}(N^{-3}). \tag{E.23}$$

Multiplying the transition rate (E.22) with its probability of occurence and subtracting common terms found in (E.8) one finds

$$\begin{aligned} &= \sum_{i=1}^{K-1} \frac{1}{N}\left[P(x_i - (1-X)) - P_i x_i(1-X)\right] \\ &+\frac{1}{N^2}\left[P(2\mathcal{M}_i - 1 + 2\hat{\mathcal{G}}(1-X)) - P_i(x_i + (2\mathcal{M}_i - 1)(1-X)) + \frac{1}{2}P_{ii}x_i(1-X)\right] + \mathcal{O}(N^{-3}) \end{aligned} \tag{E.24}$$

Next consider $T(\boldsymbol{n}|n_i + 1)$, the rate at which allele $i$ dies and allele $K$ is born with $n_i + 1$ allele $i$ in the parent population. The transition rate is given by

$$\begin{aligned} \sum_{i=1}^{K-1} T(\boldsymbol{n}|n_i+1) &= \sum_{i=1}^{K-1}(x_i + \frac{1}{N})(1 - \sum_{k=1}^{K-1}\phi_k(x_i + \frac{1}{N}, t)) \\ &= \sum_{i=1}^{K-1}\left(x_i + \frac{1}{N}\right)\left(1 - X - \frac{1}{N}(1 + 2\mathcal{M}_i) + \frac{\hat{\mathcal{G}}_i}{2N^2}\right. \\ &\qquad \left. - \frac{2}{N}\sum_{k\neq i}\left(\mathcal{M}_k(x_i - \frac{1}{N}, t) + \frac{1}{N}(U_{ik} - U_{Kk} + x_k(s_i - s_K))\right)\right) + \mathcal{O}(N^{-3}). \end{aligned} \tag{E.25}$$

Inserting

$$\mathcal{M}_k(x_i - \frac{1}{N}, t) = \mathcal{M}_k + \frac{1}{N}(U_{ik} - U_{Kk} + x_k(s_i - s_K)), \quad k \neq i, \tag{E.26}$$

and multiplying with

$$P(n_i + 1, t) = P + \frac{1}{N}P_i + \frac{1}{2N^2}P_{ii} + \mathcal{O}(N^{-3}) \tag{E.27}$$

and then subtracting common terms found in (E.8) one finds

$$
\begin{aligned}
&= \sum_{i=1}^{K-1} \frac{1}{N}\big[P((1-X)-x_i) + P_i x_i(1-X)\big] \\
&\quad + \frac{1}{N^2}\Big[P(2x_i(\hat{\mathcal{G}}_i - U_i + \hat{U}_K + \sum_{k\neq i} x_k(s_i - s_K))) - (1+2\mathcal{M}_i) \\
&\quad + P_i\left((1-X) - x_i(1+2\mathcal{M})\right) + \frac{1}{2}P_{ii}x_i(1-X)\Big] + \mathcal{O}(N^{-3}), \quad \text{(E.28)}
\end{aligned}
$$

where $\hat{U}_K = U_K - U_{Ki}$.

### E.2.4 The Fokker-Planck equation

Having obtained all the expressions in (E.6), equations (E.18), (E.24) and (E.28) are added. After some calculation one arrives at

$$
\begin{aligned}
&= \frac{2}{N^2}\sum_{i=1}^{K-1} P\left(\hat{\mathcal{G}}_i - \frac{K}{2} + (K-1)\mathcal{M}_i - \mathcal{M} + x_i\hat{U}_K - (X-x_i)U_{Ki} - x_iU_i + \sum_{j\neq i} x_jU_{ji}\right) \\
&\qquad + P_i(1 - Kx_i + \mathcal{M}_i) + \frac{1}{2}P_{ii}x_i(1-x_i) - \frac{1}{2}\sum_{j\neq i} P_{ij}x_ix_j. \quad \text{(E.29)}
\end{aligned}
$$

Noting the following cancellations,

$$
\sum_{i=1}^{K-1}\big((K-1)\mathcal{M}_i - \mathcal{M}\big) = 0, \quad \sum_{i=1}^{K-1}\big(x_i\hat{U}_K - (X-x_i)U_{Ki}\big) = 0, \quad \sum_{i=1}^{K-1}\big(x_iU_i - \sum_{j\neq i} x_jU_{ji}\big) = 0, \quad \text{(E.30)}
$$

one is left with

$$
\begin{aligned}
\frac{\partial P(\boldsymbol{x},t)}{\partial t} &= \frac{2}{N^2}\sum_{i=1}^{K-1} P(\boldsymbol{x},t)\left(\hat{\mathcal{G}}_i - \frac{K}{2}\right) + \frac{\partial P(\boldsymbol{x},t)}{\partial x_i}(1 - Kx_i + \mathcal{M}_i) \\
&\quad + \frac{1}{2}\frac{\partial^2 P(\boldsymbol{x},t)}{\partial x_i^2}x_i(1-x_i) - \frac{1}{2}\sum_{j\neq i}\frac{\partial^2 P(\boldsymbol{x},t)}{\partial x_i x_j}x_ix_j. \quad \text{(E.31)}
\end{aligned}
$$

This simplifies to

$$
\frac{\partial f(\boldsymbol{x},t)}{\partial t} = \frac{2}{N^2}\left(-\sum_{i=1}^{K-1}\frac{\partial}{\partial x_i}\mathcal{M}_i(\boldsymbol{x},t)f(\boldsymbol{x},t) + \frac{1}{2}\sum_{i,j=1}^{K-1}\frac{\partial^2}{\partial x_i x_j}C(x_ix_j)f(\boldsymbol{x},t)\right), \quad \text{(E.32)}
$$

where $P(\boldsymbol{x},t)$ has been changed to $f(\boldsymbol{x},t)$. Introducing $\tau = 2t/N^2$ and letting $N \to \infty$ one obtains the same Fokker-Planck equation as for the $K$-allele Wright-Fisher model.

## References


[1] A. McKane and D. Waxman, "Singular solutions of the diffusion equation of population genetics," *Journal of Theoretical Biology*, vol. 247, no. 4, pp. 849–858, 2007.

[2] J. Crow and M. Kimura, "An introduction to population genetics theory. 1970."

[3] H. Risken, *The Fokker-Planck Equation: Methods of Solution and Applications*. Springer, 1996.

[4] W. Ewens, "Mathematical Population Genetics. Biomathematics, Vol 9," 1979.

[5] P. Moran, *The Statistical Processes of Evolutionary Theory*. Clarendon Press, 1962.

[6] R. Blythe and A. McKane, "Stochastic models of evolution in genetics, ecology and linguistics," *J. Stat. Mech*, vol. 7018, 2007.

[7] M. Kimura, "PROCESS LEADING TO QUASI-FIXATION OF GENES IN NATURAL POPULATIONS DUE TO RANDOM FLUCTUATION OF SELECTION INTENSITIES," *Genetics*, vol. 39, no. 3, pp. 280–295, 1954.

[8] M. Kimura, "ON THE PROBABILITY OF FIXATION OF MUTANT GENES IN A POPULATION," *Genetics*, vol. 47, no. 6, pp. 713–719, 1962.

[9] I. Rouzine, A. Rodrigo, and J. Coffin, "Transition between Stochastic Evolution and Deterministic Evolution in the Presence of Selection: General Theory and Application to Virology," *Microbiology and Molecular Biology Reviews*, vol. 65, no. 1, p. 151, 2001.

[10] M. Ptashne and A. Gann, *Genes & Signals*. CSHL Press, 2002.

[11] U. Alon, "Simplicity in biology," *Nature*, vol. 446, no. 7135, p. 497, 2007.

[12] U. Alon, *An Introduction to Systems Biology: Design Principles of Biological Circuits*. Chapman & Hall/CRC, 2007.

[13] M. Savageau, "Design of molecular control mechanisms and the demand for gene expression.," *Proc Natl Acad Sci US A*, vol. 74, no. 12, pp. 5647–5651, 1977.

[14] G. Shinar, E. Dekel, T. Tlusty, and U. Alon, "Rules for biological regulation based on error minimization," *Proceedings of the National Academy of Sciences*, vol. 103, no. 11, pp. 3999–4004, 2006.

[15] M. Nowak, F. Michor, and Y. Iwasa, "The linear process of somatic evolution," *Proceedings of the National Academy of Sciences*, vol. 100, no. 25, pp. 14966–14969, 2003.

[16] D. Dingli, A. Traulsen, and F. Michor, "(A)Symmetric Stem Cell Replication and Cancer," *PLoS Computational Biology*, vol. 3, no. 4, 2007.

[17] M. Kimura and T. Ohta, "Probability of Gene Fixation in an Expanding Finite Population," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 71, no. 9, pp. 3377–3379, 1974.

[18] P. Ball, "Synthetic biology: Starting from scratch," *Nature*, vol. 431, pp. 624–626, 2004.

[19] D. Gibson, G. Benders, C. Andrews-Pfannkoch, E. Denisova, H. Baden-Tillson, J. Zaveri, T. Stockwell, A. Brownley, D. Thomas, M. Algire, *et al.*, "Complete Chemical Synthesis, Assembly, and Cloning of a Mycoplasma genitalium Genome," *Science*, vol. 319, no. 5867, p. 1215, 2008.