Utrecht University

Master thesis

Stochastic modelling of evolutionary processes

Authors:
Stefan Korenberg
Studentnumber 3709302
Utrecht University

Supervisors:
Prof. dr. ir. J.E. Frank
Utrecht University
and
Prof. dr. ir. H.T.C. Stoof
Utrecht University

30 August 2016

Abstract

Evolutionary models are applied in various branches of science. The notions of births, deaths and mutations in a population of individuals can be interpreted in many different ways. We discuss some general modelling choices and universal properties of evolutionary systems. Most importantly, when certain individuals reproduce faster than others, evolution gives rise to a dynamical system that optimizes itself.

After introducing evolutionary models in general, we focus on two specific models. Firstly, we investigate the relation of the population size to convergence properties in a population of solutions to the optimization problem 1-in-3-SAT. Secondly, we consider birth-death processes of diffusing particles. We derive an equation for the density of particles that describes the average behaviour and compare predictions to simulations of the micro-model. The average description does not account for the observed clustering of the population, so we consider methods to describe and measure deviations from the mean behaviour.

Contents

1 Evolutionary Models **1**
1.1 Evolutionary models . . . 1
1.2 Concepts of Evolutionary Modelling . . . 2
1.3 Modelling choices . . . 3
1.4 Mutation . . . 5
1.5 Dynamics . . . 7
1.5.1 Clustering . . . 8
1.5.2 Fixation . . . 8
1.6 Outline . . . 10

2 Bit-strings in 1-in-3-SAT **11**
2.1 Moran process for bit-strings . . . 11
2.2 1-in-3-SAT . . . 12
2.3 Results . . . 13
2.4 Thermodynamics . . . 15
2.5 Simulated annealing . . . 16

3 Continuum modelling of birth-death processes **17**
3.1 Micro-model description . . . 17
3.2 Fitness landscapes . . . 19
3.2.1 Peaks . . . 19
3.2.2 Double well . . . 21
3.2.3 Hydrogen potential . . . 23
3.3 Continuous description . . . 24
3.3.1 Derivation of the partial differential equation . . . 24
3.4 Solutions to the equation . . . 26
3.5 Solution for the hydrogen potential . . . 28
3.6 Noise on the PDE . . . 30

4 Comparison to the micro-model **33**
4.1 Translation between discrete populations and continuous densities . . . 33
4.1.1 Bin counting . . . 34
4.1.2 Kernel density estimation . . . 35
4.1.3 Empirical distribution function . . . 35
4.2 Comparison of the micro-model and the PDE . . . 36
4.2.1 Equilibrium . . . 36
4.2.2 Non-equilibrium . . . 36
4.3 Fluctuations . . . 38
4.3.1 Covariance . . . 40
4.3.2 Equilibrium fluctuations . . . 43
4.3.3 Non-equilibrium fluctuations . . . 45
4.4 Conclusion and discussion . . . 46

A Derivations **49**
A.1 Derivation of fixation probability 49
A.2 Poisson process 50
A.3 Clustering 51
A.4 Deterministic extinction 51

B Code **53**
B.1 Bit-strings code 53
B.2 Diffusion code 54

Chapter 1

Evolutionary Models

1.1 Evolutionary models

The ability of life to adapt emerges from a mechanism that is very simple but elegant. Individuals that flourish in their environment have the opportunity to reproduce, while individuals that struggle die without progeny. The key element is that children resemble their parents, so that the beneficial characteristics of the flourishing individuals are preserved for the next generation, while the detrimental characteristics of the struggling individuals are discarded from the population. Furthermore, parents and their offspring are not exactly equal. Instead, they differ slightly, which introduces new variability and allows for the emergence of new traits. These traits may enable some of the offspring to out-compete the others after which the story starts anew.

As humans are part of life, there has been much interest in gaining more insight into the ability of life to adapt. However, the ideas of evolution and birth-death processes have a much broader range of applicability and evolutionary models are used in several branches of science. Obviously, the models are used in biology and genetics. For example, to study the evolution of a species or the adaptation of bacteria [Nowak, 2006]. Somewhat less obvious is the application in fields like epidemics or economics. In these contexts, a birth may be an infection or a new company, while a death corresponds to a recovery of an infected person or a closure of some company. [Zhou, Liu, Bai, Chen, and Wang, 2006] [Boschma and Frenken, 2003].

In physics and chemistry, reaction-diffusion models are used that remind of birth-death processes [Wang, 2013]. Births and deaths are observed in the reactions where particles are duplicated or removed, while diffusion can be considered as continuous mutation. These processes can model for example auto-catalysis where the product is needed as reactant: $A + B \leftharpoons 2B$. In this case, the concentration of A particles determines the "reproduction rate" of B particles. To make the correspondence complete, the reaction from B to A should be spontaneous, so that this is equivalent to a spontaneous death of a B particle. Note that B could represent simple molecules, but also viruses or other complex structures. In more biologically oriented literature, this model is sometimes called the Brownian bug model [Young, Roberts, and Stuhne, 2001].

In evolutionary models, significant improvement of the population may be only seen after many generations. However, in the age of computers it is in fact feasible to generate many generations. Heuristic optimization methods that are based on evolutionary principles are known as genetic or evolutionary algorithms [Back, 1996]. Good solutions to the problem at hand are given a higher probability of reproduction than bad solutions. The solutions are mutated and recombined in order to search the solution space more efficiently than a random search. One hopes that the population converges towards some near-optimal solution of the problem, which is satisfactory in most practical cases.

Also in the field of data assimilation, methods are used that resemble evolutionary models. It is common to propagate a population (an ensemble) of solutions, for example in the Ensemble Kalman Filter [Burgers, van Leeuwen, and Evensen, 1998] and in the Particle Filter [Van Leeuwen, 2009]. However, for the Particle Filter there is a closer analogue. The technique of Sequential Importance Resampling [Gordon, Salmond, and Smith, 1993], which amounts to duplicating some of the solutions and deleting others, reminds of births and deaths. The "reproduction rate" of a solution is determined by the distance of the solution to the observed data. Mutations are introduced by model propagation. For a non-deterministic model, solutions that are equal at one instance of time are in general different at some other time, so it might be interesting to consider just these differences as mutations.

In the following, we introduce the general concepts in evolutionary systems with a finite population.

1.2 Concepts of Evolutionary Modelling

The basic ingredients of evolutionary modelling are birth, death and mutation. We consider a finite population of size M (a finite set G with $|G| = M$), which consists of members that are objects x_i[1] from some space S, so $G = \{x_i\}_{i \leq M}$ with $x_i \in S$. This population changes over time due to reproduction, death or mutation of its members. These events are random, so the population $G(t)$ is a time-dependent random variable or a stochastic process. In fact, it is a Markov process [Gardiner et al., 1985], because the current state of a population completely determines its future probabilities.
When a member reproduces it gets duplicated in the population and when a member dies it is removed from the population. Mutation means that a population member x_i is changed into some other $x' \in S$. Usually, mutation probabilities $\mathrm{P}(x_i \to x')$ are defined in such a way that a typical mutation changes the member x_i only slightly. That is, typically x' should be close to x_i in some sense. For example, S could be a metric space in which the only allowed mutations are mutations that are small with respect to this metric. The concept of mutation is illustrated in Figure 1.1 by three examples.
In the following, we will use the words "population members", "individuals" and "particles" interchangeably.

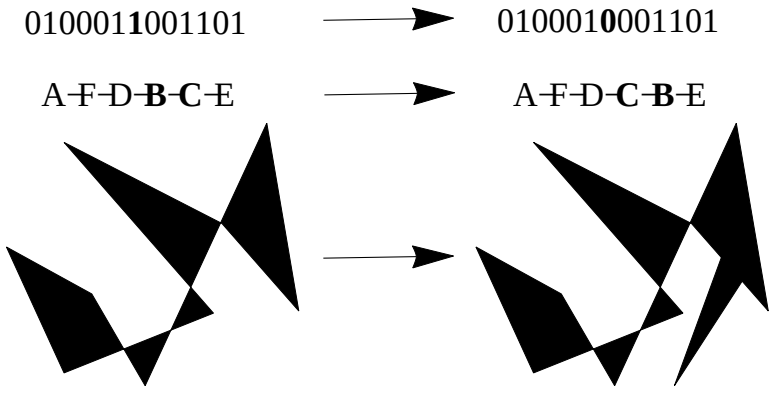

Figure 1.1 – Three examples of small mutations. Flipping one bit in a bit-string, swapping characters in a character string or adding a simple polygon to a polygon shape.

[1] Here, i is an index that labels the members, $1 \leq i \leq M$ where M is the population size.

1.3 Modelling choices

The population changes by births, deaths and mutations, so it is necessary to specify when these events happen and which members are changed. What is appropriate depends on the system we want to model. Here we discuss some of the choices that can be made.

Birth and death events can happen at fixed instances of time, but can also be distributed in time like a Poisson process with certain average rates [Gardiner et al., 1985]. Births and deaths can be independent events, but can also be coupled such that births and deaths happen simultaneously in order to ensure a fixed population size through time. A fixed population size could be due to limited resources, such as nutrients or space. Consider for example a rainforest, where every free spot in the light is immediately filled by new plants. If births and deaths are independent and the probability of death is strictly positive, there is a finite probability that the population becomes extinct. Determining this extinction probability is a central question in the theory of branching processes, which are Markov chains that consider the population size of reproducing particles [Karlin, 2014].

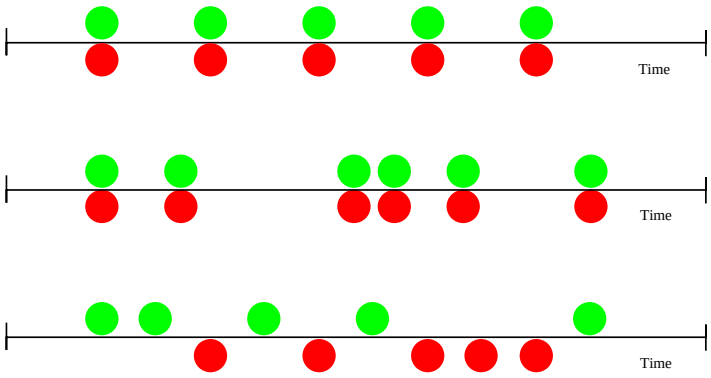

Figure 1.2 – Illustration of the distribution of births (green) and deaths (red) through time by some examples. From top to bottom: simultaneously at fixed instances of time, simultaneously at random instances of time or independent at fixed instances of time.

Now we consider two popular examples where births and deaths are coupled. In the so-called Moran process [Nowak, 2006], at each time step, one of the members x_i is chosen to reproduce and subsequently a random individual is chosen to die. The probability of death is uniform throughout the lifetime of an individual, so this process is appropriate for systems in which there is no significant ageing of its individuals. In the Moran process it is allowed that the reproducing individual is also selected for death.
As a second example, consider a population that reproduces in non-overlapping generations. The population at the next time step consists completely out of newborns and all individuals of the previous population are removed. That is, in a single time step, all individuals are replaced by their offspring. This models for example annual plants which are replaced by their seeds every year.

Reproduction rates need not be the same for every x_i in the population. The fitness of an individual is defined to be the ability of this individual to reproduce in order to propagate its genes. In fact, an individual that is good in every possible sense but is unable to reproduce, does not have a high fitness.
For every $x \in S$, define $f(x)$ as the average reproduction rate of an individual that has type x[2]. So the average reproduction rate of a population member x_i is equal to $f(x_i)$. The function f is called the fitness function and it defines the fitness landscape[3] on S.

[2]Note that we define f not only for population members x_i, but for all potential member types $x \in S$.
[3]Metaphorically, a population can move through the "fitness landscape".

If the fitness function is different for different x_i in the population, this gives rise to natural selection as described in the introduction. If the fitness function is constant, so that every type x reproduces equally fast, there might still be an interesting balance between mutations and so-called genetic drift. Genetic drift concerns the statistical properties of survival and extinction of types $x \in S$ due to random sampling of births and deaths in a finite population. Evolution in a constant fitness landscape is called neutral evolution and it is an interesting question whether neutral evolution or natural selection is more important for the evolution of species.

A similar, but slightly different definition of fitness concerns the reproduction probability instead of the reproduction rate. This definition is used when the actual reproduction times are not coupled to the fitness of the population, for example when reproduction happens at fixed instances of time. In this definition, the fitness of an individual is a weight $w_i = f(x_i)$ and the probability of being selected for reproduction is given by $\frac{w_i}{\sum w_i}$, where the sum runs over the whole population. Proper selection with these probabilities is done by roulette wheel selection, which is illustrated in Figure 1.3.
The new definition of fitness differs from the previous in that the actual reproduction rate depends on the fitness of the other individuals in the population. If time is not of interest, in most cases it is possible to rescale time in a way that the two definitions are practically equivalent.

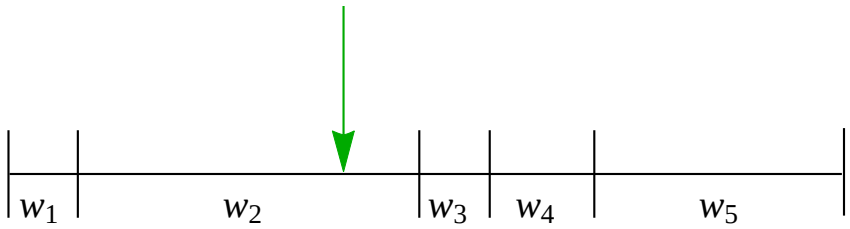

Figure 1.3 – Roulette wheel selection. The green arrow picks the reproducing individual. Its position is uniformly random on $(0, \sum_i w_i)$, so it has probability $\frac{w_i}{\sum_i w_i}$ to fall in the i-th interval, which is the correct selection probability.

It is possible and interesting to work with a fitness landscape that depends on the composition of the population. For example, Nowak investigated how cooperation could emerge by looking at agents that play the prisoners dilemma against each other [Nowak, 2006]. An individual x is defined by its strategy and the reproduction rate depends on the pay-off in the games, which depends on the encountered opponents. Although these landscapes are interesting, we will not consider them here and only use models in which there is no direct influence of the particles on the environment and on each other. The fitness landscape is fixed throughout time. Also, note that we only consider asexual reproduction. Sexual reproduction, in which two individuals, x_1 and x_2, are needed for reproduction and where the x_1 and x_2 are combined in some way, will not be treated.

Up to now, we only discussed births and deaths. For mutations, the following two choices are the most obvious and the most popular. Either mutations happen directly at birth or they happen throughout the whole lifetime of an individual. Mutations at birth correspond to the fact that an individual is slightly different from its parent. Throughout its lifetime, the individual remains the same. This is probably appropriate for most lifeforms. For other systems, it might be more appropriate to apply mutations during the entire lifetime of an individual. For example, a virus mutates between infections and not just at infections.

Often, mutations are implemented as mutation attempts. That is, in the case of mutation at birth, a newborn mutates with some probability p_m and is exactly equal to its parent with probability $1 - p_m$.

1.4 Mutation

The implementation of mutations depends on the space S in which the members live. We will now briefly discuss some examples of mutation.

As a first example, consider the case where S is the discrete space of bit-strings, so that $x \in S$ looks like $x = \{a_1 a_2 a_3 ... a_N\}$ with $a_i \in \{0, 1\}$ for $1 \leq i \leq N$. Note that practically anything can be represented by bit-strings[4] and that this space is in general high-dimensional. A mutation in this space could be a single bit flip: one bit i is selected at random and a_i is flipped ($0 \leftrightarrow 1$). This is an instance of a discrete space in which all mutations in the allowed subset are equally likely, which is a typical choice. Another possibility of mutation in S is to flip every bit with a certain probability $p_f \ll 1$. Now every x' is in principle reachable from every x in a single mutation, but the probability of such a mutation becomes very small when x and x' differ in many bits.

As a second example, we look at the case where S is the Euclidean space $\mathbb{R}^n$. In this space, mutation could be implemented as Brownian motion (or diffusion). This is an instance of continuous mutation. Instead of Brownian motion, we could also use more general Lévy processes [Gardiner et al., 1985]. Lévy processes are stochastic processes with independent and stationary increments. That is, the displacements in two non-overlapping time intervals are independent and identically distributed. Brownian motion, possibly with drift, is the only continuous Lévy process. Other Lévy processes contain discontinuous jumps, also called Lévy jumps or flights. These Lévy jumps may account for sudden larger mutations. In fact, it has been investigated if more general Lévy processes better fit the data on body mass and skull size of primates [Landis, Schraiber, and Liang, 2013].

A Lévy process is characterized by its characteristic function, which is given by the Lévy-Khintchine formula [Landis et al., 2013]. From this formula, it can be seen that a general Lévy process consists of linear displacement, Brownian motion and a jump process. The jumps arrive independently in time like a Poisson process. A Lévy process can be described by specifying these three components.

Figure 1.4 – Two-dimensional random walks based on Brownian motion (left) and the Cauchy process (right). The length scales can be set by the parameters of the distributions, so they are arbitrary and therefore not shown in the figures. The discontinuous jumps of the Cauchy process are clearly not present in the continuous Brownian motion. Note that the random walks consist of the same number of steps.

[4] Every object stored in a computer is stored in bits, so it is actually a bit-string.

To implement Lévy processes for mutation, we still need to choose a specific Lévy process. Probably the easiest choice besides Brownian motion is the Cauchy process, because it can be given analytically. The Cauchy process is a Lévy process where the magnitude of the displacement r in a time-interval Δt is given by the Cauchy distribution with parameter $\gamma\Delta t$. The probability density function of the Cauchy distribution with parameter $\gamma\Delta t$ is given by

$$p(r) = \frac{1}{\gamma\Delta t\pi\left[1 + \left(\frac{r}{\gamma\Delta t}\right)^2\right]}. \tag{1.4.1}$$

In Figure 1.4, a typical trajectory of Brownian motion is compared to a typical trajectory of the Cauchy process. It is not directly clear how to legitimately compare mutation based on Brownian motion and on the Cauchy process, because the Cauchy process has infinite variance. Since we do not have a specific application in mind, we will focus on Brownian motion in the rest of this thesis. However, it might be interesting, for example, to investigate if evolution with Lévy processes explores a rugged fitness landscape faster. Evolution in $\mathbb{R}^n$ could model continuous traits, such as propagation speed, energy consumption or the aforementioned body mass and skull size. It could also be a simple optimization method in a rough and high-dimensional parameter space with many local minima.

The third example shows that evolutionary algorithms can be very general. Consider the case where S is the space of all routes through N cities c with the restriction that we can only visit every city once. Two routes are shown in Figure 1.5. For $x \in S$, we write $x = \{c_1, c_2, \ldots c_N\}$, $c_i \neq c_j$ when $i \neq j$, which represents the order in which we visit the cities. The goal of the Travelling Salesman Problem is to find the shortest route in S. If we try to solve this problem using genetic algorithms, we could assign fitnesses based on the length of the route and implement mutation by for example cutting and reconnecting a route.

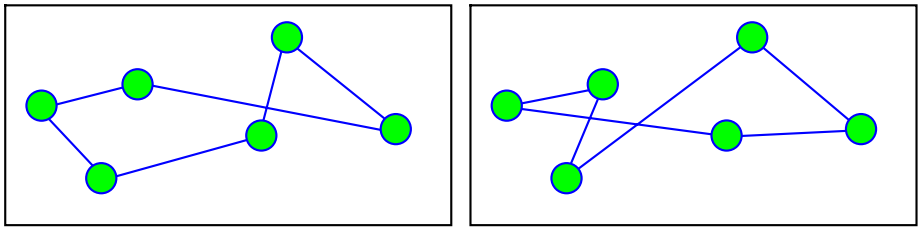

Figure 1.5 – Two possible Travelling Salesman routes through all the green cities.

1.5 Dynamics

Now that we have defined our evolutionary framework and have given some examples, we will look at some general results.

To start with, we need some way to initialize our system. For example, the initial population can be distributed randomly, according to some probability distribution. Depending on our modelling goals, we are interested in equilibrium or non-equilibrium results. If we are interested in equilibrium results, we could initialize our system randomly and wait for some time until the system has reached an equilibrium. To ensure that we are in equilibrium, we can check that certain population statistics have reached a steady value apart from some fluctuations. Also, it is interesting to look at the autocorrelation of the system, which is some measure of correlation between the system and itself at some other time. For some statistic X, the autocorrelation is defined by

$$C(t,t') = \mathrm{E}\Big[\Big(X(t)\text{–}\mathrm{E}(X)\Big)\Big(X(t')\text{–}\mathrm{E}(X)\Big)\Big], \tag{1.5.1}$$

where $\mathrm{E}(X)$ denotes the expectation of X. In equilibrium, the autocorrelation is a function of the difference in time $C(t,t') = C(\Delta t)$.

If we are interested in non-equilibrium behaviour, such as fitness increase, results depend more heavily on the initial condition of the population. Often, we are interested in some average behaviour and to study this, we need to average over an ensemble of systems that are initialized the same. In the case of equilibrium results, we can just study a single system for a very long time. This holds provided that the correlation time[5] is relatively short, which is not always the case in evolutionary systems. Sometimes it is not even clear what equilibrium means in the context of evolutionary systems.

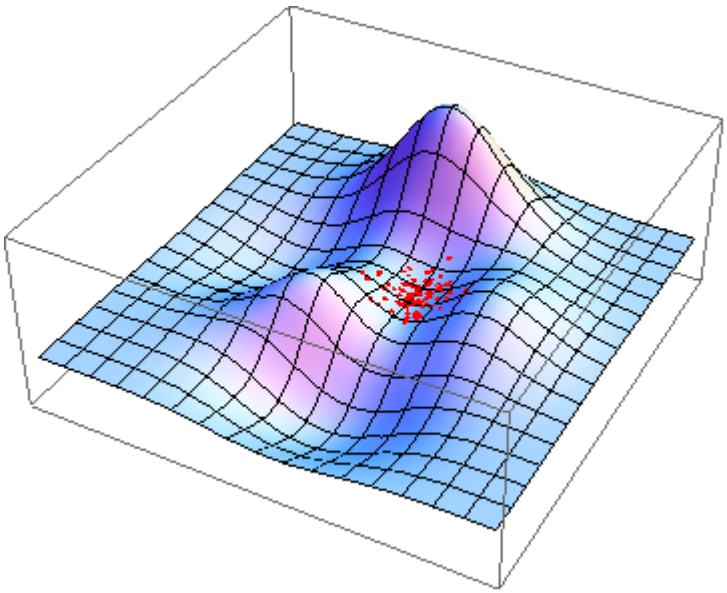

Figure 1.6 – A population $\{x_i\}_{i\leq 100}$ with $x_i \in S = \mathbb{R}^2$, represented by red dots, is clustering in $\mathbb{R}^2$. The surface represents some fitness landscape.

[5]The time after which the system is approximately uncorrelated to its previous state.

1.5.1 Clustering

A consequence of using small mutations in a bounded population is that the population clusters in the space S, which is illustrated in Figure 1.6. Consider two individuals x_1 and x_2 that are descendants from some common ancestor x_c. If the space is large compared to mutations and the time to the last common ancestor x_c is short, the descendants x_1 and x_2 did not have the opportunity to differentiate much from x_c. After a while the entire population can be traced back as descendants of some common ancestor x_c. As a result, the whole population consists of somewhat similar members.
This argument holds if the most recent common ancestor x_c lived not too far in the past. To become a common ancestor of the population, all lineages of the contemporaries of x_c need to become extinct. The larger the population size M, the longer it takes before the other $M-1$ lineages become extinct. Therefore, in a larger population, there is more time to differentiate by mutation and to spread out in the space S. On the contrary, in a smaller population, we should expect less variety in the population, because there is less time to differentiate.
We see that clustering is a finite size effect, which is related to extinction. The previous arguments also hold for sub-populations and this shows that tracing back lineages gives significant information about the population.

Another way to explain clustering is that births only occur in populated subsets of S, while death happens everywhere in S. That is, death can cause extinction in a subset $S_s \in S$, after which there are no more births in S_s[6]. This asymmetry between birth and deaths causes clustering of the population [Meyer, Havlin, and Bunde, 1996]. The shape of the population is dictated by a balance between clustering and spread due to mutations.

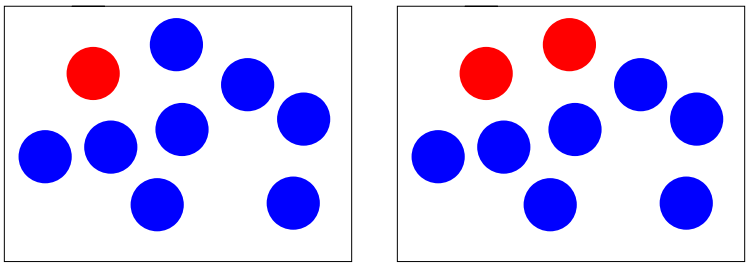

Figure 1.7 – A population of red and blue balls. In the second figure, the red ball has reproduced and its offspring has taken the place of a blue ball.

1.5.2 Fixation

To gain some understanding of natural selection and genetic drift in a population of fixed size, we perform a thought experiment in a population of M balls that can be either red or blue [Nowak, 2006]. The situation is illustrated by Figure 1.7. We use the Moran process described earlier and use fitness in the sense of selection weights. At first, there is only 1 red ball which has fitness f_R and $M-1$ blue balls that have fitness f_B. Denote by i the number of red balls. The probability that a red ball is selected for reproduction is

$$\mathrm{P}(\text{selection}) = \frac{if_R}{if_R + (M-i)f_B}, \tag{1.5.2}$$

while the probability that a red ball is selected for death is $\frac{i}{M}$. The absorbing states are given by $i = 0$ or $i = M$. We are now interested in the probability that the red balls eventually take over the entire population, so that $i = M$ after some time. In population genetics, this would be called fixation of the red balls.

[6] Until new particles are introduced in S_s by mutations from outside S_s

In Appendix A.1 it is shown that the probability of fixation of i red balls is given by

$$\mathrm{P(fixation)} = \frac{1 - r^i}{1 - r^M}, \tag{1.5.3}$$

with r the relative fitness, $r = \frac{f_B}{f_R}$. This formula is valid for both $0 \leq r < 1$ and $r > 1$. If $r = 1$, the probability of fixation is just $\lim_{r \to 1} \frac{1-r^i}{1-r^M} = \frac{i}{M}$, because in that case, the lineages of all M balls are equally probable to survive. The behaviour of this formula is shown in Figure 1.8 and Figure 1.9.

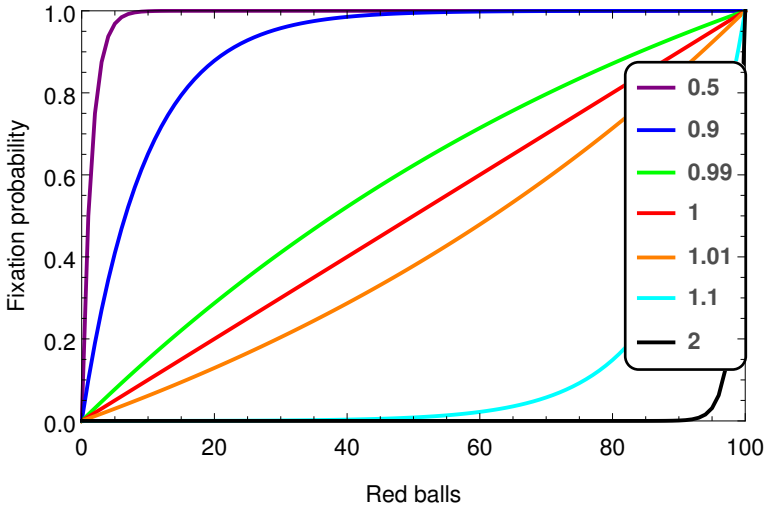

Figure 1.8 – The fixation probability in a population of size $M = 100$ as a function of initial red balls for different values of r.

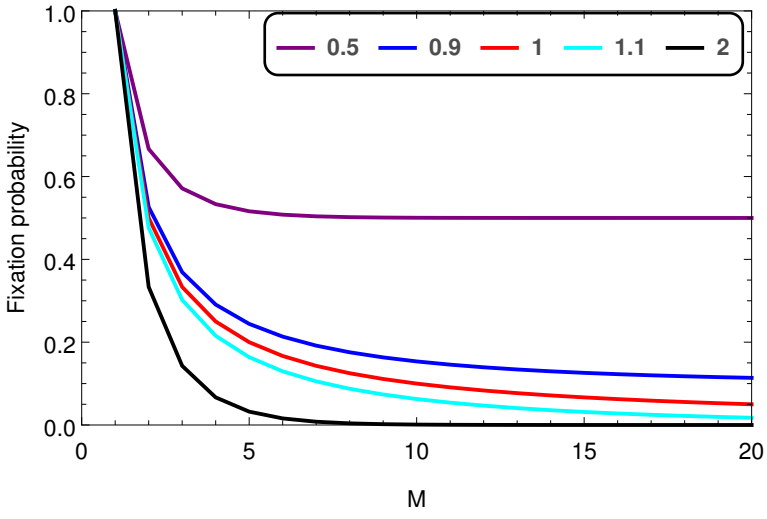

Figure 1.9 – The fixation probability of a single red ball as a function of the population size M for different values of r. The fixation probability does not decrease to zero for $r < 1$, while it does for $r \geq 1$.

As expected, the probability of fixation of the red balls is a decreasing function of the relative fitness r of the blue balls. Also, given that there is at least one red ball, $i > 0$, increasing f_R or decreasing f_B indefinitely so that $r \to 0$, results in fixation with probability $\lim_{r\to 0} \frac{1-r^i}{1-r^M} = 1$. The other way around results in fixation with probability $\lim_{r\to\infty} \frac{1-r^i}{1-r^M} = 0$, provided that $i < M$.

Other expected properties are that the fixation probability is an increasing function of i and a decreasing function of M, when keeping the other parameters constant. All these properties can be seen directly from Figure 1.8 and Figure 1.9. Figure 1.9 shows that, whenever $r < 1$, the fixation probability of a single red ball does not approach zero when the population size goes to infinity. In very large population, only beneficial mutations have a significant fixation probability. Taking the limit of infinite population size shows that fixation of lesser mutants is a finite size effect[7].

We now take a closer look at the behaviour of the fixation probability when we increase the system size by a factor $\alpha > 1$, such that $i \to \alpha i$ and $M \to \alpha M$. The fixation probability in the scaled system is given by $\frac{1-r^{\alpha i}}{1-r^{\alpha M}} = \frac{1-\tilde{r}^i}{1-\tilde{r}^M}$, with $\tilde{r} = r^\alpha$. If $r < 1$, $\tilde{r} < r$ and if $r > 1$, $\tilde{r} > r$, which shows that the scaled fixation probability is an increasing function of α for $r < 1$ and a decreasing function of α for $r > 1$[8]. This implies that scaling the system favours the fitter individuals. The reason for this is that it is less probable that a fluctuation accidentally makes the fitter individuals become extinct. Clearly, scaling the system changes its dynamics. This insight shows that it is difficult to take a limit to infinite population size for analytical purposes. Many interesting effects in evolutionary systems are finite size effects.

1.6 Outline

In the next chapter, we look at a specific evolutionary system and look at its convergence properties. Also, we describe that it is possible to make a link with statistical physics. In Chapter 3, we look at a continuous evolutionary system which we try to describe by a stochastic partial differential equation. In Chapter 4, we compare the analytical results from Chapter 3 to simulations of the micro-model.

[7] We will later see that finite population size corresponds to non-zero temperature when comparing evolutionary systems with physical systems.

[8] Because the normal fixation probability is a decreasing function of r

Chapter 2

Bit-strings in 1-in-3-SAT

2.1 Moran process for bit-strings

In order to make our discussion more concrete, we will focus on a specific model in this section. Our population consists of M bit-strings and we use the Moran process. For clarity, we will repeat their definitions. Every member x_i from the population

$$G = \{x_i\}_{1 \leq i \leq M} \tag{2.1.1}$$

is a bit-string of the form

$$x_i = \{a_1, a_2, a_3, ..., a_N\} \tag{2.1.2}$$

with $a_j \in \{0, 1\}$ for $1 \leq j \leq N$. We use bit-strings of length $N = 200$ and our population sizes range from $M = 100$ to $M = 5000$. The initial population is chosen completely random, so every bit of every bit-string is either 0 or 1 with equal probability. In the Moran process, at every time step a bit-string is selected for reproduction and simultaneously a bit-string is selected for death. The bit-strings are selected for reproduction according to their fitness, which we will define below, and are selected for death with uniform probability $\frac{1}{M}$. Mutation occurs only at birth: every bit of the newborn bit-string is flipped[1] with probability

$$\mu_m = \frac{1}{10N}, \tag{2.1.3}$$

so the number of bit flips is a binomial random variable with expected value of only

$$\text{E(number of bit flips)} = N\mu_m = \frac{1}{10}. \tag{2.1.4}$$

Most newborns are therefore exactly equal to their parent. In fact, a newborn equals its parent with probability

$$\text{P(zero bit flips)} = (1 - \mu_m)^N = \left(1 - \frac{1}{10N}\right)^N \approx \frac{9}{10}. \tag{2.1.5}$$

Mutation probabilities should be small enough so that good individuals are preserved in the population instead of mutated randomly. In genetic algorithms, good individuals can be preserved by simply excluding them from deaths, but we will not consider that case here.

[1] That is, $0 \to 1$ or $1 \to 0$.

2.2 1-in-3-SAT

We continue by defining the fitness landscape[2]. Every population member x_i is assigned a fitness value $f(x_i) > 0$ by a fitness function f. The probability that x_i is selected for reproduction is given by $\frac{f(x_i)}{\sum_j f(x_j)}$, where the sum runs over the population members.

The fitness function f is based on the boolean satisfiability problem 1-in-3-SAT, which is similar to the more familiar 3-SAT [Motoki and Uehara, 2000]. The problem 3-SAT is well-studied and was shown to be NP-complete. This means that any other problem in NP[3] can be reduced to it in polynomial time. 1-in-3-SAT is also NP-complete [Schaefer, 1978], so if the well-known conjecture P $\neq$ NP holds, there is no efficient algorithm to find a solution to 1-in-3-SAT, because this would imply P = NP.

We will now proceed by explaining 1-in-3-SAT. We are given so-called clauses A_J, with $1 \leq J \leq N_{\text{clauses}}$, that make up a formula or proposition

$$A_1 \wedge A_2 \wedge ... \wedge A_{N_{\text{clauses}}}, \tag{2.2.1}$$

which is satisfied if all clauses are satisfied[4]. The clauses consist of three so-called literals. That is, the clauses are of the form[5]

$$A_J = (a_i \vee \bar{a_j} \vee a_k), \tag{2.2.2}$$

where the indices $1 < i, j, k < N$ specify the bits. In 3-SAT, a clause is satisfied if one out of the three literals is satisfied by the bit-string. In 1-in-3-SAT, the clause is only satisfied if exactly one out of the three literals is satisfied. The goal of 3-SAT and 1-in-3-SAT is to determine if there exists a bit-string that satisfies all clauses and therefore satisfies the given formula.
If there is no efficient algorithm to solve 1-in-3-SAT, there is also no efficient algorithm to find the optimal bit-string that satisfies as many clauses as possible. For if such an algorithm would exist, we could simply check if the optimal bit-string satisfies all clauses or not, which would give an efficient algorithm for 1-in-3-SAT. The fitness of a bit-string will be based on the number of satisfied clauses. We could therefore view our evolutionary system as a genetic algorithm that tries to find bit-strings that satisfy many clauses.

We use $N_{\text{clauses}} = 6N$ and choose the clauses randomly. That is, the indices i, j and k are chosen uniformly random from $\{1, 2, ..., N\}$ and the literals are negated with probability $\frac{1}{2}$. The clauses are chosen once and are reused for every simulation, so that the fitness landscape based on 1-in-3-SAT will be the same for every simulation. Because there are $6N = 1200$ clauses, it is very improbable that there exists a bit-string that satisfies all clauses. By using many clauses, we create an interesting and complex landscape with many local optima.
The actual fitness function $f(x)$ is given by $\exp\left(-\frac{N_{\text{unsatisfied}}(x)}{N}\right)$, where $N_{\text{unsatisfied}}(x) = N_{\text{clauses}} - N_{\text{satisfied}}(x)$ is the number of unsatisfied clauses. Every satisfied clause increases the fitness by a factor $\exp\left(\frac{1}{N}\right) = \exp\left(\frac{1}{200}\right) \approx 1.005$ and makes it more probable that the particular bit-string is selected for reproduction. This is only a slight advantage, but note that we are interested in the behaviour of the model instead of the actual optimal bit-string.

[2] This fitness landscape was taken from [Brotto, Bunin, and Kurchan, 2015].
[3] NP is the class of decision problems for which a solution can be checked in polynomial time, while P is the class of problems that can be solved in polynomial time [Korte and Vygen, 2012]. By polynomial time, we mean that the computational time or the number of operations needed to solve a problem is bounded by a polynomial in the problem size n. For purposes of illustration, it may well be that a polynomial algorithm solves an instance of $n = 10^6$ in the same computational time as an exponential algorithm solves an instance of $n = 50$.
[4] *Here*, $\wedge$ means "AND".
[5] Here, $\vee$ means "OR" and $\bar{a_j}$ means negation of a_j.

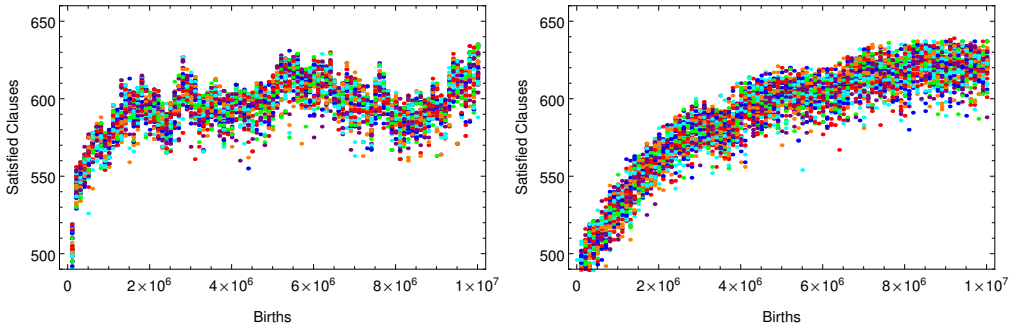

Figure 2.1 – The number of satisfied clauses of (a random subset of) the population members as a function of the number of births in the system. On the left, the population size is $M = 500$, while on the right $M = 5000$.

2.3 Results

Now that the model has been completely defined, we can look at some results. We use population sizes $M \in \{10, 100, 500, 1000, 2000, 5000\}$ and propagate the populations for 10^7 time steps of the Moran process, so there are 10^7 births and deaths in every system. This way, the computational effort needed for a simulation is fairly comparable for different population sizes, although selection takes longer in a larger population. Pseudo-code can be found in Appendix B.1. We average results over an ensemble of 100 populations for every population size.

The evolution of two populations of sizes 500 and 5000 is shown in Figure 2.1, where the number of satisfied clauses of a random subset of the population is shown as a function of time. We see that the fitness of the population members indeed increases over time. The system with population size 500 has reached an equilibrium around which it fluctuates, while the system with population size 5000 is still slowly improving.
Note that the best bit-string found by random sampling in the same computational time as all the simulations, typically does not exceed 540 satisfied clauses. This shows that most of the 2^{200} bit-strings satisfy less than 540 clauses. From the figures, we see that natural selection is able to find better bit-strings, which validates using genetic algorithms for optimization purposes. Evolution searches locally without getting stuck in local minima. It exploits the fact that similar bit-strings often have similar fitness, so that good bit-strings are close to other good bit-strings.

From Figure 2.1, we see that a large population converges slowly compared to a smaller population when measuring time in births[6]. To investigate the rate of change in the population, we now inspect the autocorrelation of the system in equilibrium for different population sizes. We define the autocorrelation in the following way.[7]

$$C(\Delta t) = \frac{4}{NM}\mathrm{E}\left[\sum_{i=1}^{M}\sum_{j=1}^{N}\left(a_j^i(t) - \frac{1}{2}\right)\left(a_j^i(t+\Delta t) - \frac{1}{2}\right)\right], \tag{2.3.1}$$

where the double sum is over all bits j of all bit-strings i. We measure time in generations instead of births[8], so $\Delta t = A$ means that every population member has reproduced A times on average. Formulated differently, $\Delta t = A$ means that there have been AM births. Therefore, the total number of births per unit of time is different for different population sizes.

[6] In genetic algorithms, births are a good measure of computational time.
[7] Note that with this definition a population is an ordered list and that we compare bit-strings that have the same index.
[8] When we are not considering computational effort, measuring time in generations is more natural, because the unit of time is equal to the average time that a single particle needs to reproduce.

The autocorrelation C has the property that a population has autocorrelation $C(0) = 1$ with itself. Also, if the population is completely uncorrelated to itself after some time, $C(\infty) = 0$. To measure the autocorrelation, we take the average over the ensemble and over time, after an initialization period of 5000 generations:

$$C_{\text{measured}}(\Delta t) = K \sum_{l=1}^{100} \sum_{t=5000}^{t_{\text{final}}} \sum_{i=1}^{M} \sum_{j=1}^{N} (a_j^i(t) - \frac{1}{2})(a_j^i(t+\Delta t) - \frac{1}{2}), \tag{2.3.2}$$

where K is a constant which ensures that $C_{\text{measured}}(0) = 1$. We take a measurement every 100 generations[9], so t increases in steps of 100.

The measured autocorrelation is shown in Figure 2.2 for different population sizes. Even though there are more births per generation[10], a larger population is more correlated to its previous state than a smaller population in the same number of generations. That is, the change per generation is smaller.

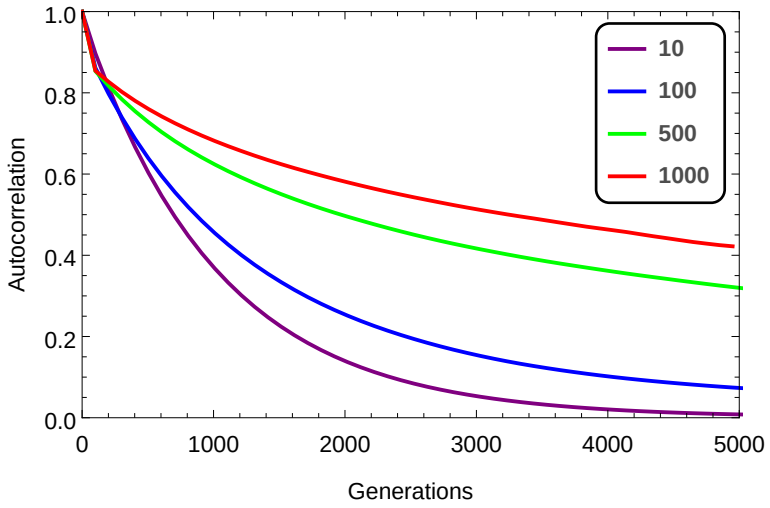

Figure 2.2 – The autocorrelation C as a function of generations for different population sizes. Large populations remain correlated to their previous state for a very long time.

[9] These measurements are correlated, but this does not pose a problem for computing the mean. We just do not have the same accuracy of the mean compared to averaging over the same number of independent measurements.

[10] The number of births per generation is M, so it already takes more computational effort to generate a single generation step.

2.4 Thermodynamics

There have been several attempts to explain evolutionary systems using ideas from statistical physics, for example by Brotto et al. [2015], Peliti [1997] and Sella and Hirsh [2005].

Fitness is similar to energy in the sense that a physical system tries to minimize its energy, while an evolutionary system tries to maximize its fitness. However, both systems are restricted from attaining the absolute minimum or maximum by probabilistic considerations. Because random mutations in a finite population can drive the population to a great many other states, the probability of finding the population in the state of absolutely maximum fitness is small. This resembles the notion of entropy in a physical system at finite temperature[11].

It can be shown that $\frac{1}{M}$ plays the role of temperature in an evolutionary system. Brotto et al. [2015] show that the framework of statistical physics applies to evolutionary systems under certain conditions. They show that transition rates between coarse-grained states of the evolutionary system are equivalent to the transition rates of a physical system at temperature proportional to $\frac{1}{M}$. Analogous to energy-increasing transitions in the physical system, fitness-decreasing transitions are more probable in small populations (high temperature) than in large populations (low temperature). This comes with the side effect that a large population can get stuck in a local optimum more easily than a small population.

The correspondence can be exploited by studying the behaviour of evolutionary systems using theory or techniques from statistical physics, for example by using parallel tempering to study evolutionary systems by simulations [Brotto et al., 2015]. We will now compare the method of simulated annealing to genetic algorithms using the connection of the population size to temperature.

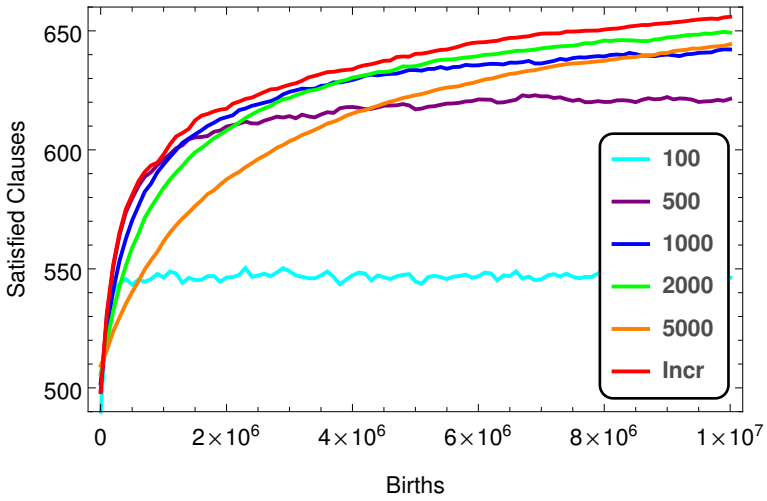

Figure 2.3 – The average highest number of satisfied clauses as a function of the number of births in the system for different population sizes. "Incr" stands for increasing population size as described in Section 2.5. In Figure 2.4, the same data is shown with a log-scale for the x-axis.

[11] Note that this discussion refers to systems that are in equilibrium. In evolutionary systems, however, we are often interested in evolving non-equilibrium systems. This would lead to a comparison with non-equilibrium physics.

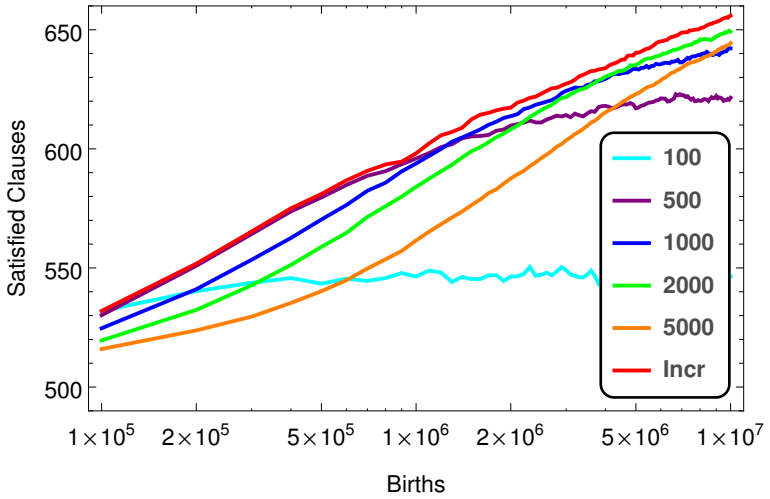

Figure 2.4 – Equivalent to Figure 2.3, but with a log-scale for the x-axis. The approximately straight lines in this figure show that the time to improve by $\Delta N_{\text{satisfied}}$ clauses increases approximately exponentially until evolution slows down near equilibrium.

2.5 Simulated annealing

In Figure 2.1, we have seen that a small population adapts quickly, while a large population eventually obtains a higher equilibrium fitness. To get more solid results, we now look at average results over an ensemble of 100 systems. In Figure 2.3, the average highest number of satisfied clauses[12] given by

$$\frac{1}{100}\sum_{j=1}^{100}\max_i N_{\text{satisfied}}(x_i^j) \tag{2.5.1}$$

is shown as a function of time. Here, the sum runs over the populations j in the ensemble and i labels the population members of population j.
We see that smaller populations indeed converge faster in the beginning, while larger populations attain a higher equilibrium fitness. Note that all simulations are done in the same fitness landscape. The only differences are the random components in initializations, births, deaths and mutations. Results for the minimum or maximum of the ensemble are similar to that of the average.

Now we implement a procedure similar to simulated annealing. In simulated annealing, the temperature of the system is decreased gradually, so that the system can find a global rather than a local optimum. At the beginning, the system explores the landscape very roughly but quick, because the high temperature means that many states of the system are acceptable. By decreasing the temperature, the system becomes less likely to accept higher energy states and the system hopefully converges towards the global minimum.
By recognizing that $\frac{1}{M}$ plays the role of temperature, the corresponding procedure for our evolutionary system would be to increase the population size gradually. We start at 500 particles and increase it in steps of 250 until 2750. We would not expect better equilibrium results than by using 2750 particles from the start, but the equilibrium might be reached faster. From Figure 2.3, we see that this is indeed the case.
Instead of increasing the population size in steps of 250, we could have also created the possibility of births without deaths, so that the population size increases gradually.

[12]The best population member represents the fitness of the population.

Chapter 3

Continuum modelling of birth-death processes

The rest of this thesis focuses on birth-death processes in a continuous space with diffusion as mutation. We will try to describe the population by a continuous density and its evolution by a partial differential equation.

3.1 Micro-model description

We start with a complete description of the micro-model. Consider a population of M particles defined by their position x in the d-dimensional Euclidean space, $x_i \in \mathbb{R}^d$ for $1 \leq i \leq M$. These particles reproduce and die according to a Moran process. That is, at a single reproduction event there is one birth and one death. This ensures constant population size. In between reproduction events, all particles diffuse independently.

The reproduction events are randomly distributed in time like a Poisson process with rate[1] $\lambda(t)M$, where the rate per particle $\lambda(t)$ will be specified later. The Poisson process is appropriate for discrete events that happen completely independently, but with a well-defined average rate. In Appendix A.2, we elaborate on our choice for the Poisson process. For now, the important result is that the time intervals between reproduction events, denoted by Δt and illustrated in Figure 3.1, are exponentially distributed with parameter $\lambda(t)M$. That is, the probability density function of the time increments Δt between reproduction events is given by

$$p(\Delta t) = \lambda(t)M \exp\Big(- \lambda(t)M\Delta t \Big), \tag{3.1.1}$$

which results in the expected value of Δt given by

$$\mathrm{E}(\Delta t) = \int_0^\infty \Delta t p(\Delta t) d(\Delta t) = \frac{1}{\lambda(t)M}. \tag{3.1.2}$$

The rate $\lambda(t)M$ is the expected number of reproduction events per unit of time, so the expected value of the time increments is the inverse of the rate, as it should be.

It is necessary to keep track of time explicitly, because mutation of the particles is implemented by diffusion, which happens continuously in time. The strength of the diffusion is represented by the diffusion constant D. Each particle x performs a continuous random walk. More specifically, the displacement $\Delta x = x(t+\Delta t) - x(t)$ follows a normal distribution with expectation $\mathrm{E}(\Delta x) = 0$ and covariance $\mathrm{Cov}(\Delta x) = 2D\Delta t I_d$. Here I_d is the d-dimensional identity matrix reflecting d independent and identically distributed components with variance $2D\Delta t$.

[1] The rate is the average number of events per unit of time.

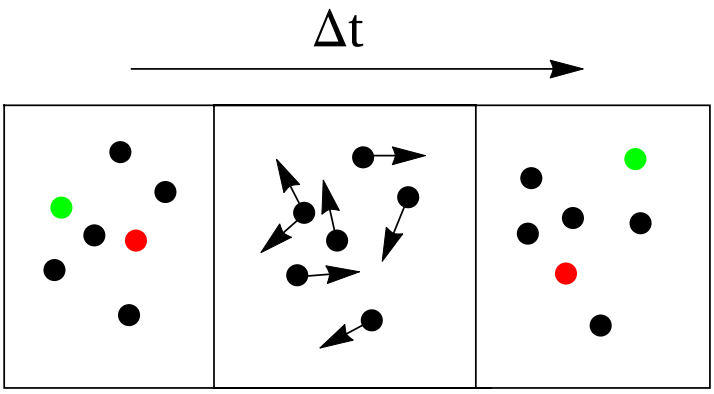

Figure 3.1 – Illustration of diffusion between two consecutive reproduction events. The green particles reproduce while the red particles die. The arrows represent diffusion. The time increment Δt follows an exponential distribution.

Recall that Δt between reproduction events is exponentially distributed. We can generate an exponentially distributed variable by the method of inverse transform sampling [Devroye, 1986]. Given a uniform random variable Y on $[0, 1)$[2], the variable $\frac{\log(1-Y)}{\lambda(t)M}$ is distributed exponentially with parameter $\lambda(t)M$.[3] After we generated Δt, we can update the locations of the particles to account for the displacement due to diffusion. Because $x(t + \Delta t) = x(t) + \Delta x$ and Δx is normally distributed, we obtain $x(t + \Delta t)$ by adding a normal variable with standard deviation equal to $\sqrt{2D\Delta t}$ to every component of x. For clarity, pseudo-code of the core of the micro-process is given in Appendix B.2.

In our model, the particles need not reproduce equally fast. We define a fitness landscape, $f(x) > 0$ for every $x \in \mathbb{R}^d$, which is the reproduction rate of a particle at x. Without loss of generality we can write $f(x) = \exp(-kU(x))$, where U is similar to a potential and k can be used to set the selection strength. Specific fitness landscapes are defined later on.

The fitness of a particle $f(x_i)$ represents its reproduction rate in the sense of the Poisson process. Because $f(x_i)$ is the average number of reproductions per unit of time of particle i, the sum $\sum_i f(x_i)$ is the average number of reproductions per unit of time of the entire population. So the rate at which reproduction events occur in the population is given by $\lambda(t)M = \sum_i f(x_i)$[4]. We sample reproduction events at the rate $\sum_i f(x_i)$ and at every reproduction event, particle x_i is selected for reproduction with probability

$$\mathrm{P}(\text{particle } i \text{ is selected}) = \frac{f(x_i)}{\sum_i f(x_i)}. \tag{3.1.3}$$

This way, every particle x_i reproduces with the correct rate $f(x_i)$, because out of the $\sum_i f(x_i)$ reproductions in a unit of time, on average there are

$$\left(\frac{f(x_i)}{\sum_i f(x_i)}\right) \sum_i f(x_i) = f(x_i) \tag{3.1.4}$$

reproductions of particle x_i per unit of time.

[2] To generate a uniform random variable, we used a pseudo-random number generator by Matsumoto and Nishimura [2002] also known as the Mersenne Twister.

[3] Note that we calculate the parameter of the exponential distribution, $\lambda(t)M$, at t and implicitly assume that it is constant during $(t, t + \Delta t)$. This introduces a slight bias, because $\lambda(t)$ is actually continuously changing due to continuously diffusing particles. However, the variation of $\lambda(t)$ is only small during a short time interval Δt, so it is a minor flaw. Moreover, it is impossible to update $\lambda(t)$ continuously in a computer simulation, although it is possible to split Δt into smaller time steps.

[4] Recall that we defined $\lambda(t)$ such that $\lambda(t)M$ was the reproduction rate of the population.

When a particle at x reproduces, a new particle is created at the same location x. The new particle diffuses independently of its parent afterwards. At every reproduction there is also a particle that dies. A randomly chosen individual is removed from the population. This means that all particles die equally likely with a rate given by

$$\lambda(t) = \frac{1}{M} \sum_i f(x_i), \tag{3.1.5}$$

so $\lambda(t)$ is the death rate of particles. Because the number of deaths is determined by the number of births, the death rate depends on the fitness of the population. Indeed, $\lambda(t)$ is equal to the average fitness of the population. The fitness of one particle influences the death rate of another particle, which is the only interaction between particles in this system.

Our current model contains the dimensions of time and space. By making our model dimensionless, we could scale the diffusion constant D to $D = 1$. However, it is more convenient to define some fitness landscape and keep D as a parameter to scale diffusion compared to the length scale of the fitness landscape. We should just acknowledge that scaling the fitness landscape $f(x) \to \alpha f(\beta x)$ with $\alpha > 0$ and $\beta > 0$ gives essentially the same model[5].
The point here is that we can choose different units of time and length[6] such that $\alpha f(\beta x)$ and $f(x)$ have exactly the same numerical dynamics. Measuring time in hours instead of seconds or measuring the positions in inches instead of meters changes the numerical values we work with, but does not actually change the system. Because we just need to choose specific units of time and length to find the same numerical dynamics, the systems are equivalent for analytical purposes.

3.2 Fitness landscapes

We now introduce some fitness landscapes. Fitness is defined to be the reproduction rate, which is a positive quantity. We only consider reproducing particles, so the fitness is assumed to be strictly positive. This is the only restriction on our fitness function f, but we will focus on continuous fitness landscapes.

3.2.1 Peaks

As a first example to illustrate the behaviour of our model, we use a population of $M = 100$ particles in a two-dimension space, $d = 2$, with a fitness landscape that possesses several local minima. It is defined by $f(x) = \exp(-\frac{1}{10}U(x))$ where[7]

$$U(x) = 3(1-x_1)^2 e^{(-x_1^2-(x_2+1)^2)} - \frac{1}{3} e^{(-(x_1+1)^2-x_2^2)} - 10(-x_1^3+x_1/5-x_2^5) e^{(-x_1^2-x_2^2)}. \tag{3.2.1}$$

The fitness potential U is shown in Figure 3.2. Depending on the diffusion constant D, we find different behaviour. If the diffusion constant D is too large, the population will spread beyond the length-scales of the interesting part of this landscape. Eventually, the population will escape the interesting region entirely and will evolve to parts where U is nearly flat. If D is too small, the population will get stuck in the local minimum where it started. In between these extremes, the population will be stuck in a local minimum for most of the time, while it is still able to switch to another local minimum by stochastic escape. Such a transition is shown in Figure 3.2. The transition time is small compared to the time spent in the minima.

[5] In general with a different value of D

[6] Different in the sense that the first system is measured in other units than the second system. For example, in one system length is measured in micrometres and in the other in kilometres.

[7] U is the peaks function from MATLAB

In any case, the population behaves as a cluster that moves as a whole. The clustering behaviour in a flat landscape is analysed in an article by Meyer et al. [1996][8]. The main results are stated in Appendix A.3. Instead of describing statistics, we would like to represent the population by a population density $u(x)$ at every x and describe its evolution. This way, we do not need to describe every single particle, which might be useful in very large populations where keeping track of all particles is impossible. This could be viewed as an instance of multi-scale modelling.

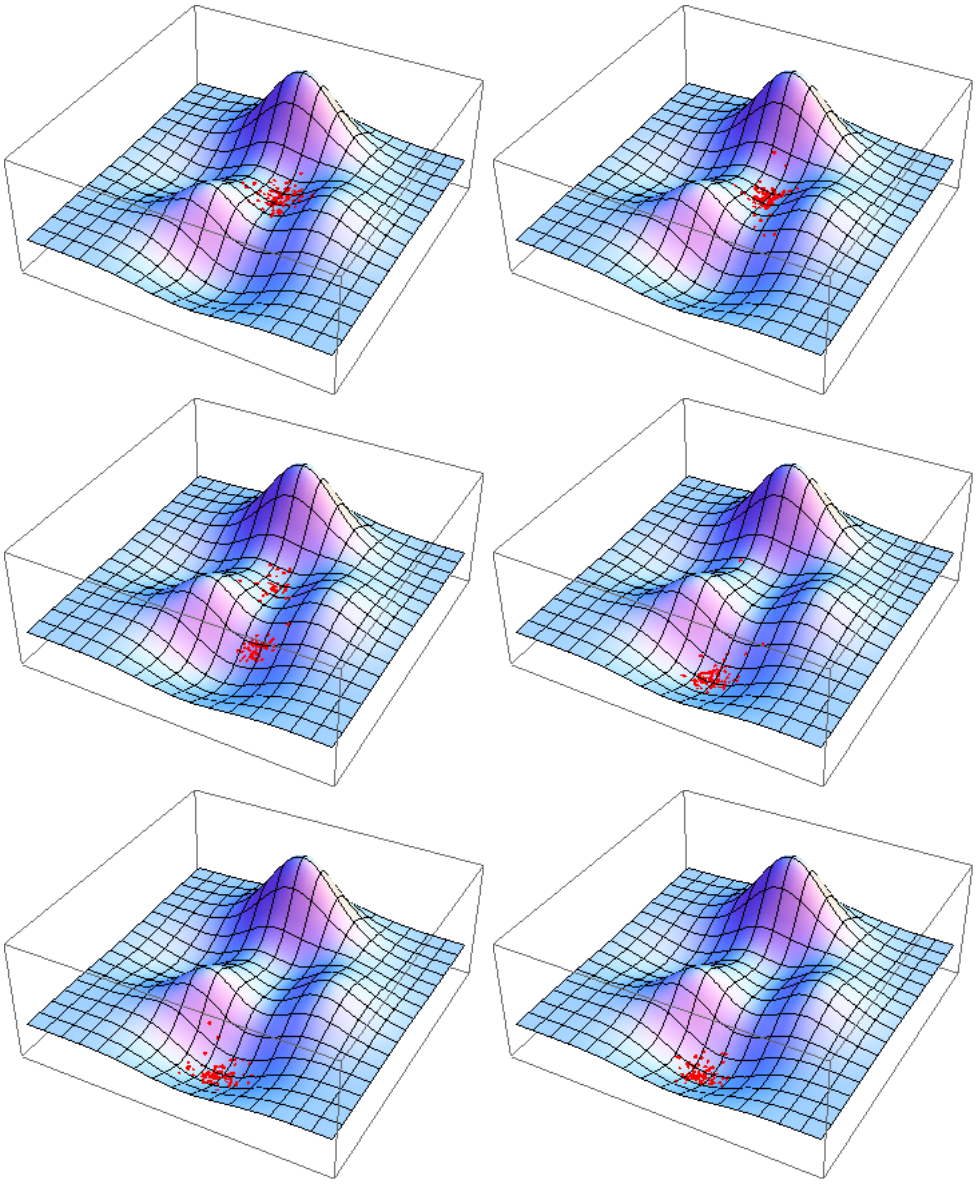

Figure 3.2 – A birth-death process with diffusion in the peaked landscape defined in Equation (3.2.1).

[8]For other articles on similar models, see [Kessler, Levine, Ridgway, and Tsimring, 1997] and [Zhang, Serva, and Polikarpov, 1990]

3.2.2 Double well

In the previous landscape, we have seen that the population can get stuck in local optima. To study this in a simple setting, we use the one-dimensional double well. The fitness function is given by $f(x) = \exp(-kU(x))$, where

$$U(x) = x^4 - 2x^2. \tag{3.2.2}$$

This potential exhibits minima at $x = \pm 1$[9]. To determine the selection probabilities $\frac{f(x_i)}{\sum_i f(x_i)}$, we only need to know the relative fitness of the particles. Therefore, we can scale the fitness function without changing the selection probabilities, $f \to \alpha f$ with $\alpha > 0$. The potential U and the scaled fitness functions f are shown in Figure 3.3 for several values of k. The selection parameter k determines the depth of the fitness barrier.

To illustrate behaviour that is possible in the double well, Figure 3.4 and Figure 3.5 show the population as a function of time for different values of k and D. We used a population of $M = 100$ particles. From the figures, we see that the population can get stuck in one of the two wells if the diffusion constant D is small enough or if the selection strength k is strong enough.

The average shape of the population density is shown in Figure 3.6. We see that on average most of the population is near the centre of the population, while there are some particles that are further away. Note that the actual shape of the population at some instance of time does not need to be equal or even similar to the average shape.

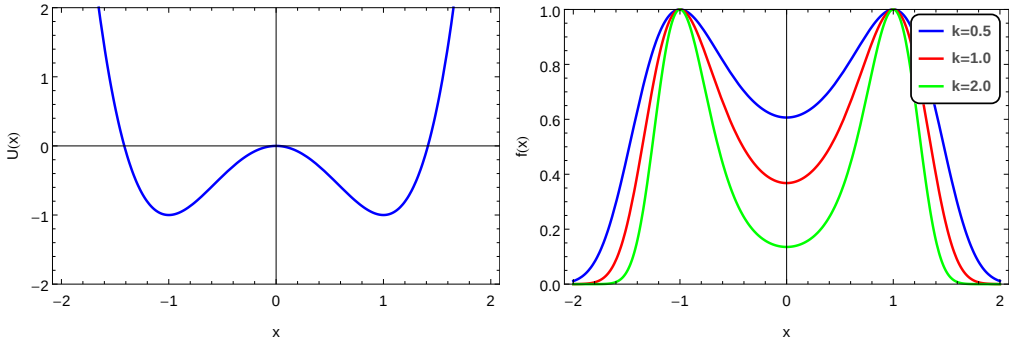

Figure 3.3 – The double well potential $U(x)$ (left) and fitness $f(x)$ (right) for several values of k.

[9]Note that for any $c > 0$, there exist $a > 0$ and $b > 0$ such that $a((bx)^4 - c(bx)^2) = x^4 - 2x^2$. Therefore, by scaling the length scale by b and the selection strength by a, we can turn any potential of the form $U(x) = x^4 - cx^2$ with $c > 0$ into $U(x) = x^4 - 2x^2$, which makes our discussion more generally applicable.

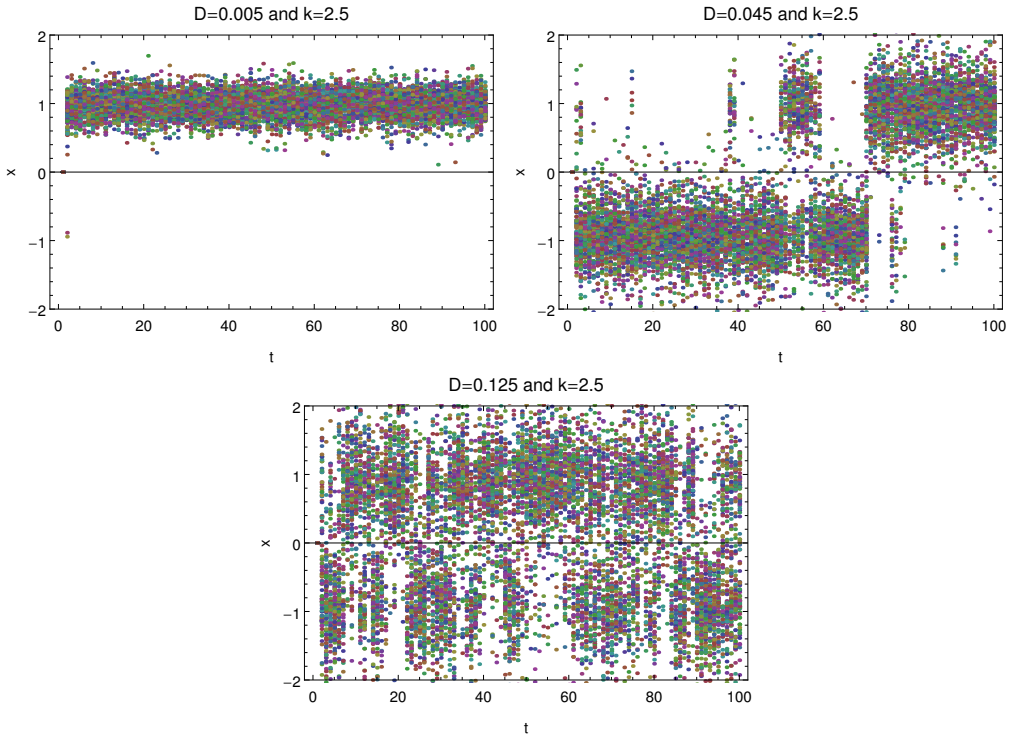

Figure 3.4 – Typical evolution of particles in the double well potential. The positions of the particles are shown as a function of time for different values of D and $k = 2.5$. With a larger value of D, the population becomes more spread. The values of D are chosen such that $\sqrt{2D}$ is simple.

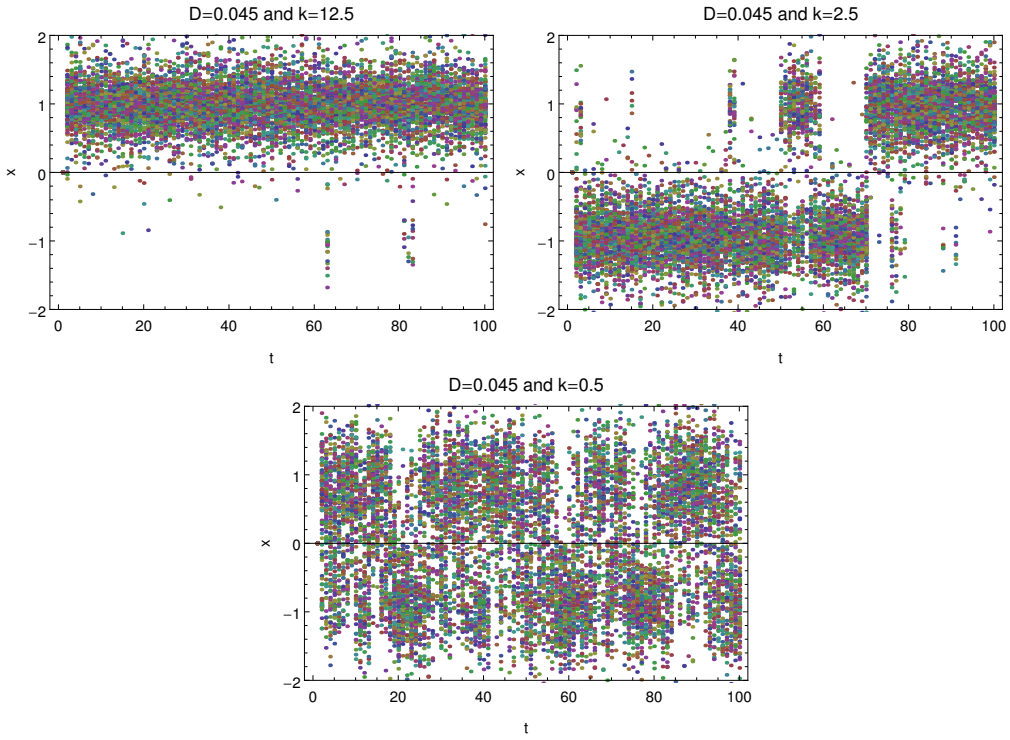

Figure 3.5 – Similar to Figure 3.4, but with several values of k and $D = 0.045$. With a larger value of k, it is harder to cross the fitness barrier of the double well potential.

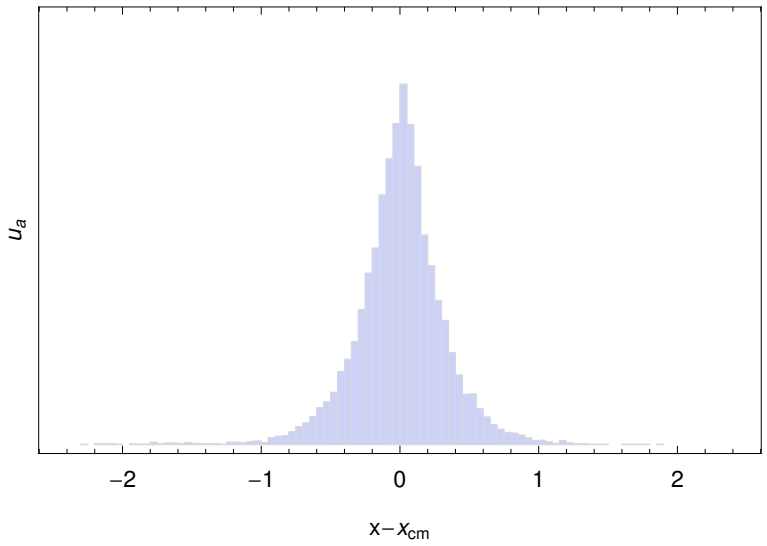

Figure 3.6 – Average shape of the population for $k = 12.5$ and $D = 0.045$. The positions are shifted such that 0 corresponds to the centre of mass of the population.

3.2.3 Hydrogen potential

Our final example of a fitness landscape is based on the three-dimensional hydrogen potential. In spherical coordinates, the fitness function f is given by[10]

$$f(r, \theta, \phi) = \frac{c}{r}, \tag{3.2.3}$$

for some $c > 0$, where $r = \sqrt{x_1^2 + x_2^2 + x_3^2} \geq 0$. We are interested in this landscape, because it turns out that we can use theory from the hydrogen atom.

[10]Note that we drop the U notation.

3.3 Continuous description

3.3.1 Derivation of the partial differential equation

We now stop looking at individual particles. Instead, we take a coarser approach by looking at the density of particles as a function of x, which we denote by $u(x)$. We will now derive a partial differential equation for this density in one dimension. The equation we will derive is given by

$$u_t = (f(x) - \lambda(t))u + Du_{xx}, \tag{3.3.1}$$

where $u_t = \frac{\partial u}{\partial t}$ and $u_{xx} = \frac{\partial^2 u}{\partial x^2}$. We see that there is a term for births and death, $(f(x) - \lambda(t))u$, and another term for diffusion.

To arrive at this equation, we write down equations in discretized time and space after which we take a limit to make the equations continuous. This is a common procedure to arrive at partial differential equations [Holmes, 2009].
Let $w(m, n)$ be the number of particles in the interval $I_m = \left(x_m - \frac{1}{2}\Delta x, x_m + \frac{1}{2}\Delta x\right)$ averaged over the time interval $\left(t_n - \frac{1}{2}\Delta t, t_n + \frac{1}{2}\Delta t\right)$. We consider diffusion first. Assume that in a single time step, the number of particles that leave the region is equal to $\alpha w(m, n)$ with $0 < \alpha \ll 1$. Assume that particles do not have a preference for a direction and that particles can only reach adjacent regions in a single time step. This results in the equation

$$w(m, n) = \frac{\alpha}{2} w(m - 1, n - 1) + (1 - \alpha)w(m, n - 1) + \frac{\alpha}{2} w(m + 1, n - 1). \tag{3.3.2}$$

The number of particles $w(m, n)$ is equal to the sum of the remaining number of particles, $(1 - \alpha)w(m, n - 1)$, and the number of particles that enter from adjacent regions, which is given by $\frac{\alpha}{2} w(m - 1, n - 1) + \frac{\alpha}{2} w(m + 1, n - 1)$.

Now we consider reproduction and death, with rates given by f and λ. Assume there are $f(x)$ reproductions and $\lambda(t)$ deaths per unit of time per particle for a particle at position x at time t. Define $\bar{f}(m)$ and $\bar{\lambda}(n)$ on the grid by averaging f and λ over the grid cells. Now we assume that the time-steps are small enough so that $\Delta t \bar{f}(m) \ll 1$ and $\Delta t \bar{\lambda}(n) \ll 1$. In this regime, we can interpret $\Delta t \bar{f}(m)$ and $\Delta t \bar{\lambda}(n)$ as the probability that a particle reproduces or dies. In that case, $\Delta t \bar{f}(m) w(m, n)$ and $\Delta t \bar{\lambda}(n) w(m, n)$ approximates the number of particles that reproduce or die. Adding $\Delta t \bar{f}(m) w(m, n) - \Delta t \bar{\lambda}(n) w(m, n)$ to Equation (3.3.2) and rewriting, we find[11]

$$w(m,n) = \Big(1+\Delta t(\bar{f}(m)-\bar{\lambda}(n))\Big)\left[\frac{\alpha}{2}w(m-1,n-1)+(1-\alpha)w(m,n-1)+\frac{\alpha}{2}w(m+1,n-1)\right]. \tag{3.3.3}$$

In order to derive a continuous equation, we scale $w(m, n)$ to a density $u(x, t)$ and explicitly write down the equation in terms of Δt and Δx, which gives

$$u(x,t) = \Big(1+\Delta t(\bar{f}(x)-\bar{\lambda}(t))\Big)\left[\frac{\alpha}{2}u(x-\Delta x,t-\Delta t)+(1-\alpha)u(x,t-\Delta t)+\frac{\alpha}{2}u(x+\Delta x,t-\Delta t)\right]. \tag{3.3.4}$$

[11] This can be interpreted as applying a birth-death operator after a diffusion operator.

Now we perform a Taylor expansion of $u(x-\Delta x, t-\Delta t)$, $u(x, t-\Delta t)$ and $u(x+\Delta x, t-\Delta t)$ for small Δx and Δt. This results in

$$\begin{aligned} u = &\left(1+\Delta t(\bar{f}(x)-\bar{\lambda}(t))\right) * .. \\ &\left[\frac{\alpha}{2}\left(u - \Delta x u_x - \Delta t u_t + \frac{1}{2}(\Delta x^2 u_{xx} + 2\Delta x \Delta t u_{xt} + \Delta t^2 u_{tt}) + \mathcal{O}(\Delta t^3 + \Delta x^3)\right) + .. \right. \\ &(1-\alpha)\left(u - \Delta t u_t + \frac{1}{2}\Delta t^2 u_{tt} + \mathcal{O}(\Delta t^3)\right) + .. \\ &\left.\frac{\alpha}{2}\left(u + \Delta x u_x - \Delta t u_t + \frac{1}{2}(\Delta x^2 u_{xx} - 2\Delta x \Delta t u_{xt} + \Delta t^2 u_{tt}) + \mathcal{O}(\Delta t^3 + \Delta x^3)\right)\right]. \end{aligned} \tag{3.3.5}$$

Here, $\mathcal{O}(\Delta t^2 + \Delta x^3)$ represents higher order terms in Δt or Δx that go to zero at least as fast as $(\Delta t^2 + \Delta x^3)$.[12] This simplifies to

$$0 = \Delta t(\bar{f}(x) - \bar{\lambda}(t))u - \Delta t u_t + \frac{\alpha}{2}\Delta x^2 u_{xx} + \mathcal{O}(\Delta t^2 + \Delta x^3) \tag{3.3.6}$$

or

$$u_t = (\bar{f}(x) - \bar{\lambda}(t))u + \frac{\alpha \Delta x^2}{2\Delta t} u_{xx} + \mathcal{O}\left(\Delta t + \frac{\Delta x^3}{\Delta t}\right) \tag{3.3.7}$$

We now take the limit $\Delta t \to 0$ and $\Delta x \to 0$. The most interesting limit is obtained by keeping $\frac{\Delta x^2}{\Delta t}$ fixed. However, the fraction of leaving particles α is actually a function of Δt and Δx. Therefore, we want to take $\Delta t \to 0$ and $\Delta x \to 0$ in such a way that α is constant. Fortunately, for a standard random walk this should be such that $\frac{\Delta x^2}{\Delta t}$ is fixed.[13] Defining $D = \frac{\alpha \Delta x^2}{2\Delta t}$ and taking the limit, we find

$$u_t = (f(x) - \lambda(t))u + D u_{xx}, \tag{3.3.8}$$

which we set out to derive. We will from now on call this equation the birth-death-diffusion equation.

The death rate λ should be such that the population size is conserved. Setting

$$\frac{\partial}{\partial t}\int u(x,t)dx = \int \left(f(x) - \lambda(t) + D\frac{\partial^2}{\partial x^2}\right)u(x,t)dx = 0, \tag{3.3.9}$$

we find that the death rate should be equal to the average fitness,

$$\lambda(t) = \frac{\int f(x)u(x,t)dx}{\int u(x,t)dx}. \tag{3.3.10}$$

To derive this, note that $\int \frac{\partial^2}{\partial x^2}u(x,t)dx = \frac{\partial}{\partial x}u(x\to\infty,t) - \frac{\partial}{\partial x}u(x\to-\infty,t) = 0$, because the population is localized and vanishes at infinity.

The first term of the birth-death-diffusion equation is similar to the so-called replicator equation [Ohtsuki and Nowak, 2006], which is given by $\partial_t u_i = (f_i - \sum_i u_i f_i)u_i$ with $\sum_i u_i = 1$, but in a continuum setting.

[12] By definition, $f(x) = \mathcal{O}(g(x))$ as $x \to 0$ if there exist $M > 0$ and $\delta > 0$ such that $|f(x)| < M|g(x)|$ for every $x \in (-\delta, +\delta)$.

[13] The displacement of a random walker in time Δt follows a normal distribution with variance equal to $2D\Delta t$. This probability distribution is proportional to $\exp(\frac{x^2}{2D\Delta t})$, which is a function of $\frac{x^2}{\Delta t}$. Therefore, to obtain a constant α, Δx should go to zero in such a way that $\frac{\Delta x^2}{\Delta t}$ is constant.

3.4 Solutions to the equation

In this section, we look at solutions of the just derived birth-death-diffusion equation

$$\partial_t u = (f(x) - \lambda(t) + D\partial_{xx})u. \tag{3.4.1}$$

This equation resembles the time-dependent Schrödinger equation for a single non-relativistic particle,

$$i\hbar\partial_t \Psi(x,t) = \frac{-\hbar^2}{2m}\nabla^2\Psi(x,t) + V(x,t)\Psi(x,t), \tag{3.4.2}$$

and we can solve the equation using techniques from physics[14]. First we look at solutions to the related equation[15]

$$\partial_t \tilde{u} = (f(x) + D\partial_{xx})\tilde{u}. \tag{3.4.3}$$

Using separation of variables, writing $\tilde{u}_n(x,t) = v_n(t)w_n(x)$ and plugging this in Equation (3.4.3), we find a simple equation for v_n, namely $\partial_t v_n = -E_n v_n$ for some constant E_n, which is solved by

$$v_n(t) = C\exp(-E_n t). \tag{3.4.4}$$

The spatial part $w_n(x)$ should then satisfy the eigenvalue equation

$$(-f(x) - D\partial_{xx})w_n(x) = E_n w_n(x). \tag{3.4.5}$$

These w_n are stationary states of the original system if we set $\lambda(t) = -E_n$. Now assume that the w_n form a basis of the function space, so that for any initial condition $\tilde{u}(x,0) = u_0(x)$ we can write $u_0(x) = \sum c_n w_n(x)$ for some constants c_n[16]. By linearity of the differential equation, Equation (3.4.3), and applying the defining equation for the w_n, Equation (3.4.5), it follows that

$$\tilde{u}(x,t) = \sum c_n w_n(x)\exp(-E_n t). \tag{3.4.6}$$

solves Equation (3.4.3) with initial condition $\tilde{u}(x,0) = u_0(x)$.

For the solution of the original equation, Equation (3.4.1), we write

$$u(x,t) = \sum c_n(t) w_n(x)\exp(-E_n t), \tag{3.4.7}$$

with time-dependent coefficients $c_n(t)$. Since the w_n are a basis of the function space, this expression is completely general. By inserting it in Equation (3.4.1),

$$\partial_t\left[\sum c_n(t)w_n(x)\exp(-E_n t)\right] = (f(x) - \lambda(t) + D\partial_{xx})\left[\sum c_n(t)w_n(x)\exp(-E_n t)\right], \tag{3.4.8}$$

we hope to get simpler relations for the $c_n(t)$. Indeed, working out the derivative[17] and using the relation in Equation (3.4.5), terms cancel and rewriting results in

$$\sum\left[\partial_t c_n(t) + \lambda(t)c_n(t)\right]w_n(x)\exp(-E_n t) = 0. \tag{3.4.9}$$

[14] We will see that the absence of the imaginary i results in convergence towards the lowest-energy state. This can be used to compute the ground state for quantum systems. These methods are called imaginary time propagation methods. By the results of this thesis, one could try to calculate the ground state by the micro-model.

[15] Note that this equation is linear in u, while the original equation is non-linear in u, because $\lambda(t)$ depends on u.

[16] For notational convenience, we assume that the spectrum of the operator $(-f(x) - D\partial_{xx})$ is discrete, so that we can use a sum to represent $u_0(x)$. This will be the case for $f(x)$ used in Section 3.5. The arguments in this section also hold if the operator has a continuous spectrum, but with an integral.

[17] We assume that we can interchange the order of the derivative and the infinite sum. This is not always true, but it is true when the convergence of the sum meets some conditions. For example, uniform convergence of the sum of derivatives is sufficient, because we already assume that the sum for u converges.

This would be true, and it would therefore solve the equation, if

$$\partial_t c_n(t) = -\lambda(t) c_n(t). \tag{3.4.10}$$

for all n. Note that if the w_n are orthogonal, it is also a necessary condition. The equation for $c_n(t)$ is the same for every n, which implies that $c_n(t) = c_n c(t)$ for some function $c(t)$. Plugging this in, we find the most important equation of this section,

$$u(x,t) = c(t) \sum c_n w_n(x) \exp(-E_n t). \tag{3.4.11}$$

The function $c(t)$ can be found by the condition that $c(t)$ ensures normalization. We see that the solution tends to converge towards the w_n with smallest E_n, the ground state.

The function $c(t)$ is such that $\int u(x,t)dx = 1$, so[18]

$$c(t) \sum \left(c_n \exp(-E_n t) \int w_n(x) dx \right) = 1, \tag{3.4.12}$$

from which we see that

$$c(t) = \frac{1}{\sum \left(c_n \exp(-E_n t) \int w_n(x) dx \right)}. \tag{3.4.13}$$

Taking the derivative, using $\partial_t \left(\frac{1}{g(t)} \right) = - \left(\frac{1}{g(t)} \right)^2 \partial_t g(t)$, we find

$$\partial_t c(t) = -c(t)^2 \sum \left(-c_n E_n \exp(-E_n t) \int w_n(x) dx \right). \tag{3.4.14}$$

Comparing to the Equation (3.4.10), we see that

$$\lambda(t) = c(t) \sum \left(-c_n E_n \exp(-E_n t) \int w_n(x) dx \right) \tag{3.4.15}$$

$$= \sum \left(-c_n(t) \exp(-E_n t) \int E_n w_n(x) dx \right) \tag{3.4.16}$$

$$= \sum \left(c_n(t) \exp(-E_n t) \int (f(x) + D\partial_{xx}) w_n(x) dx \right), \tag{3.4.17}$$

where we used the defining equation for $E_n w_n$, Equation (3.4.5). Assuming that the w_n are smooth and vanish at infinity, it follows that $\int \partial_{xx} w_n(x) dx = 0$ and we find

$$\lambda(t) = \sum \left(c_n(t) \exp(-E_n t) \int f(x) w_n(x) dx \right) \tag{3.4.18}$$

$$= \int f(x) \left(\sum c_n(t) \exp(-E_n t) w_n(x) \right) dx \tag{3.4.19}$$

$$= \int f(x) u(x) dx, \tag{3.4.20}$$

which is the average fitness, as we have derived before.

To summarize, we need to solve the time-independent Schrödinger equation

$$-D\partial_{xx} w_n(x) + V(x) w_n(x) = E_n w_n(x) \tag{3.4.21}$$

with potential $V(x) = -f(x)$. If we can expand the initial condition $u(x,0) = u_0(x)$ into the eigenfunctions w_n, $u_0(x) = \sum_n c_n w_n(x)$, we can solve the equation by inserting the coefficients c_n in Equation (3.4.11). The time-independent Schrödinger equation is well-studied and we can use results from physics. However, the reproduction rate is positive, so we need to ensure that $f(x) > 0$, which means that $V(x) < 0$ for all x. Also, for a sensible interpretation, we should make sure that the density u is non-negative everywhere.

[18] To be able to interchange the order of an infinite sum and an integral, again we need some conditions on the convergence of the sum. The following derivation is just to show consistency, so we do not worry about this too much.

3.5 Solution for the hydrogen potential

We just found that if we can solve

$$(-f(x) - D\partial_{xx})w_n(x) = E_n w_n(x). \tag{3.5.1}$$

and get a basis of solutions w_n, we can write down the solution of the original birth-death-diffusion equation, Equation (3.4.1). Note that the derivation also holds in several dimensions without changing the arguments.

With physical constants, the time-independent Schrödinger equation is given by

$$\frac{-\hbar^2}{2m}\nabla^2\psi_n(x,t) + V(x,t)\psi_n(x,t) = E_n\psi_n(x,t). \tag{3.5.2}$$

We can identify $D = \frac{\hbar^2}{2m}$ and $f(x) = -V(x,t)$, so we see that Equation (3.5.1) and Equation (3.5.2) are the same equation. The hydrogen potential, $V(r,t) = -\frac{e^2}{4\pi\epsilon_0}\frac{1}{r}$ with $r = \sqrt{x_1^2 + x_2^2 + x_3^2}$, satisfies the condition $f(r) = \frac{e^2}{4\pi\epsilon_0}\frac{1}{r} > 0$ and with this potential, the time-independent Schrödinger equation has been solved exactly in terms of generalized Laguerre polynomials and spherical harmonics [Griffiths, 2005]. We can use these exact solutions to find an exact solution for our birth-death equation, Equation (3.4.1).

The exact solutions of the hydrogen atom are given by

$$\psi_{nlm}(r,\theta,\phi) = \sqrt{\left(\frac{2}{na}\right)^3 \frac{(n-l-1)!}{2n((n+l)!)^3}} \exp\left(\frac{-r}{na}\right)\left(\frac{2r}{na}\right)^l L_{n-l-1}^{2l+1}\left(\frac{2r}{na}\right) Y_l^m(\theta,\phi), \tag{3.5.3}$$

where

$$a = \frac{4\pi\epsilon_0\hbar^2}{me^2} = \frac{2D}{f(1)}, \tag{3.5.4}$$

$$L_{q-p}^p(x) = (-1)^p\left(\frac{\partial d}{\partial x}\right)^p\left[e^x\left(\frac{\partial d}{\partial x}\right)^q\left(e^{-x}x^q\right)\right] \tag{3.5.5}$$

are the generalized Laguerre polynomials and Y_l^m are the spherical harmonics. For convenience, to be able to effectively reduce the problem to a one dimensional problem, we only use spherically symmetric solutions, for which $l = 0$ and $m = 0$. In that case, the spherical harmonic $Y_l^m = Y_0^0$ is just a constant. The integer $n \geq 1$ determines the eigenvalue of the solution in Equation 3.5.2, $E_n = \frac{-1}{n^2}\frac{m}{2\hbar^2}\left(\frac{e^2}{4\pi\epsilon_0}\right)^2 = \frac{-1}{n^2}\frac{f(1)^2}{4D}$.

We choose $D = \frac{1}{2}$ and $f(r) = \frac{1}{r}$ so that $a = 1$[19]. The eigenvalues are then given by $E_n = \frac{-1}{2n^2}$. For simplicity, we build an initial density from $\psi_{100} \propto e^{-r/a} = e^{-r}$ and $\psi_{200} \propto \left(1 - \frac{r}{2a}\right)e^{-r/2a} = \left(1 - \frac{r}{2}\right)e^{-r/2}$. Note that ψ_{100} is the ground state, which is the only stable equilibrium state. Any randomly perturbed solution will tend to this ground state after some time. Specifically, we use the initial condition $u_0(r) = C\left(e^{-r} - \left(1 - \frac{r}{2}\right)e^{-r/2}\right)$, where C is a normalization constant. The system is entirely spherically symmetric, so we can just look at the r component. Note that the density of particles that have radius r is equal to the $p_{\text{rad}}(r,t) = \int_0^{2\pi}\int_0^{\pi} u(r,t)r^2\sin(\theta)d\theta d\phi = 4\pi r^2 u(r,t)$, which we will call the radial density. The radial density p_{rad} is most easily compared to a micro model population, because we can just look at the distribution of r values in the population.

[19]The resulting discussion is still completely general, because for any $f(r) = \frac{c}{r}$, we can choose a unit of time such that $f(r) \to \frac{1}{r}$. Also, we can choose a unit of length such that $D = \frac{1}{2}$. Note that there are two parameters and two dimensions, so that by the Buckingham π theorem, we can get rid of all the parameters through non-dimensionalization. Also for the micro-model it can be checked that choosing different parameters merely scales the variables and does not change the dynamics.

From Section 3.4, we know that the solution of the equation

$$\partial_t u(r,\theta,\phi,t) = \left[\frac{1}{r} - \int \frac{u(r,\theta,\phi,t)}{r} + \frac{1}{2}\nabla^2(r,\theta,\phi)\right] u(r,\theta,\phi,t). \tag{3.5.6}$$

with initial condition $u(r,\theta,\phi,0) = u_0(r)$ is given by

$$u(r,\theta,\phi,t) = c(t)\left(e^{-r}e^{-E_1 t} - \left(1-\frac{r}{2}\right)e^{-r/2}e^{-E_2 t}\right) = c(t)\left(e^{-r}e^{t/2} - \left(1-\frac{r}{2}\right)e^{-r/2}e^{t/8}\right), \tag{3.5.7}$$

where $c(t)$ is the time-dependent normalization constant that ensures

$$\int_0^\infty \int_0^{2\pi} \int_0^\pi u(r,\theta,\phi,t) r^2 \sin(\theta) d\theta d\phi dr = 1. \tag{3.5.8}$$

Explicit integration shows that $c(t) = \frac{1}{8\pi(e^{t/2}+16e^{t/8})}$, so that the solution is

$$u(r,\theta,\phi,t) = \frac{1}{8\pi(e^{t/2}+16e^{t/8})}\left[e^{-r}e^{t/2} - \left(1-\frac{r}{2}\right)e^{-r/2}e^{t/8}\right]. \tag{3.5.9}$$

Plugging this expression in Equation (3.5.6) shows that it is indeed the normalized exact solution with initial condition $u_0(r)$.

Multiplying by $4\pi r^2$ to get the radial density, we find

$$p_{\text{rad}}(r,t) = \frac{r^2}{2(e^{t/2}+16e^{t/8})}\left[e^{-r}e^{t/2} - \left(1-\frac{r}{2}\right)e^{-r/2}e^{t/8}\right]. \tag{3.5.10}$$

and its evolution[20] is shown in Figure 3.7. The solution is attracted to the equilibrium state. In Section 4.2.2, we compare the solution in Equation (3.5.10) to the evolution of the micro-model with the same initial density.

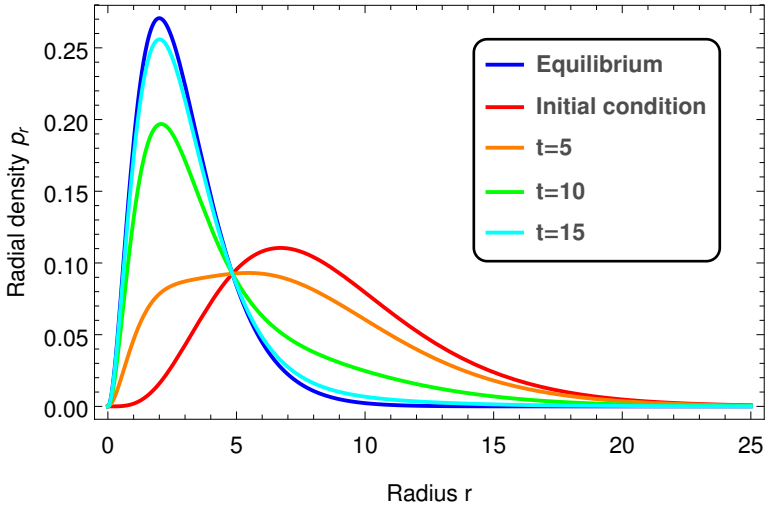

Figure 3.7 – Evolution of the exact solution of the radial density from the initial condition (red) to the equilibrium state (blue).

[20] Note that there is a radius r for which the density does not change in time. By inspecting the solution in Equation (3.5.9), we see that this happens when $-(1-\frac{r}{2})e^{-r/2} = 16e^{-r}$, because in that case the time dependence cancels out. The corresponding value for r is approximately $r \approx 4.84$, which is in agreement with Figure 3.7.

3.6 Noise on the PDE

We derived a partial differential equation for the average behaviour. Unfortunately, this description does not contain the clustering effect yet. The movement of the population as a cluster is not deterministic but random. Therefore, we need to add a stochastic part to the partial differential equation. In this section, we derive the form of this stochastic part by looking at the micro model more closely. The derivations in this section will not be very rigorous, but they serve to support the chosen form of the noise, which is a modelling choice.

Consider some region V in $\mathbb{R}^d$ and denote by $N_V(t)$ the number of particles in V at time t. Births and deaths happen according to a Poisson process with rates $\bar{f}(V)$[21] and $\lambda(t)$ per particle, so the total rates of births and deaths are given by $N_V(t)\bar{f}(V)$ and $N_V(t)\lambda(t)$. If these rates are constant through time, the number of births B and deaths D in time Δt follows a Poisson distribution with parameter $N_V(t)\bar{f}(V)\Delta t$ and $N_V(t)\lambda(t)\Delta t$. For a Poisson distribution with parameter γ, we have that both the expected value and the variance are equal to γ. Therefore, the expectations and variances are

$$\mathrm{E}(B) = \mathrm{Var}(B) = N_V(t)\bar{f}(V)\Delta t, \tag{3.6.1}$$

and

$$\mathrm{E}(D) = \mathrm{Var}(D) = N_V(t)\lambda(t)\Delta t. \tag{3.6.2}$$

For a moment, we neglect diffusion[22] so that the change in $N_V(t)$ is only due to birth and deaths:

$$N_V(t+\Delta t) = N_V(t) + B - D; \tag{3.6.3}$$

Unfortunately, although the sum of two Poisson variables is Poisson, the difference of two Poisson variables is not Poisson[23]. However, we can still compute the expectation of $\Delta N_V = N_V(t+\Delta t) - N_V(t)$, which is given by

$$\mathrm{E}(\Delta N_V) = \mathrm{E}(B) - \mathrm{E}(D) = N_V(t)(\bar{f}(V) - \lambda(t))\Delta t, \tag{3.6.4}$$

and if we assume that B and D are independent[24], the variance is given by

$$\mathrm{Var}(\Delta N_V) = \mathrm{Var}(B) + \mathrm{Var}(D) = N_V(t)(\bar{f}(V) + \lambda(t))\Delta t. \tag{3.6.5}$$

A Poisson variable with a large rate is approximately normal. The difference of two normal variables is again normal. Therefore, ΔN_V, which is the difference of two Poisson variables, is approximately normal when $N_V(t)\bar{f}(V)\Delta t$ and $N_V(t)\lambda(t)\Delta t$ are large. That is, when $N_V(t)\bar{f}(V)\Delta t \gg 1$ and $N_V(t)\lambda(t)\Delta t \gg 1$. The assumptions we have made up to now are that the rates are large and nearly constant over the time interval $(t, t+\Delta t)$.

The Wiener process $W(t)$ satisfies that $\Delta W = W(t+\Delta t) - W(t)$ is normally distributed with variance $\sigma\Delta t$. Similarly, we just showed that the process for $N_V(t)$ satisfies that ΔN_V is approximately normally distributed with variance proportional to Δt. So we see that the fluctuations for $N_V(t)$ are approximately described by a Wiener process when the assumptions hold.

[21] We assume that the fitness does not vary much over the region V and take the average $\bar{f}(V)$ of f over the region V.

[22] We could also take the view that a particle that enters the region by diffusion is a birth and a particle that leaves the region is a death. This way, we could include diffusion into this discussion, but we would need rates at which particles enter and leave the region V.

[23] For example, the difference can be negative, while a Poisson variable can not.

[24] Which is true by approximation if the number of particles in V is small compared to the population size

Splitting off the expectation and taking the limit, combining Equation (3.6.4) and Equation (3.6.5), we find as an approximation that

$$dN_V = N_V(t)(f-\lambda)dt + \sqrt{(f+\lambda)N_V(t)}dW. \tag{3.6.6}$$

The first term is equal to the first term of the birth-death-diffusion equation. The second term is an approximation of the fluctuations from the average behaviour. Note that the standard deviation of the fluctuations is proportional to $\sqrt{N_V(t)}$.

The previous discussion motivates the following form of the stochastic birth-death-diffusion equation

$$\partial_t u(x,t) = \Big(f(x) - \lambda(t) + D\nabla^2\Big)u(x,t) + c\sqrt{u(x,t)}\eta(x,t), \tag{3.6.7}$$

where $\eta(x,t)$ is uncorrelated Gaussian noise[25] and c is a constant that determines the strength of the noise. Note that c should actually depend on f, λ and D, which makes it a function of x and t. As a simplification, we take it to be constant and hope that the important features are still present.

Now we consider the justification of the assumptions that were made. The first assumption was that the rates $N_V(t)\bar{f}(V)$ and $N_V(t)\lambda(t)$ are constant during the time interval $(t, t+\Delta t)$. The fitness $\bar{f}(V)$ does not change in time and the death rate $\lambda(t)$ is nearly constant if Δt is small. The number of particles $N_V(t)$ changes due to births and deaths, but the assumption is approximately correct when $\Delta N_V \ll N_V(t)$. If $N_V(t)$ is small, the discrete nature of $N_V(t)$ could result in violation of this condition. For example, when $N_V(t) = 1$, a single birth or death causes a large relative change in $N_V(t)$. Even worse, when $N_V(t) \to 0$ at some instance of time, it stays zero, because births are no longer possible. These arguments show that we need to be careful when $N_V(t)$ is small. We need to take into account the discrete nature of $N_V(t)$ and the possibility of extinction. If $N_V(t)$ is large, we can safely assume that $\Delta N_V \ll N_V(t)$ holds when both

$$\mathrm{E}(\Delta N_V) = N_V(t)(\bar{f}(V) - \lambda(t))\Delta t \ll N_V(t), \tag{3.6.8}$$

which is equivalent to $(\bar{f}(V) - \lambda(t))\Delta t \ll 1$, and[26]

$$\sqrt{\mathrm{Var}(\Delta N_V)} = \sqrt{N_V(t)(\bar{f}(V) + \lambda(t))\Delta t} \ll N_V(t). \tag{3.6.9}$$

We could take Δt small enough such that both conditions hold, so it is valid in the limit of $\Delta t \to 0$.

To show that ΔN_V is normally distributed, we made the second assumption that $N_V(t)\bar{f}(V)\Delta t$ and $N_V(t)\lambda(t)\Delta t$ are large. This conflicts with taking the limit of $\Delta t \to 0$. The change ΔN_V in a very small time Δt is not normally distributed, as is already clear by considering the discrete nature of ΔN_V. However, we already assumed that $N_V(t)$ is large, so $N_V(t)\bar{f}(V)\Delta t$ and $N_V(t)\lambda(t)\Delta t$ are in fact large when $\bar{f}(V)$, $\lambda(t)$ and Δt are not too small. The rates per particle $\bar{f}(V)$ and $\lambda(t)$ are not necessarily small, so in most cases, when Δt is not too small, ΔN_V is approximately normally distributed.
This shows that although taking normal increments at very short times Δt is incorrect, it results in the correct behaviour at somewhat larger time-scales, where the discrete nature of ΔN_V is less important.

[25] The noise $\eta(x,t)$ is a normal random variable with variance equal to 1 for every x and t such that $\mathrm{E}(\eta(x,t)\eta(\tilde{x},\tilde{t})) = 0$ if $x \neq \tilde{x}$ or $t \neq \tilde{t}$. This is chosen for simplicity. Note that this form of the noise can destroy continuity of the solution, although diffusion smooths the solution again. Also note that the noise does not conserve the population size, so $\lambda(t)$ should be adapted so that the population size is constant again.

[26] The standard deviation $\sqrt{\mathrm{Var}(\Delta N_V)}$ represents the spread of ΔN_V. This should be small compared to $N_V(t)$, because we want that all likely realisations of ΔN_V are small compared to $N_V(t)$.

As we already discussed in Section 1.5.1, extinction is the main reason for clustering of the population. If $N_V(t) \to 0$ at some instance of time, the process $N_V(t)$ is absorbed into the state $N_V(t) = 0$ until new particles are introduced into V by diffusion. The approximation by a Wiener process does not hold for small $N_V(t)$ and this approximation does not grasp the effects of extinction.

Lawson and Jensen [2008] use field theoretic techniques based on [Täuber, Howard, and Vollmayr-Lee, 2005] to derive a stochastic partial differential equation for a flat fitness landscape in order to explain clustering. For a system that is very similar to ours, they report[27]

$$\partial_t u(x,t) = D\nabla^2 u(x,t) + c\sqrt{u(x,t)}\eta(x,t), \tag{3.6.10}$$

which is the same stochastic partial differential equation as Equation (3.6.7), but in a flat fitness landscape. Correctly simulating such a stochastic partial differential equation numerically is not trivial. They refer to an approach by Moro [2004], which incorporates the Poissonian nature of the fluctuations at small densities in the numerical scheme to solve the stochastic partial differential equation. In their simulations, Lawson and Jensen [2008] found clustering qualitatively, but the width of the clusters are not correctly represented.
In Appendix A.4, we propose a deterministic term to account for extinction, which might improve the agreement between the micro-model simulations and the stochastic partial differential equation.

[27] Here, $\eta(x,t)$ is also defined as uncorrelated Gaussian noise. It is mentioned that this is done for simplicity.

Chapter 4

Comparison to the micro-model

In the following, we compare simulations of the micro-model to predictions of the density description.

4.1 Translation between discrete populations and continuous densities

In the micro-model, the population is given by a set of positions $\{x_i\}$. In the continuous description, the population is given by a density function $u(x)$ defined for every $x \in \mathbb{R}^d$. In order to compare results, we need to make a translation between these two descriptions of a population.

The most obvious way to make a translation from a density to a population is to sample from the density. That is, given a scaled density u such that[1]

$$\int_{\mathbb{R}^d} u = 1, \tag{4.1.1}$$

we can view u as a probability density on $\mathbb{R}^d$ and take a sample of M particles from this probability density[2].
Note that there is not a single discrete population $\{x_i\}$ that exactly corresponds to the density u. Sampling from u gives a random population and in principle the sampled population can be any set of points from regions where $u > 0$. However, it is more probable to find populations that resemble the density u, in a sense we will explain later. Larger samples approximate u better, but the sample size M is dictated by the micro-model. Instead, we can use an ensemble of N_e populations of size M. The larger the ensemble, the better the average population will approximate u.

[1] In the normal definition of a density, $\int_{\mathbb{R}^d} u = M$, so to scale it in this way, we divide by M.
[2] We can use the rejection method to sample from any distribution.

4.1.1 Bin counting

To translate a discrete population to a density, there are several possibilities. The first well-known method we discuss is based on the histogram[3]. In order to make a histogram, we divide the space into bins b of volume V_b and count the number of particles N_b in every bin. The density is equal to the number of particles per volume, so $\frac{N_b}{V_b}$ gives the average density in the bin. To scale it in the same way as u, we should divide by M so that the value is given by $u^* = \frac{N_b}{V_b M}$.

If these translations are proper, we should be able to turn a density u into a population $\{x_i\}$ and turn this population into a density u^* similar to u again. The density obtained by bin counting, u^*, is a discontinuous step function and it is based on a population that is random, so u^* is random. Now what are the statistics of $u^*(b)$ and how are these related to u?

The value of $u^*(b)$ is determined by the number of particles N_b in the region V_b. The number of particles N_b in V_b is determined by the sampling procedure from the distribution u. Specifically, the probability that a particle is sampled in the region V_b is given by $p = \int_{V_b} u$. The number of sampled particles is equal to M so there are $n = M$ independent trials to sample the particle in the region V_b. This implies that N_b is a binomial variable with $n = M$ trials and $p = \int_{V_b} u$ probability of success, which has expectation value and variance equal to[4]

$$\mathrm{E}(N_b) = np = M \int_{V_b} u, \tag{4.1.2}$$

$$\mathrm{Var}(N_b) = np(1-p) = M \left(\int_{V_b} u \right) \left(1 - \int_{V_b} u \right). \tag{4.1.3}$$

The expected value of the translated density u^* in bin b is then given by

$$\mathrm{E}(u^*(b)) = \mathrm{E}\left(\frac{N_b}{V_b M} \right) = \frac{\mathrm{E}(N_b)}{V_b M} = \frac{\int_{V_b} u}{V_b} = \bar{u}, \tag{4.1.4}$$

where $\bar{u}$ is the average value of u in the region V_b. This means that u^* is an unbiased approximation of u within the constraint of the discontinuous histogram.

More interesting is the variance, given by

$$\mathrm{Var}(u^*(b)) = \mathrm{Var}\left(\frac{N_b}{V_b M} \right) = \frac{\mathrm{Var}(N_b)}{V_b^2 M^2} = \frac{(\int_{V_b} u)(1 - (\int_{V_b} u))}{V_b^2 M} = \frac{\bar{u}}{M} \left(\frac{1}{V_b} - \bar{u} \right) \approx \frac{\bar{u}}{M V_b}, \tag{4.1.5}$$

where the last approximation holds when the probability $p = \int_{V_b} u$ is small, which can be accomplished by using a small bin-size V_b. We see that the approximation indeed becomes better with increasing M. The variance increases with increasing $\bar{u}$, but the relative standard deviation $\frac{1}{\bar{u}} \sqrt{\frac{\bar{u}}{M V_b}}$ decreases with increasing $\bar{u}$. When V_b is too large, the discrete nature of the histogram results in a bad approximation that gives little information[5]. When V_b is too small, the variance will increase and the resulting histogram will be very erratic.

There are two messages here. Firstly, it is important to choose proper bin-sizes. Secondly, we should keep in mind that there is variance associated with the translation of discrete populations and densities, because the discrete population is sampled randomly.

[3] Usually, a "histogram" refers to the graphical representation of data by bin counting. For notational convenience, we will use the word histogram for the function obtained by bin counting.

[4] See, for example, [Pitman, 1993].

[5] Note that in general, the approximation in Equation (4.1.5) does not hold in this regime.

4.1.2 Kernel density estimation

Another approach to make a density out of a discrete population $\{x_i\}$ is by using kernel density estimation. We take some non-negative function $h(x)$, symmetric around the origin, that integrates to $\int h(x)dx = 1$, for example the probability density of the normal distribution. We then use as a density approximation

$$u^*(x) = \frac{1}{M}\sum_i h(x - x_i). \tag{4.1.6}$$

In order for u^* to approximate u, the function h should be chosen properly[6]. This approach results in the same picture as with the histogram. The width of h should be chosen properly and we should expect variance associated to the sampling procedure. Density estimation is a topic from statistics and there are other methods to estimate a density, for example by estimating the characteristic function. In all these methods, there is some way of smoothing the data.

4.1.3 Empirical distribution function

The purest approach to process data into a density is by using $h(x) = \delta(x)$, a delta distribution that satisfies $\delta(x) = 0$ for $x \neq 0$ and $\int_{V_\epsilon} \delta(x)dx = 1$ for any V_ϵ containing the origin. This results in an estimated density given by

$$u^*(x) = \frac{1}{M}\sum_i \delta(x - x_i). \tag{4.1.7}$$

This is the purest approach because we do not need artificial smoothing and we use the complete information in the population. However, this density estimation is not always useful, because it is still hard to compare to a normal continuous density u.

If we look at data points from a one-dimensional space, we can compare densities properly by looking at the cumulative distribution function F instead of the probability density function p. The cumulative distribution function $F(x) = \int_{-\infty}^{x} p(\tilde{x})d\tilde{x}$ gives the probability to sample a particle with a value smaller than x. Given a discrete data set $\{x_i\}$, the empirical distribution function F_e is defined to be

$$F_e(x) = \frac{1}{M}\sum_{i=1}^{M} I(x_i \leq x) \tag{4.1.8}$$

where $I(x_i \leq x)$ is the indicator function returning 1 if $x_i \leq x$ and 0 if $x_i > x$[7]. When sampling from a probability distribution with cumulative distribution function F, the empirical distribution F_e of the sample tends to F[8]. Therefore, we can compare F_e from the population to F from the density. It is possible to devise statistical tests based on the empirical distribution function. The tests determine whether the hypothesis that the population $\{x_i\}$ was taken from the density u is accepted or rejected. The main virtue of the empirical distribution function is that it takes all information from the population into account. Note that its probability distribution function is given by the δ distributions of Equation (4.1.7), so that the empirical distribution function unfortunately does not give a method to compute appropriate continuous densities.

[6] We could take the view that a histogram fits in this description, with h a non-central block function.
[7] Also known as the Heaviside step function.
[8] Specifically, the Glivenko-Cantelli theorem says that F_e converges uniformly to F when $M \to \infty$. That is, $\sup_x(F(x) - F_e(x)) \to 0$.

4.2 Comparison of the micro-model and the PDE

We now compare the predictions in Section 3.5 to the actual results from the micro-model. We initialize the population by sampling from the initial density

$$u_0(r,\theta,\phi) = \frac{1}{136\pi}\left(e^{-r} - \left(1-\frac{r}{2}\right)e^{-r/2}\right), \tag{4.2.1}$$

following the procedure from the previous section. We use the same values $D = \frac{1}{2}$ and $f(r,\theta,\phi) = \frac{1}{r}$, so that we use the same units of time and space. The system is then evolved according to the micro-process defined in Section 3.1.

4.2.1 Equilibrium

Before we use our specific initial condition, let us first look at equilibrium results. We simulate a system of $M = 1000$ particles for $T = 500,000$ and measure the state of the population every 10 units of time. We discard the first 1000 population measurements[9] so that the remaining 49000 measurements are done in equilibrium and do not depend on the initial condition. To measure the population state, we record the radius r of every population member, so a measurement looks like $\{r_i\}_{1\leq i\leq M}$. Although we lose information on angular fluctuations, this is convenient because the problem is spherically symmetric and by looking at the r values we reduce the problem to one dimension.

To translate discrete populations to densities, we use bin counting to obtain a histogram as described in Section 4.1.1. Our measurements consist of r values, so it is most easily compared to the predicted distribution of r values, which is the radial density $p_{\text{rad}}(r) = 4\pi r^2 u(r)$. In Figure 4.1, we see that the predicted $p_{\text{rad}} \propto r^2 e^{-r}$ is in perfect agreement with the measured distribution of r values. For completeness, we have also transformed the measured radial densities p^*_{rad} into a measured densities u^* by multiplying by $4\pi r^2$.

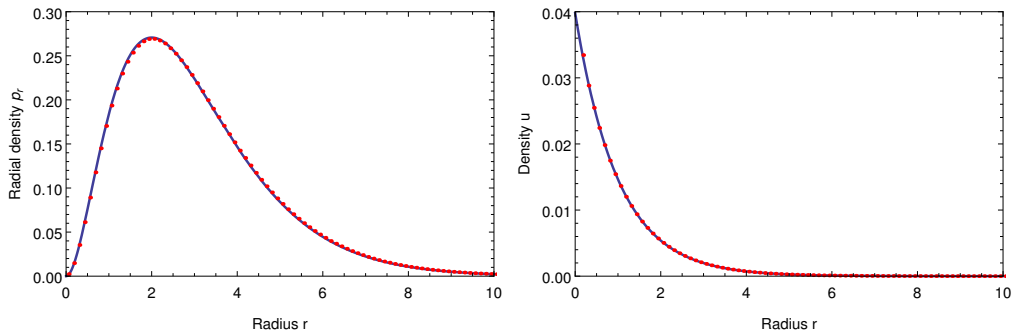

Figure 4.1 – The average measured equilibrium radial density (left) and density (right) as a function of radius r given by the red dots compared to the exact equilibrium distribution of the birth-death-diffusion equation given by the blue line.

4.2.2 Non-equilibrium

We have just seen that the equilibrium state is predicted correctly. However, the birth-death-diffusion equation from Equation (3.4.1) actually describes time-dependent dynamics. We now investigate if the equation predicts the non-equilibrium dynamics properly. Specifically, we compare the evolution from the initial condition in Equation (4.2.1) by the micro-model and the birth-death-diffusion equation. For the birth-death-diffusion equation, we have derived the exact solution $u(r,t)$, which is given by Equation (3.5.9).

[9] It turns out that the system is close to equilibrium after 30 units of time, so we are very safe here.

For the micro-model, we take an ensemble of $N_e = 5000$ populations of $M = 1000$ particles. These populations are propagated by the micro-model as described above. We compare the ensemble average of these simulations to the exact solution of the birth-death-diffusion equation in Figure 4.2. We see that also for time-dependent dynamics, the birth-death-diffusion equation describes the average behaviour very well.

Only at $t = 10$ and $t = 15$, we see that the micro-model lags behind a bit on the birth-death-diffusion equation. This discrepancy is eliminated when the system moves closer to equilibrium. The difference appears where the change is most rapid. An explanation for this is that in the birth-death-diffusion equation, births and deaths happen continuously in time, while in the micro-model, births and deaths happen at discrete instances of time. This difference is most apparent when the change in time is most rapid[10].

We also applied the other methods of Section 4.1. Kernel density estimation with normal distributions gives approximately the same results. The exact solution for the radial distribution function p_{rad} of the birth-death-diffusion equation can be integrated to find $F(r) = \int_0^r p_{\text{rad}}(\tilde{r})d\tilde{r}$, which can be compared to the the average empirical distribution F_e from the ensemble. Again, the results indicate that the birth-death-diffusion equation describes the micro-model very well.

By simulating the same set-up with a different population size M, we found that the birth-death-diffusion equation approximates the evolution of the system best when we use many particles. For example, with a smaller population of $M = 100$, the discrepancy found around $t = 10$ in Figure 4.2 gets larger. This is to be expected, because the more particles, the better the continuous approximation of the births and deaths. When the population size M is taken smaller and smaller, the approximation by the birth-death-diffusion equation should evidently break down at some point.

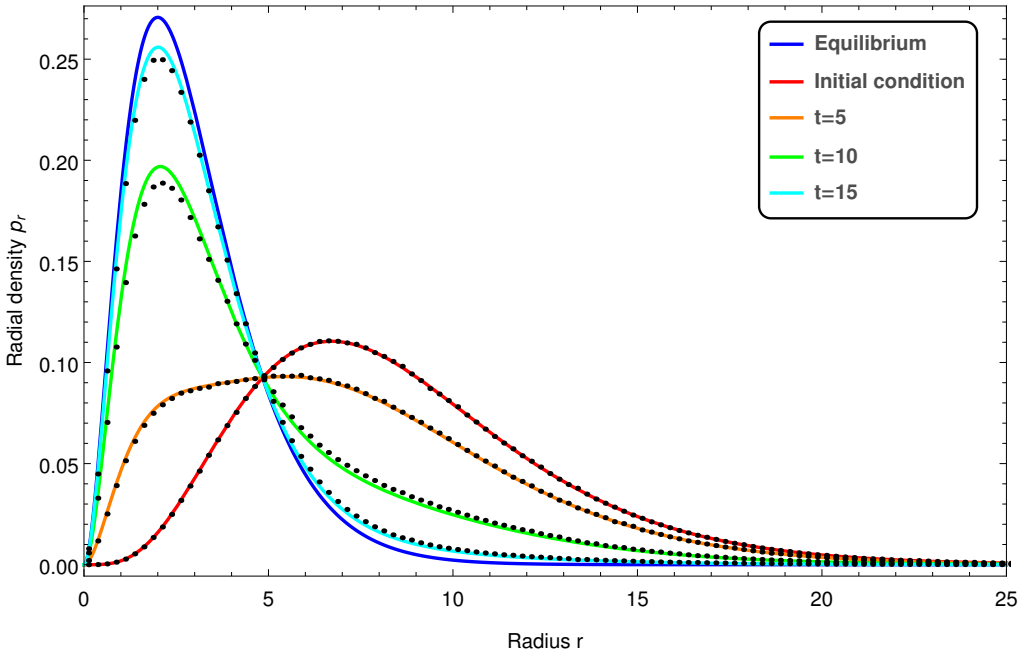

Figure 4.2 – Results from the micro-model added to Figure 3.7. The average measured radial density as a function of radius r for different times t (given by black dots) compared to the exact solution of the birth-death-diffusion equation given by the solid lines. It is clear what measurements belong to what time, because the descriptions agree very well. The measurements in equilibrium are not shown here, but in Figure 4.1.

[10]For comparison. When discretizing a differential equation to solve it numerically, errors are introduced most where the system changes fastest.

4.3 Fluctuations

We have seen that the average behaviour is described very well by the birth-death-diffusion equation. However, the actual populations are different from the average population. We could take the view that this is due to fluctuations from the mean. In order to make further progress, we need to quantify in what way the populations differ from the average population.

By a histogram, a discrete population $G = \{x_i\}_{1 \leq i \leq M}$ is turned into "measurements" of the radial density $p_{\text{rad}}(r)$ and the density $u(r)$, which we will denote by $p^G_{\text{rad}}(r)$ and $u^G(r)$[11]. The procedure is illustrated in Figure 4.3. The measured $p^G_{\text{rad}}(r)$ and $u^G(r)$ are discontinuous step-functions of r, because the histogram is a discontinuous step-function. The bins of the histogram are denoted by $b_j = (r_j, r_j + |b|)$ where the bin-size $|b|$ is some constant. The measured radial density $p^G_{\text{rad}}(r)$ is given by

$$p^G_{\text{rad}}(r) = \frac{N^G_{b(r)}}{M|b|}, \tag{4.3.1}$$

where $N^G_{b(r)}$ is the number of particles from G in the bin $b(r)$ corresponding to r, that is $r \in b(r)$[12]. The measured density $u^G(r)$ is given by

$$u^G(r) = 4\pi r^2 p^G_{\text{rad}}(r). \tag{4.3.2}$$

Figure 4.1 and Figure 4.2 were made by averaging p^G_{rad} and u^G over many populations. To quantify the strength of the fluctuations from the mean, we can look at the variance of $p^G_{\text{rad}}(r)$ and $u^G(r)$ in the ensemble of measurements.

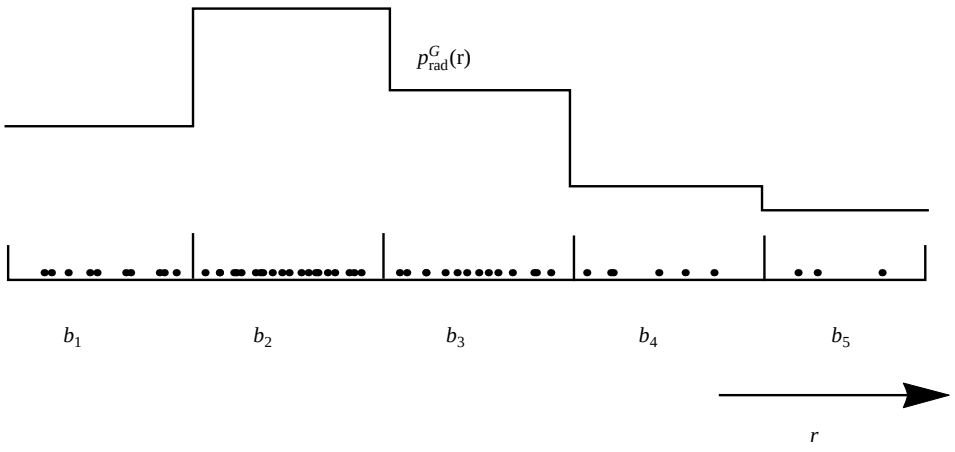

Figure 4.3 – Illustration of the measurement of $p^G_{\text{rad}}(r)$ by means of a histogram. The black dots represent the population G. The measured radial density $p^G_{\text{rad}}(r)$ is created by simply counting the number of members from G in the bin that belongs to r.

[11] Note that $p^G_{\text{rad}}(r)$ and $u^G(r)$ are functions of the random population G, as indicated by the superscript.

[12] Note that we consider $p^G_{\text{rad}}(r)$ as a function of r, even though we can represent it by a discrete set of values at the bins.

A major complication with this idea is that we already expect to find non-zero variance because of the sampling procedure, as described in Section 4.1.1. Even if the birth-death-diffusion equation would exactly determine the evolution of the micro-model, so that there is only one appropriate density $u(r)$, we should still expect to see variance in the measured $p^G_{\text{rad}}(r)$ and $u^G(r)$. The derived variance based on the binomial distribution gives a lower bound against which the measured variance can be compared.

Ideally, we should choose our system parameters[13] such that this lower bound is negligible compared to the measured variance, but we found that this is not possible. Recall that the variance in the histogram could be reduced by increasing the number of particles. However, the more particles we use, the more the system behaves like the average system. As we found in Section 1.5, most interesting behaviour is due to the finite size of the population. By increasing the population size to decrease the lower bound of the variance, the finite size effects we are looking for also decrease in size.
This leaves only the bin-size to play with, but the range of decent bin-sizes is only small and there is not much to gain. Therefore, we should try other approaches.

Another complication that arises when measuring fluctuations is that these fluctuations spread over space due to diffusion. If there are more than an average number of particles around x at t, these extra particles diffuse towards regions nearby x. This results in highly correlated bins. Also, it is not directly clear where the extra particles came from. Imagine the case where there is one main source of fluctuations in the system, say at position x_f, at which there are large fluctuations of artificial births. The measured variance at positions x close to x_f will be large because of fluctuations in x_f. These extra fluctuations have nothing to do with the dynamics at x itself.

[13] The free system parameters are the population size M, the ensemble size N_e and the bin-size $|b|$. The ensemble size N_e just leads to more accurate results, so cannot be used.

4.3.1 Covariance

We will now look at another approach to determine the fluctuations in the density based on studying the covariance of $p^G_{\text{rad}}(r_1)$ and $p^G_{\text{rad}}(r_2)$. The covariance is defined by

$$\text{Cov}\left(p^G_{\text{rad}}(r_1), p^G_{\text{rad}}(r_2)\right) = \text{E}\left[\left(p^G_{\text{rad}}(r_1) - \text{E}(p^G_{\text{rad}}(r_1))\right)\left(p^G_{\text{rad}}(r_2) - \text{E}(p^G_{\text{rad}}(r_2))\right)\right]. \tag{4.3.3}$$

Note that $\text{Cov}(p^G_{\text{rad}}(r), p^G_{\text{rad}}(r)) = \text{Var}(p^G_{\text{rad}}(r))$. If $\text{Cov}\left(p^G_{\text{rad}}(r_1), p^G_{\text{rad}}(r_2)\right)$ would be a continuous function, taking the limit of $r_2 \to r_1$ would result in $\text{Var}(p^G_{\text{rad}}(r_1))$.
The covariance is defined in terms of expected values. A finite average of continuous functions $g_i(x)$, $\frac{1}{N}\sum_{i=1}^{N} g_i(x)$ is continuous. The expected value is approximated by a finite average, so intuitively, we would expect that the expected value of a continuous random function is continuous[14]. If p^G_{rad} would be a continuous function[15], by the previous arguments, we would expect that the covariance $\text{Cov}\left(p^G_{\text{rad}}(r_1), p^G_{\text{rad}}(r_2)\right)$ is a continuous function of r_1 and r_2.

The measured covariance between $p^G_{\text{rad}}(r_1)$ and $p^G_{\text{rad}}(r_2)$ is defined to be

$$\text{Cov}_m\left(p^G_{\text{rad}}(r_1), p^G_{\text{rad}}(r_2)\right) = \frac{1}{N_m - 1}\sum_{i=1}^{N_m}\left(p^i_{\text{rad}}(r_1) - \text{E}_m(p_{\text{rad}}(r_1))\right)\left(p^i_{\text{rad}}(r_2) - \text{E}_m(p_{\text{rad}}(r_2))\right) \tag{4.3.4}$$

where N_m is the number of measured populations and $\text{E}_m(p_{\text{rad}}(r_1)) = \frac{1}{N_m}\sum_{i=1}^{N_m} p^i_{\text{rad}}(r_1)$ is the measured mean of $p_{\text{rad}}(r_1)$. The measured mean $\text{E}_m(p_{\text{rad}}(r_2))$ is defined equivalently. The values of $p^i_{\text{rad}}(r)$ are obtained as defined in Equation (4.3.1).

The measured covariance in equilibrium[16] is shown in Figure 4.4. Here, we use the population measurements in equilibrium from Section 4.2.1. Recall that the diagonal, where $r_1 = r_2$, shows the measured variance $\text{Var}(p^G_{\text{rad}}(r))$. We see that the measured covariance is indeed continuous, except at the diagonal, where the covariance becomes the variance. Also, we see that the variance at the diagonal depends on the bin-size[17], which is artificially chosen, while the covariance elsewhere does not. This shows that we cannot trust direct measurements of the variance.

Now we come back to the complication that we discussed before, the lower bound of the variance. Figure 4.5 shows that the measured variance is only slightly higher than the lower bound for the variance[18]. Most of the measured variance is simply due to the sampling procedure. This explains why the measured variance depends on the bin-size.
Before we continue, let us take a look again at the lower bound of the variance. As described before, we consider sampling from a deterministic density u_0 and translating the sampled population $\{x_i\}$ back to a density u^* again. The measured density is given by

$$u^*(r) = \bar{u_0}(r) + \eta, \tag{4.3.5}$$

[14] Counter-examples exist. However, the necessity of producing a complicated counter-example shows that usually it does hold that the expectation is continuous. If $p^G_{\text{rad}}(r)$ admits a probability density for p_{rad} at every r, $f(r, p_{\text{rad}})$, the expectation is given by $\int p_{\text{rad}} f(r, p_{\text{rad}}) dp_{\text{rad}}$. To be continuous, we should have that $\lim_{r\to r0}\int p_{\text{rad}} f(r, p_{\text{rad}}) dp_{\text{rad}} = \int p_{\text{rad}} f(r_0, p_{\text{rad}}) dp_{\text{rad}}$, which is true when f is continuous in r and when the integral converges uniformly.

[15] Note that this discussion is not at all precise, because there is no such thing as an exact function p^G_{rad} for which G is the representation. However, we do try to describe the micro-model in such a way, which is intrinsically not an exact description.

[16] In equilibrium, the extra variance is due to a balance between fluctuations away from the equilibrium and the average drive towards the equilibrium. The system can be viewed as a dissipative system and it might be possible to apply fluctuation-dissipation theorems.

[17] The variance decreases for a larger bin-size, as expected from Equation (4.1.5).

[18] If we found exactly the lower bound, this would imply that there is a single appropriate density. This is not the case because the micro-model does not behave deterministically, but random and introduces some more variability which is reflected in the higher variance.

where $\bar{u_0}$ is the average of u_0 over the bin b with $r \in b$ and η is random noise constructed from a binomial distribution. The average is split off, so that $\mathrm{E}(\eta) = 0$. The variance of the measured density is given by

$$\mathrm{Var}(u^*(r)) = \mathrm{Var}(\bar{u_0}(r) + \eta) = \mathrm{Var}(\eta). \tag{4.3.6}$$

Now consider the case that u_0 is not a deterministic density, but a random density U_0[19]. According to Equation (4.1.5), the size of the variance η depends on $\bar{U_0}$. However, the actual realisation does not. Therefore, the realisation of U_0 and the realisation of the noise η are approximately independent. Because $\mathrm{Var}(X + Y) = \mathrm{Var}(X) + \mathrm{Var}(Y)$ if X and Y are independent, we have that

$$\mathrm{Var}(u^*(r)) = \mathrm{Var}(\bar{U_0}(r) + \eta) \approx \mathrm{Var}(\bar{U_0}(r)) + \mathrm{Var}(\eta). \tag{4.3.7}$$

This shows that we can obtain an approximation for $\mathrm{Var}(\bar{U_0}(r))$ by subtracting $\mathrm{Var}(\eta)$[20] from the measured variance, which we will call the modified variance. Translated to the radial density, the modified variance is given by

$$\mathrm{Var}_{\mathrm{mod}}\left(p^G_{\mathrm{rad}}(r)\right) = \mathrm{Var}\left(p^G_{\mathrm{rad}}(r)\right) - \mathrm{Var}_{bin}\left(p^G_{\mathrm{rad}}(r)\right) \tag{4.3.8}$$

where $\mathrm{Var}\left(p^G_{\mathrm{rad}}(r)\right)$ is the measured variance and $\mathrm{Var}_{bin}\left(p^G_{\mathrm{rad}}(r)\right)$ is the lower bound of the variance as determined by the binomial variance of the histogram counts from a deterministic density, as discussed in Section 4.1.1.

In Figure 4.4, the modified variance is shown in black. We see that the modified variance fits in the covariance landscape and seems to make the covariance continuous everywhere. In Figure 4.5, the modified variance and the neighbour covariance are shown, where the neighbour covariance is defined to be $\mathrm{Cov}_m\left(p^G_{\mathrm{rad}}(r), p^G_{\mathrm{rad}}(r + |b|)\right)$. That is, the neighbour covariance is calculated using neighbouring bins and it represents the best approximation of the covariance to the variance[21]. Note that it is the off-diagonal in Figure 4.4.
We see that the modified variance and the neighbour covariance are approximately the same and that they do not depend on the bin-size used. As expected, the modified variance is larger than the neighbour covariance, because the neighbour covariance underestimates the variance, while the modified variance overestimates the variance[22].

The previous discussion makes it plausible that the modified variance and the neighbour covariance approximate the actual variance in the density distribution of $p^G_{\mathrm{rad}}(r)$, while direct measurements of the variance cannot be trusted because the results depend on the bin-size used. Because the procedure for the modified variance is not very well substantiated, we will focus on obtaining an approximation of the variance in $p^G_{\mathrm{rad}}(r)$ directly from the covariance by measuring the neighbour covariance.
Up to now we have only considered the variance and the mean. The actual distribution of $p^G_{\mathrm{rad}}(r)$ for some specific r is illustrated in Figure 4.6. We see that the distribution is approximately Gaussian, which corresponds to the fact that the main contribution to the variance is due to the sampling procedure, which follows a binomial distribution, which is approximately Gaussian.

[19] The following derivation assumes that there exists an underlying distribution of densities U_0 that describes the ensemble of populations of the micro-model. The ensemble of populations could be realized from the density distribution by sampling. This is not exact mathematics but rather modelling, which is about how to interpret a distribution of densities.

[20] The variance of the noise $\mathrm{Var}(\eta)$ depends on the realisation of U_0, but as an approximation, we can subtract $\mathrm{Var}(\eta)$, with η corresponding to the average value of U_0.

[21] The neighbour covariance is expected to be almost zero in the case of sampling from a deterministic density, because the binomial variables of the bin counts are almost independent if the bins are small. This means that the neighbour covariance is not polluted by a lower bound.

[22] Because the assumption that U_0 and η are independent underestimates the variance of u^*.

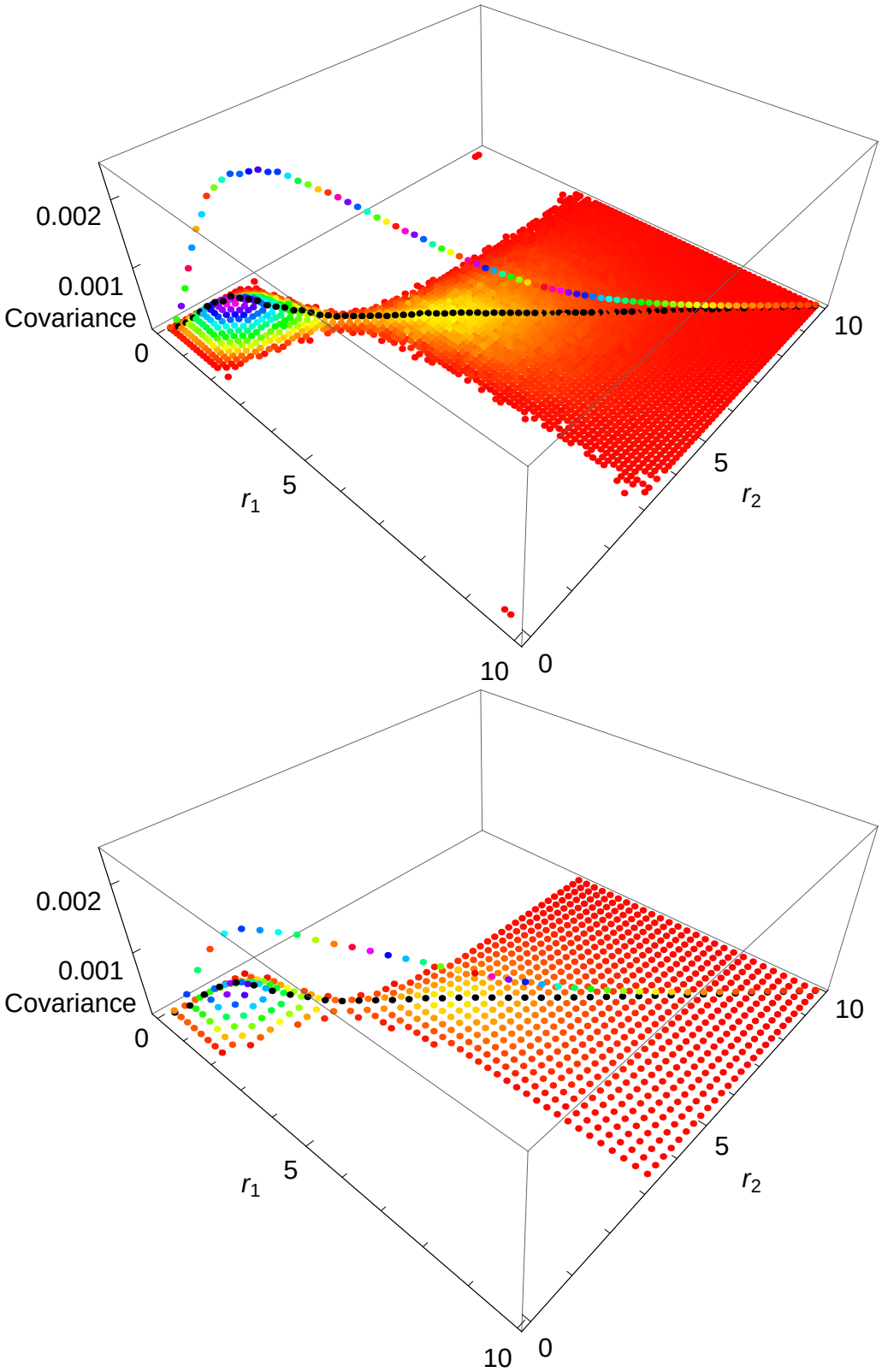

Figure 4.4 – The measured covariance of $p^G_{\text{rad}}(r_1)$ and $p^G_{\text{rad}}(r_2)$ as defined in Equation (4.3.4) with bin-size $|b| = \frac{1}{8}$ (top) and $|b| = \frac{1}{4}$ (bottom). The colours indicate the height. Only positive covariance is shown. In the blank regions $p^G_{\text{rad}}(r_1)$ and $p^G_{\text{rad}}(r_2)$ are negatively correlated. From the correlation, we see that if $p^G_{\text{rad}}(r)$ is larger than average for some small r, it tends to be larger for all small r and smaller for all large r. This results in a covariance function that is like a saddle.

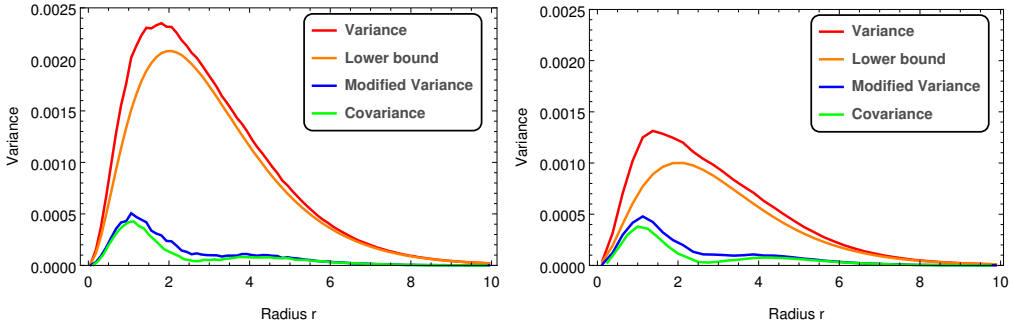

Figure 4.5 – The measured variance, the lower bound of the variance, the modified variance and the neighbour covariance as a function of r for bin-size $|b| = \frac{1}{8}$ (left) and $|b| = \frac{1}{4}$ (right). Note that the covariance has a local minimum close to $r = 2.7$, which is at the saddle point in Figure 4.4.

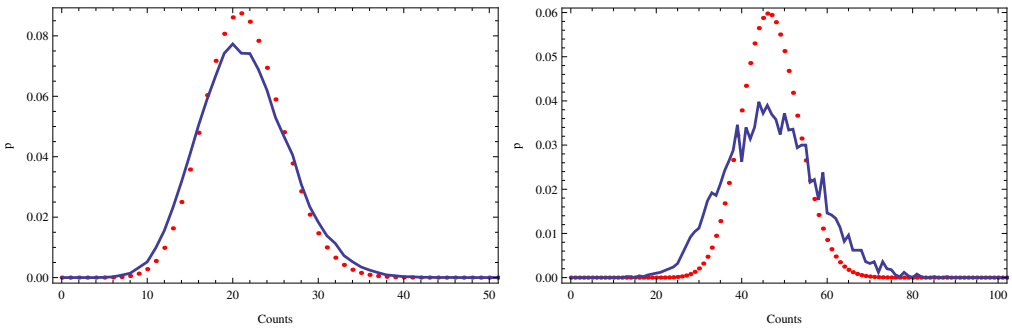

Figure 4.6 – The measured distribution of p^U_{rad} at a single r and t (blue) compared to the expected binomial distribution from sampling alone (red). In equilibrium at $r = 1$ with bin-size $\frac{1}{8}$ (left) and at $t = 10$, $r = 2$ with bin-size $\frac{1}{4}$ (right).

4.3.2 Equilibrium fluctuations

We can do the same for the variance of u^G as we did for the variance of p^G_{rad}. The neighbour covariance of u^G is shown in Figure 4.7. Because the neighbour covariance appears to decrease exponentially, we plot the natural logarithm of the neighbour covariance. This shows two distinct regions. For small r, the variance goes like $c_1 \exp(-4r) \propto u^4_{\text{eq}}$, while for larger r it goes like $c_2 \exp(-r) \propto u_{\text{eq}}$, where $u_{\text{eq}} = \exp(-r)$ is the exact equilibrium density.

If we ignore the fact that fluctuations spread through the system due to diffusion[23], in Section 3.6 we predicted that the variance should go like u. We see that this is approximately the case for larger r. For smaller r, the neighbour covariance follows a different functional form of u.

It is interesting to see why there are two different regions. To investigate this, we look for some qualitative change close to the transition spot around $r = 2.65 \pm 0.1$. The transition happens at the saddle point in Figure 4.4, so we cross the boundary of the two correlated regions in the covariance landscape. The behaviour of the neighbour covariance of u^G is different in the two different regions. Note that such a saddle point should be there if there are two correlated regions that are negatively correlated with each other. However, it still raises the question why the saddle point is at that particular value of r.

[23] To avoid complications as mentioned in Section 4.3.

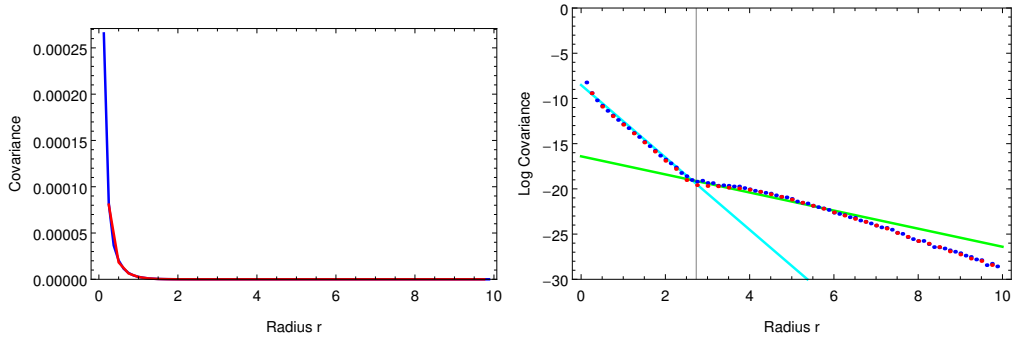

Figure 4.7 – The neighbour covariance of u and the natural logarithm of the neighbour covariance of u as a function of r. Two different regions can be distinguished. The slope of the cyan line is -4, so the covariance goes approximately like $c_1 \exp(-4r)$ for small r. The slope of the green line is -1, so the covariance for large r goes approximately like $c_2 \exp(-r)$. The vertical line is drawn at $r = 1 + \sqrt{3}$, at which the diffusion term has a maximum. Compare to Figure 4.8.

In Figure 4.8, we see that the diffusion term has a maximum[24] at $r_b = 1 + \sqrt{3} \approx 2.73$, which is very close to the transition value of r. Note that there is a balance between reproductions, deaths and diffusion in the system. Close to zero, there are much more reproductions than deaths and particles leave the region by diffusion. Up to r_b, the net incoming flux of particles increases. Above r_b, it decreases again. We do not have an adequate explanation of the different behaviour for small r, but it might have to do with fluctuations due to transportation and short distances. Also the fitness function $f = \frac{1}{r}$ varies very fast near the origin, so there is much extra noise from neighbours. On top of that, we derived that the fluctuations should also depend on the average number of births and deaths, which is a lot higher at small r.
At large r, the fitness $\frac{1}{r}$ changes much slower than the density $\exp(-r)$. Also, for large r, u^G is averaged over a larger region in space, because the bin b corresponds to a larger spherical region. This might explain why the variance does go like u_{eq} for large r.

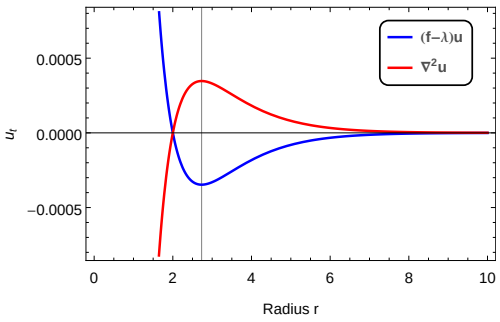

Figure 4.8 – The terms $(f - \lambda)u$ and $\nabla^2 u$ cancel in equilibrium, but their values might give insight into the measured variance of u.

[24] the birth-death term has a minimum as they should add to 0, because it is in equilibrium

4.3.3 Non-equilibrium fluctuations

We can do the same with the non-equilibrium measurements from Section 4.2.2. In Figure 4.9, the average and the variance is shown compared to the prediction of the birth-death-diffusion equation and the lower bound of the variance. As we already found, the average is very well described by the birth-death-diffusion equation. The difference between the lower bound is shown in Figure 4.10 and is compared to $\partial_t u$. We see that the extra variance increases fastest when $\partial_t u$ is large. Some systems are ahead of the average and some systems are behind and if $\partial_t u$ is large, a small lag in time results in a large difference.
After the system has come to equilibrium, there is still extra variance, which we discussed in the previous section. Note that at $t = 0$ there is no extra variance, because the populations are in fact sampled from a single deterministic density.

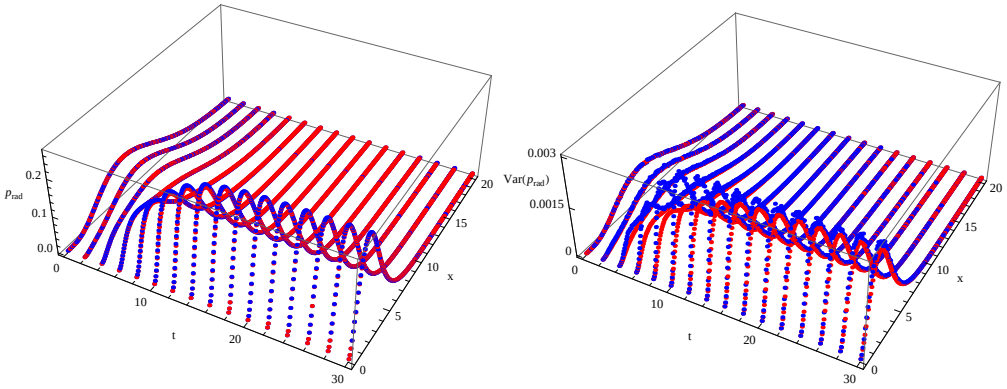

Figure 4.9 – On the left, the measured average (blue) is compared to the exact solution (red). On the right, the measured variance (blue) is compared to the lower bound for the variance (red). The difference is the modified variance and is shown in Figure 4.10.

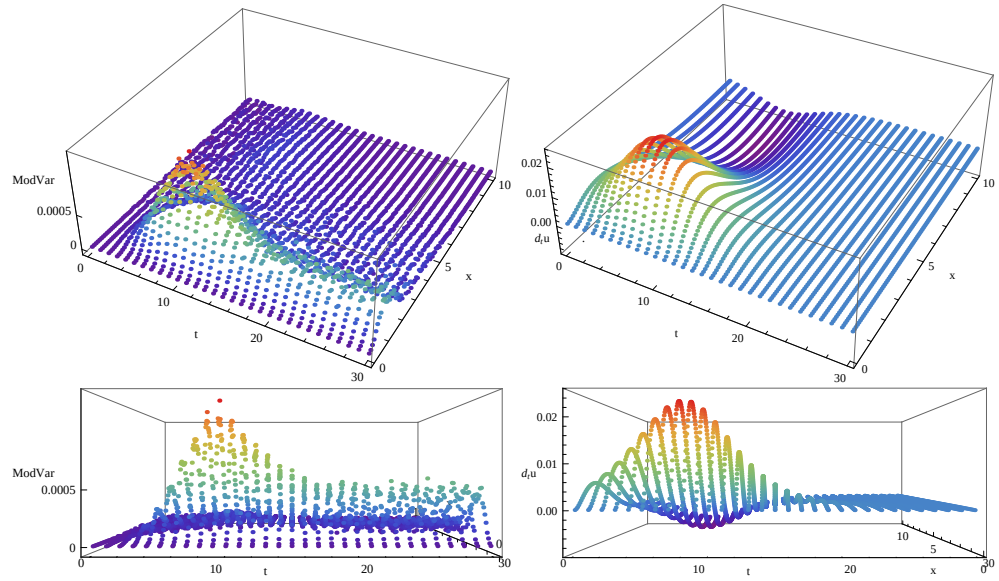

Figure 4.10 – The modified variance (left) compared to the exact time derivative $\partial_t u$ of the exact solution (right), viewed from two different angles.

4.4 Conclusion and discussion

We described evolutionary models in general and discussed some modelling choices that can be made. For example about the distribution of births, deaths and mutations through time. We defined the notion of a fitness landscape, explained how clustering can appear and discussed the competition between natural selection and genetic drift by conducting a thought experiment. In finite populations, evolving according to a Moran process, a lesser mutant is able to fixate in the population with a non-zero probability. This is a finite size effect and disappears in the limit of infinite population size. Natural selection persists for all population sizes and in large populations, only superior mutants have a significant probability of fixation.

Our first specific model, with a fitness landscape based on 1-in-3-SAT, demonstrated that evolution finds much better individuals than a random search. Also, it showed that larger populations achieve a higher equilibrium fitness than smaller populations. Viewing the population size as the inverse of a temperature clarifies why small populations explore the fitness landscape faster, while large populations obtain a higher fitness. The autocorrelation of a large population remains large for a long period of time compared to small populations. We implemented a procedure similar to simulated annealing and found that it indeed speeds up evolution towards the equilibrium state.

The second model considered a population of diffusing particles in a continuous space. Three fitness landscapes were defined that illustrate different aspects of the general dynamics. We derived the birth-death-diffusion equation to describe the average evolution of the micro-model. We studied how the birth-death-diffusion equation can be solved and derived the exact solution for the hydrogen fitness landscape. We found that the exact solution describes the average behaviour of the micro-model very well by comparing with simulations of the micro-model.

Unfortunately, the birth-death-diffusion equation does not account for clustering of the population. Since clustering is due to fluctuations that cause extinction, we explored how to add these fluctuations to the birth-death-diffusion equation. By analysing the Poissonian nature of births and deaths, we motivated the definition of the stochastic birth-death-diffusion equation. The assumptions made are only justified when the density is high, so to model extinction correctly, a different approach is necessary.

In order to measure fluctuations in the micro-model, we consider deviations from the exact average solution in the hydrogen fitness landscape. For this fitness landscape, the population is confined in a small region of space. As a result, we can accurately measure deviations from the average.
To compare a density to a micro-model population, we need some way of translating these descriptions. However, there is no one-to-one correspondence of a continuous density to a micro-model population. Therefore, taking an ensemble of populations from a single deterministic continuous density u already shows variety. If we want to measure deviations from some average density u by looking at an ensemble of micro-model populations, the task is to determine how much the measured ensemble statistics deviate from the expected ensemble statistics.

We proceeded by defining the modified variance and the neighbour covariance. These independent approaches to approximate the variance of the density distribution produce similar results, which validates the methods to some extent.
Applying these methods, we found that the variance scales differently in different regimes of the hydrogen fitness landscape. Also, we found evidence that the density distribution spreads fastest when the average density changes fastest, that is when $\partial_t u$ is largest.

The next step would be to properly adapt and simulate the stochastic birth-death-diffusion equation, after which statistics can be compared to the micro-model simulations. For example, it would be interesting to search for a density description that correctly represents the spread and average movement of a population cluster in a flat landscape. If this is achieved, it would be interesting to compare transition rates in the double well potential for the stochastic birth-death-diffusion equation and the micro-model.
We proposed a method to account for extinction at small densities, which might be tuned in such a way that the density acquires the correct spread and movement.

Appendix A

Derivations

A.1 Derivation of fixation probability

Here, we calculate the probability of fixation of i red balls with fitness f_R in a population that further consists of $M - i$ blue balls with fitness f_B using the Moran process, as discussed in Section 1.5.2. This derivation is based on [Nowak, 2006]. We give some more details on some steps, while referring to their derivation on other steps. The relevant probabilities are given by

$$\text{P(Red ball is selected for reproduction)} = \frac{if_R}{if_R + (M-i)f_B} \tag{A.1.1}$$

$$\text{P(Blue ball is selected for reproduction)} = \frac{(M-i)f_B}{if_R + (M-i)f_B} \tag{A.1.2}$$

$$\text{P(Red ball is selected for death)} = \frac{i}{M} \tag{A.1.3}$$

$$\text{P(Blue ball is selected for death)} = \frac{M-i}{M} \tag{A.1.4}$$

The probability that the number of red balls decreases by one is given by

$$P_{i\to i-1} = \text{P(blue birth)P(red death)} = \frac{(M-i)f_B}{if_R + (M-i)f_B}\frac{i}{M}, \tag{A.1.5}$$

while the probability that the number of red balls is increased by one is given by

$$P_{i\to i+1} = \text{P(red birth)P(blue death)} = \frac{if_R}{if_R + (M-i)f_B}\frac{M-i}{M}. \tag{A.1.6}$$

The probability that the number of red balls does not change is given by

$$P_{i\to i} = 1\text{–}P_{i\to i-1}\text{–}P_{i\to i+1}. \tag{A.1.7}$$

Now we denote the probability of fixation of i red balls by x_i, which is the probability we set out to derive. Obviously, $x_0 = 0$ and $x_M = 1$. Now note that the following relation holds between the x_i

$$x_i = P_{i\to i-1}x_{i-1} + P_{i\to i}x_i + P_{i\to i+1}x_{i+1}, \tag{A.1.8}$$

because we express the probability of fixation using fixation probabilities conditional on what happens in the first step.

Now we define the relative fitness

$$r = \frac{f_B}{f_R} \tag{A.1.9}$$

of blue balls compared to red balls. Note that

$$\frac{P_{i\to i-1}}{P_{i\to i+1}} = \frac{\frac{(M-i)f_B}{if_R+(M-i)f_B}\frac{i}{M}}{\frac{if_R}{if_R+(M-i)f_B}\frac{M-i}{M}} = \frac{f_B}{f_A} = r. \tag{A.1.10}$$

Equation (A.1.8) can be written like $\vec{x} = P\vec{x}$ for all i simultaneously, so we see that $\vec{x}$ is an eigenvector of the stochastic matrix P with eigenvalue 1. On page 99 in [Nowak, 2006] it is shown that the solution x_i is given by[1]

$$x_i = \frac{1+\sum_{j=1}^{i-1}\prod_{k=1}^{j}\frac{P_{k\to k-1}}{P_{k\to k+1}}}{1+\sum_{j=1}^{M-1}\prod_{k=1}^{j}\frac{P_{k\to k-1}}{P_{k\to k+1}}} = \frac{1+\sum_{j=1}^{i-1} r^j}{1+\sum_{j=1}^{M-1} r^j} = \frac{1+\frac{r^i-r}{r-1}}{1+\frac{r^M-r}{r-1}} = \frac{1-r^i}{1-r^M} \tag{A.1.11}$$

It can be worked out that

$$\mathrm{P}(\text{fixation of blue balls}) = 1 - \mathrm{P}(\text{fixation of red balls}) = 1 - \frac{1-r^i}{1-r^M} = \frac{1-\frac{1}{r}^{M-i}}{1-\frac{1}{r}^{M}}, \tag{A.1.12}$$

from which we see that the formula is valid for both $r > 1$ and $0 < r < 1$, because the equation is the same when we interchange the role of red and blue balls[2].

A.2 Poisson process

In this section, we elaborate on the use of the Poisson process for the model in Section 3.1. In a Poisson process with rate $\lambda(t)$ the number of events in a time interval dt follows a Poisson distribution with parameter $\lambda(t)dt$,

$$\mathrm{P}(N_{\text{events}} = k) = \frac{(\lambda(t)dt)^k}{k!}\exp(-\lambda(t).dt), \tag{A.2.1}$$

The number of events in non-overlapping time intervals are independent. The Poisson distribution is obtained by taking a limit of the binomial distribution, namely by increasing the number of trials $n \to \infty$ while decreasing the probability of success $p \to 0$ such that the average number of successes np is constant. This limit corresponds to indefinitely splitting a time interval Δt in independent smaller intervals, while maintaining the same average number of events in Δt.
It can be shown that the time between events follows an exponential distribution with parameter $\lambda(t)$. The exponential distribution is memoryless, which corresponds to the independence of the events in different time intervals.

[1] Here, we used that $\sum_{j=1}^{i-1} r^j = \frac{r^i-r}{r-1}$, because $(r-1)\sum_{j=1}^{i-1} r^j = r^i - r$.
[2] Either $r > 1$ or $\frac{1}{r} > 1$.

A.3 Clustering

The critical case in [Meyer et al., 1996] is equivalent to our model in Section 3.1 with a flat fitness landscape, $f(x) = f_0$. By considering times since the last common ancestor of particles, it is shown that in equilibrium the mean squared pair distance is equal to

$$\mathrm{E}(r_2^2) = \int d^d r r^2 P_2(r) = \frac{2dD(M-1)}{f_0}. \tag{A.3.1}$$

Here, $P_2(r)$ is the pair correlation function, which is the probability that the distance between two randomly chosen individuals is r. The constants are defined in Section 3.1. If the mean squared pair distance $\mathrm{E}(r_2^2)$ is larger, this implies that the population is more spread out. The length scale of the region covered by the population is typically of the order $\sqrt{\mathrm{E}(r_2^2)} = \sqrt{\frac{2dD(M-1)}{f_0}}$. From this, we see that the population is more spread when $\frac{2dD}{f_0}$ is larger, which is the averaged squared distance covered by diffusion between births, because $\frac{1}{f_0}$ is the average time between births. Also, if the population size M is larger, the population is more spread. We already mentioned this in Section 1.5.1.

They also show that, in equilibrium, the centre of mass of the population performs a random walk with $E((x_{\mathrm{cm}}(t+\Delta t) - x_{\mathrm{cm}}(t))^2) = 2dD\Delta t$. We checked that the centre of mass of the population indeed performs a random walk by simulation of the micro-model. We found that the measured average displacement $\mathrm{Mean}((x(t+\Delta t) - x(t))^2) \propto \Delta t$ exactly.

If we want to describe the micro-model on a different scale, we should at least make sure that the spread and average movement of the population are in agreement with the results in this section.

A.4 Deterministic extinction

In this section we propose a way to model extinction effects that are not taken into account by the stochastic birth-death-diffusion equation as described in Section 3.6. Note that extinction of a group of N particles occurs when all N lineages of these particles become extinct. Extinction of different lineages is approximately independent when N is much smaller than the population size M. The probability of extinction within time Δt of a single lineage could be written as π_0. The probability that all N lineages become extinct in Δt is therefore equal to

$$\mathrm{P}(\text{extinction of all } N \text{ lineages}) = \pi_0^N. \tag{A.4.1}$$

Now imagine an ensemble of N_e identical systems. We expect that in about $\pi_0^N N_e$ systems the group of N particles becomes extinct. This motivates to define the extinction process given by

$$\frac{\partial N}{\partial t} = -c\pi_0^N N \tag{A.4.2}$$

As N gets larger, the probability of extinction decreases. This is represented in the extinction process because the relative change

$$\frac{1}{N}\frac{\partial N}{\partial t} = -c\pi_0^N \tag{A.4.3}$$

decreases as N gets larger, because $\pi_0 < 1$ is a probability. Compared to the birth-death-diffusion equation, the extinction term of Equation (A.4.2) is negligible when N is large, while it has a noticeable effect when N is small.

Extinction in some place leads to extra births somewhere else, because the population size is constant. The extinction process can be seen as a refinement to the normal death process. A better interpretation might be to consider it as a compensation for expected births that are not possible due to extinction. To illustrate this, for the sequence of births and deaths $B - D - D - D - B - D - B - B$, with B a birth and D a death, we would expect that the N does not change, because there are as many births and deaths. However, such a sequence is only possible when there are more than 2 particles. For 1 or 2 particles, this sequence would result in 0 particles, because the particles become extinct. The extinction process is a model to account for sequences that are expected but not possible due to extinction.

We now add the extinction process to the stochastic birth-death-diffusion equation. By adding terms to ensure constant population size, we find

$$\partial_t u = \Big(f(x) - \langle f \rangle(t) - c(\pi_0^u - \langle \pi_0^u \rangle(t)) + D\partial_{xx} \Big) u + k\sqrt{u}\eta - \langle k\sqrt{u}\eta \rangle, \tag{A.4.4}$$

where $\langle f \rangle = \int f u dx$ and $\langle \pi_0^u \rangle = \int \pi_0^u u dx$ denote the population averages. We can put every mechanism that ensures constant population together and write

$$\partial_t u = \Big(f(x) - c\pi_0^u - + D\partial_{xx} \Big) u + k\sqrt{u}\eta - \lambda(t) u \tag{A.4.5}$$

where $\lambda(t)$ forces the constraint of constant population size.

Appendix B

Code

In this appendix we show the core of the programs used to simulate the models[1]. We call it pseudo-code, because the variables and routines are not defined, while their names are suggestive.

B.1 Bit-strings code

```
for (t=0; t<NBIRTHS; t++) {

  /* Death */
  replaced = (int) M*genrand();               //Randomly choose some individual to be replaced (death)

  /* Reproduction */
  parent = select_individual(fitness);        //Select parent: selection probabilities according to fitness
  for (j=0; j<N; j++) {                       //The parent state is copied into the spot of the replaced individual
    population[replaced][j] = population[parent][j];
  }

  /* Mutation */
  for (j=0; j<N; j++) {
    if (genrand()<MUTPROB) {                  //Flip every bit with probability MUTPROB
      population[replaced][j] ^= 1;           //1->0 and 0->1 (using bitwise XOR)
    }
  }
  compute_fitness(population[replaced]);      //Compute the fitness of the new particle (possibly mutated)

}
```

Listing B.1 – The core of the C program used for the model in Section 2.1.

[1] Mutation for the first model could be implemented more efficiently by generating a binomial variable to determine how many bit flips there are, after which these are randomly distributed over the bits. However, this is just a minor improvement in performance, because the routine to select an individual dominates the computational costs.

B.2 Diffusion code

```
while (t<tmax) {

  deltat = exprand(sumfitness);                          //The time increment is an exponential random variable
  t += deltat;

  /* Diffusion process */
  sigma = sqrt2D*sqrt(deltat);                           //Pre-calculate for efficiency
  for (i=0; i<M; i++) {                                  //Every particle diffuses
    for (j=0; j<DIMENSION; j++) {                        //DIMENSION components j for every particle i
      population[i][j] += sigma*normalrand();            //Random walk (diffusion)
    }
    fitness[i] = compute_fitness(population[i],K);       //Calculate new fitness
  }

  /* Reproduction process */
  replaced = (int) M*genrand();                          //Randomly choose some individual to be replaced
  parent = select_individual(M,fitness,&sumfitness);     //Select parent: selection probabilities according to fitness
  for (j=0; j<DIMENSION; j++) {
    population[replaced][j] = population[parent][j];     //The state of the parent is copied into the spot of the
      replaced individual
  }
  sumfitness = sumfitness - fitness[replaced] + fitness[parent];  //Update the sum of the fitnesses necessary to compute
       selection
  fitness[replaced] = fitness[parent];                   //Update the fitness of the new particle

}
```

Listing B.2 – The core of the C program used for the model in Section 3.1.

Bibliography

M. A. Nowak, *Evolutionary dynamics* (Harvard University Press, 2006).

T. Zhou, J.-G. Liu, W.-J. Bai, G. Chen, and B.-H. Wang, Physical Review E **74**, 056109 (2006).

R. A. Boschma and K. Frenken, Jahrbuch für Regionaiwissenschaft **23**, 183 (2003).

C.-J. Wang (2013).

W. Young, A. Roberts, and G. Stuhne, Nature **412**, 328 (2001).

T. Back, *Evolutionary algorithms in theory and practice: evolution strategies, evolutionary programming, genetic algorithms* (Oxford university press, 1996).

G. Burgers, P. J. van Leeuwen, and G. Evensen, Monthly weather review **126**, 1719 (1998).

P. J. Van Leeuwen, Monthly Weather Review **137**, 4089 (2009).

N. J. Gordon, D. J. Salmond, and A. F. Smith, in *IEE Proceedings F-Radar and Signal Processing* (IET, 1993), vol. 140, pp. 107–113.

C. W. Gardiner et al., *Handbook of stochastic methods*, vol. 3 (Springer Berlin, 1985).

S. Karlin, *A first course in stochastic processes* (Academic press, 2014).

M. J. Landis, J. G. Schraiber, and M. Liang, Systematic biology **62**, 193 (2013).

M. Meyer, S. Havlin, and A. Bunde, Physical Review E **54**, 5567 (1996).

T. Brotto, G. Bunin, and J. Kurchan, arXiv preprint arXiv:1507.07453 (2015).

M. Motoki and R. Uehara, SAT pp. 293–305 (2000).

B. Korte and J. Vygen, *Combinatorial optimization*, vol. 2 (Springer, 2012).

T. J. Schaefer, in *Proceedings of the tenth annual ACM symposium on Theory of computing* (ACM, 1978), pp. 216–226.

L. Peliti, arXiv preprint cond-mat/9712027 (1997).

G. Sella and A. E. Hirsh, Proceedings of the National Academy of Sciences of the United States of America **102**, 9541 (2005).

L. Devroye, *Non-uniform random variate generation* (1986).

M. Matsumoto and T. Nishimura, *mt19937.c* (2002).

D. A. Kessler, H. Levine, D. Ridgway, and L. Tsimring, Journal of statistical physics **87**, 519 (1997).

Y.-C. Zhang, M. Serva, and M. Polikarpov, Journal of Statistical Physics **58**, 849 (1990).

M. H. Holmes, *Introduction to the foundations of applied mathematics*, vol. 56 (Springer Science & Business Media, 2009).

H. Ohtsuki and M. A. Nowak, Journal of theoretical biology **243**, 86 (2006).

D. J. Griffiths, *Introduction to quantum mechanics* (Pearson Education, 2005).

D. J. Lawson and H. J. Jensen, Bulletin of mathematical biology **70**, 1065 (2008).

U. C. Täuber, M. Howard, and B. P. Vollmayr-Lee, Journal of Physics A: Mathematical and General **38**, R79 (2005).

E. Moro, Physical Review E **70**, 045102 (2004).

J. Pitman, *Probability* (1993).