

UNIVERSITY OF CALIFORNIA, DAVIS

PHYSICS OF COMPUTATION: SPRING PHY 256B

---

# Spontaneous Mutations in Population Dynamics and Phase Transitions

---

Arthur PLAUD  
Year 2022

*Professor* : Jim CRUTCHFIELD  
*Assistant* : Mikhael SEEMAN

## Abstract

This project aims to study the influence of general natural selection but mainly spontaneous mutations on population growth. It accordingly follows the usual framework for mutation-selection models, but includes time varying population size as an interesting feature of the dynamics, instead of fixing it and only studying relative genetic dominance. A model is built accordingly as an usual discrete time mutation-selection model, but with the selection step deviating from the fixed-population size work. We studied the influence of mutation rate on the population dynamics and ended up with an optimal mutation rate for population growth. This optimum results from the competition between fitness exploration (that grows with mutation rate) and genetic fixation (that behaves in the opposite way). We also recover previous mutation-selection results. Finally, an alternative version of the model is considered, that leads to interesting and yet-to-be-explored behaviours.

# Contents

<b>1</b>	<b>Introduction and Context</b>	<b>1</b>
1.1	Mutation-Selection Models . . . . .	1
1.2	Exploration of the space of fitnesses through spontaneous mutations	2
1.3	Transition between order and disorder . . . . .	3
<b>2</b>	<b>The system and description of the model</b>	<b>4</b>
2.1	The system . . . . .	4
2.2	The selection step . . . . .	5
2.3	The mutation step . . . . .	6
<b>3</b>	<b>Results</b>	<b>6</b>
3.1	An optimum for population growth . . . . .	6
3.2	Stepwise increase in fitness . . . . .	8
3.3	Modifying the model by suppressing correlations during the mutation step . . . . .	9
<b>4</b>	<b>Conclusion</b>	<b>12</b>
<b>5</b>	<b>References</b>	<b>12</b>

## 1 Introduction and Context

### 1.1 Mutation-Selection Models

There is a long history of mutation-selection models. They can describe both discrete time or continuous time processes but there is always two components in the dynamics. On the one side, the selection term increases the relative proportion of individuals with high genetic advantage at the expense of the ones with genetic disadvantage. This step can be either deterministic or stochastic. On the other side, the mutation term is a stochastic change of the genotype of certain individuals inside the population.

The first occurrence of a mutation-selection model is traditionally attributed to Fisher in 1930 [1], where he also introduced the notion of fitness as it's used today in evolutionary dynamics. The fitness of an individual quantifies the ability of this individual to survive and reproduce. Formally, it is a function that outputs a real number from the underlying genetic structure of the individual. It is actually a way to avoid the details of genetic information storage and the complex biologic processes that are processing this information, while keeping the influence of the genotype on population growth. It has been so widely used that fitness can today be considered as a synonym to reproduction rate in evolutionary dynamics. Discrete time processes are describing non-overlapping generations and the selection step can then be interpreted in term of mean number of offspring for an individual with a given genotype. When the process is taken to be continuous, the description can get more complicated when considering overlapping generations and some dependence of the evolution properties (reproduction rate, mutation rate) on the age of the individual. This is for example achieved in [2].

Despite a very rich literature on mutation-selection dynamics, few links have

been made with population growth. Indeed, population growth is usually considered as part of another kind of dynamics, ecosystem dynamics. Models associated with ecosystem dynamics typically don't present any genetic structure (not even a fitness function), and include many different species that compete for limited resources and interact mainly in a deterministic way. An overview of these historical models can be found in [3]. Mutations and selection are not considered in these models. On the other hand, when studying genetic dynamics, the population size is typically fixed or even taken as infinite. Here, in order to study population growth, population size will be a time-varying parameter. This has already been done but is non standard for mutation-selection models.

## 1.2 Exploration of the space of fitnesses through spontaneous mutations

The historic role of the mutation step in mutation-selection models is to ensure that underrepresented genotypes are given a chance to fix themselves inside the population if they give a fitness advantage. This is mainly the case for finite size models, as in this case some genotypes are not represented at all, and they could never appear (despite potentially being high fitness) without mutations. In the infinite size case, mutations are less required but they still describe real-life phenomena. In some cases, mutations can even help to describe complex genetic recombinations that do not occur in fitness-based reproduction. For example, two individuals with the same fitness can engender a less-fit child. In selection-mutation models, mutations help modelling these marginal phenomena.

This exploration of the space of fitnesses can be seen in the variation of the mean fitness with time. It increases due to favorable mutations taking over the population. This increase is typically done in different ways depending on the values of the parameters  $N$  (the population size), and  $p$  (the mutation rate). When  $Np \ll 1$ , favorable mutations are rare and there is time to fix the genotype inside the population before the next mutation. We end up with a stepwise increase in fitness, as seen in [4], from which are taken Figure 1 and 2.

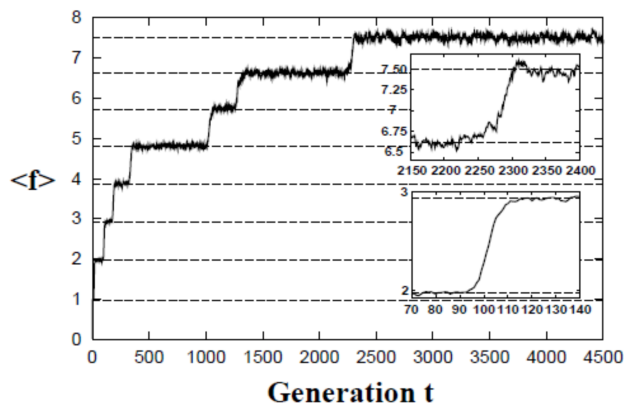


Figure 1: Steep increases in fitness separated by stagnating epochs for  $Np = 0.5$

When increasing population size or mutation rate, the epochs last fewer time, eventually leading up to a smooth increase in fitness when  $Np \gg 1$ . We

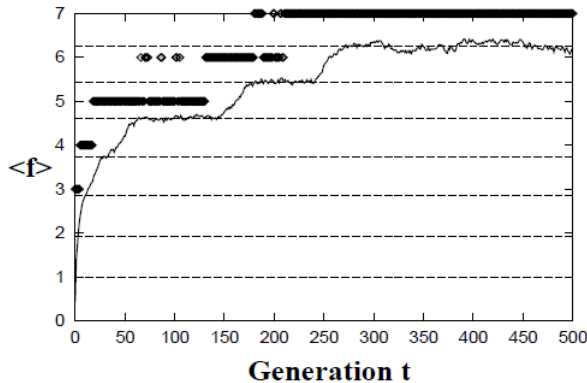


Figure 2: Here, periods of increase in fitness and stagnating periods last comparable time for  $Np = 2$

can also compare the time scales. To reach the same mean fitness, it takes 10 times less iterations in the second case. This correctly relates to the fact that exploration of the space of fitnesses is more efficient when the mutation rate increases.

### 1.3 Transition between order and disorder

In a purely selective model with fixed population size, the long term behaviour is just a fully homogenous population where the highest fitness existing in the initial state took over. This describes an ordered phase. The effect of mutations is then to drive the system away from homogeneity. This has also been studied a lot and gave birth to the notion of error threshold. The error threshold is actually a mutation rate threshold, over which the system is described as disordered, with a broad range of fitnesses being relevant inside the population, and below which the population is typically dominated by one fitness value. As for every phase transition, the change of behaviour actually becomes steep in the limit of infinite population size. Very simple mutation-selection model exhibit such behaviours, as in [5]. Another exemple is shown in Figure 3, which is taken from [6].

On this figure, the mean fitness decreases linearly from 1 to 0 when the (normalized) mutation rate reaches a value close to 0.6. Moreover, the standard deviation in fitnesses, described by  $\hat{x}$  (it is actually the input of the fitness function here and not the output, so it describes directly the genotype) jumps from a low value before the treshold (less than 0.1) to a high value after (around 0.9). We are here in the infinite size case, so the transition is infinitely steep, we can talk about a phase transition from order to disorder with the control parameter being the mutation rate. Regarding the order parameter, here the standard deviation was taken. In this project, we will use the entropy of the fitness distribution as one of the order parameters, renormalized by population size (which is time-varying here). This mesures, as for the standard deviation, the level of heterogeneity in the population.

The decrease in mean fitness is also very interesting : we see that increasing the mutation rate leads to less fit individuals at equilibrium. Despite the population

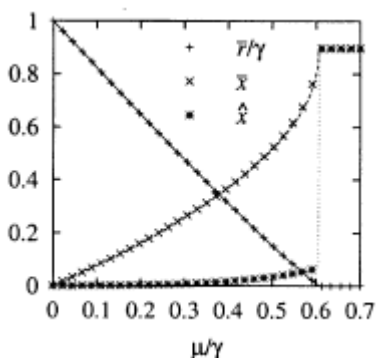


Figure 3: Variation of both the mean and the standard deviation of the stationary distribution of fitnesses as a function of the mutation rate

being fixed here, we can imagine the influence of this phenomena on population growth : this actually limits population growth for high mutation rate. When competing with the exploration aspect, this tends towards the existence of an optimum for population growth, as will be confirmed later.

As previously explained, this influence of mutation rate on both fitness increase and heterogeneity of the population has been since long established. But there is some reluctance to apply this to population growth, instead of taking a fixed-size population (finite or infinite). There is however good experimental proof that this could be in some cases a correct description of real-life population dynamics, as is the notion of optimal mutation rate for population growth. This has for example been developed in [7].

## 2 The system and description of the model

### 2.1 The system

The model is taken to be a discrete time mutation-selection model. Accordingly, the generations do not overlap and each generation encounters two steps successively, the selection step followed by the mutation step. There is no detailed genetic structure and individuals are completely described by their fitness value. The updating rules will be described later and the initial conditions are taken as a gaussian distribution of fitnesses over  $N$  individuals. The important parameters are the initial population size  $N$ , the mutation rate  $p$  and the distribution width  $\sigma$ .  $\sigma$  quantifies the initial distribution of fitnesses but also plays a role during the mutation step, as will be detailed later. For the whole duration of the project, we only studied the influence of  $p$  the mutation rate.  $N$  was fixed to 2000 and  $\sigma$  to 1.

## 2.2 The selection step

The selection step describes how individuals reproduced inside a generation. The mean of reproduction considered here is cloning : individuals reproduce independently of each other and the child has exactly the same fitness value. For each individual, there are 3 possible outcomes :

1. the individual died without reproducing, it is erased of the population.
2. the individual gave birth to 1 child before dying, which is implemented as a status quo for the population.
3. the individual gave birth to 2 children before dying, it duplicates itself inside the population.

The probabilities of each outcome depend of the fitness value of the individual, and more precisely of the gap between its fitness and the mean fitness of the population. These probabilities are given by the pseudo-step functions below :

- $P_0(x) = \frac{1}{1 + 10e^x}$
- $P_2(x) = \frac{1}{1 + 10e^{-x}}$
- $P_1(x) = 1 - P_0(x) - P_2(x)$

where  $x$  is the gap between the fitness  $f$  and the mean fitness :  $x = f - \langle f \rangle$ . Figure 4 shows the shape of these outcome probabilities.

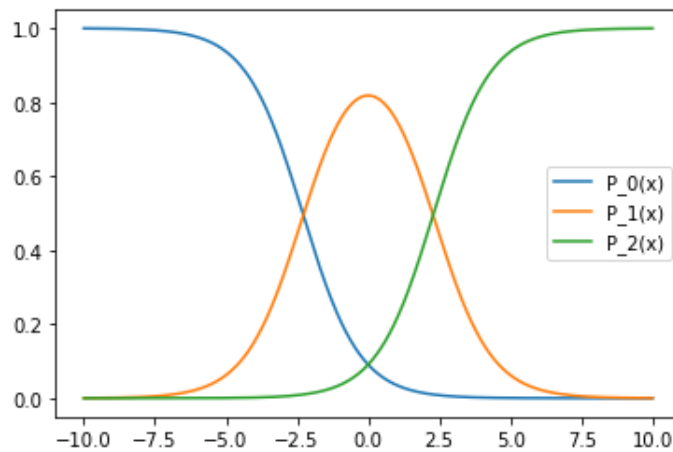


Figure 4: Outcome probabilities as a function of  $x$

As you can see, an individual very advantaged compared to the rest of the population will nearly always have 2 children, and another one very disadvantaged will nearly always die without any children. This leads to a selection of the most fit individual inside the population, and the disappearance of the least fit ones.

This step also quantifies the growth of the population : it is the one where the

population size can change in the model. However, the symmetry properties of the functions  $P_i(x)$  are restricting regarding this growth rate. For example, if the fitness distribution in the population is symmetric, we expect the population size to remain constant at this step. The symmetry property being not conserved and the presence of the mutation step means that despite symmetric initial conditions, we can still observe population growth (or decay) with this model for the selection step.

### 2.3 The mutation step

The mutation step is actually much simpler and very classic for mutation-selection models, but there still is a choice to make regarding the value of the fitness after the mutation. After the selection step, every new individual has a fixed probability  $p$  of mutating. If it mutates, the new fitness value is taken from a gaussian distribution of width  $\sigma$  (same width as for the initial distribution) and of mean  $f$ , the previous fitness value of the individual. Here, we made the choice to have correlations between the previous level of fitness and the next. This makes sense but will lead to some concerns adressed in 3.3, regarding the increase in fitness of the population.

## 3 Results

### 3.1 An optimum for population growth

The main result of the project is the highlighting of an optimum in mutation rate for population growth. We have 3 different behaviours for the time-varying population size, depending of the value of the mutation rate.

When  $p \rightarrow 0$ , there is very little exploration in fitness and the fitness increases slowly. This leads to a very homogenous population, with very little growth of the population size, and highly relevant fluctuations. This is shown in figure 5.

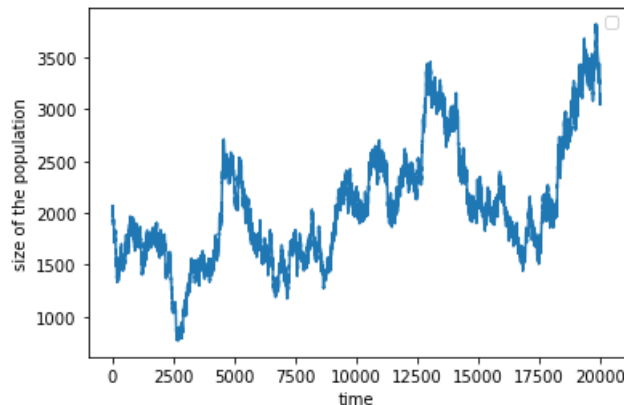


Figure 5: Evolution of the population size for  $p = 10^{-6}$

When  $p \rightarrow 1$ , the fitness also cannot increase because the high-fitness individuals don't take over the population. This leads to a very heterogenous population of general low fitness, and relatively quick extinction. This is shown in figure 6.

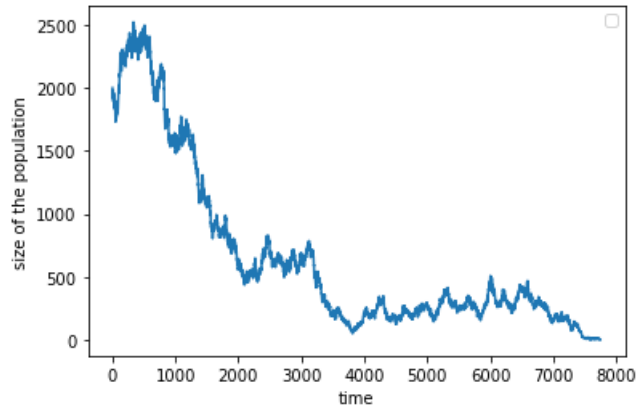


Figure 6: Evolution of the population size for  $p = 0.3$

Lastly, there is an intermediate range of mutation rate, where we observe exponential growth of the population size : the mutation rate is high enough to significantly increase the fitness while allowing the fittest individuals to take over. This is shown in figure 7.

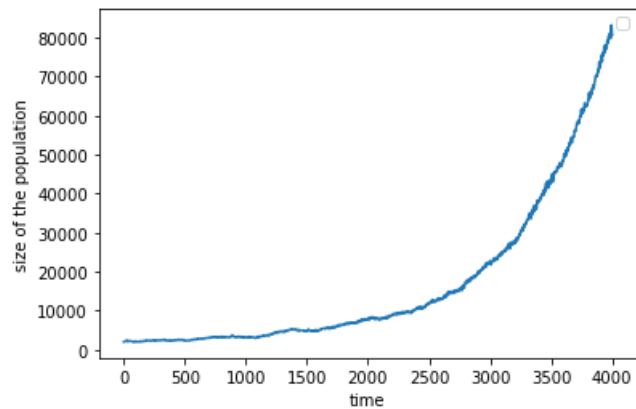


Figure 7: Evolution of the population size for  $p = 0.01$



This 3 different behaviours can be summed up by looking at the effective growth rate of the population as a function of the mutation rate. We see in figure 8, the increase in growth rate until the optimum at around 0.01, followed by a decrease in growth rate. The extinction region begins when  $p > 0.1$ . For very low mutation rate, the population is actually growing but it does so very slowly and very irregularly, while the growth is steady in the optimal range.

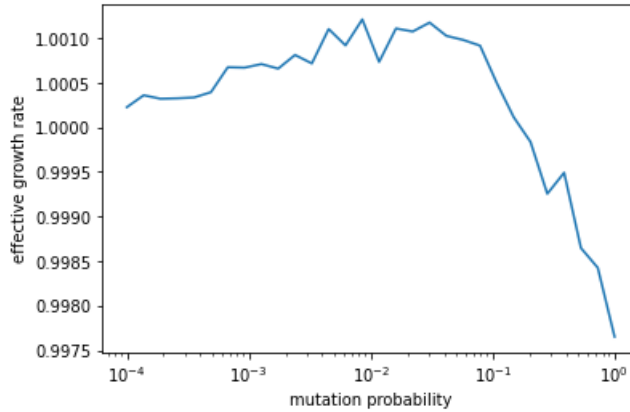


Figure 8: Effective growth rate as a function of the mutation rate

### 3.2 Stepwise increase in fitness

In this project, we also recover some of the results displayed in [4]. The relevant region is the very low mutation rate region. As in [4], we see a stepwise increase in fitness, and stagnating epochs between each step increase. This is shown in figure 9.

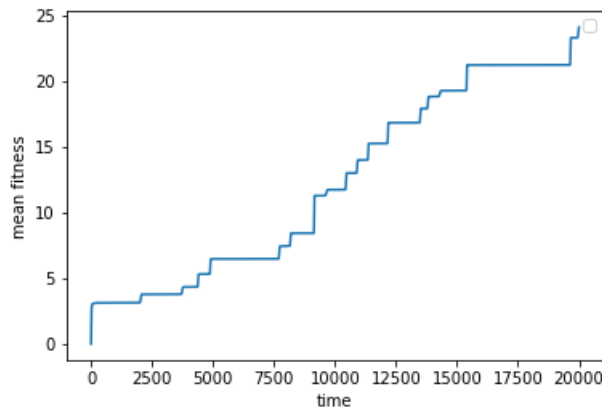


Figure 9: Mean fitness of the population for  $p = 10^{-6}$

To interpret this behavior, we also monitor the entropy of the distribution of fitnesses. This is a measure of the level of heterogeneity in the population. We see a great correlation between heterogeneity and growth. Each growth phase is visible on figure 10 as a spike in entropy. This actually gives us a detailed description of what's happening for the population : in a fully homogenous population, a favorable mutation appear and takes over the entire population (period of growth in fitness) then the population is stationary until the next favorable mutation (this is the stagnating epoch).

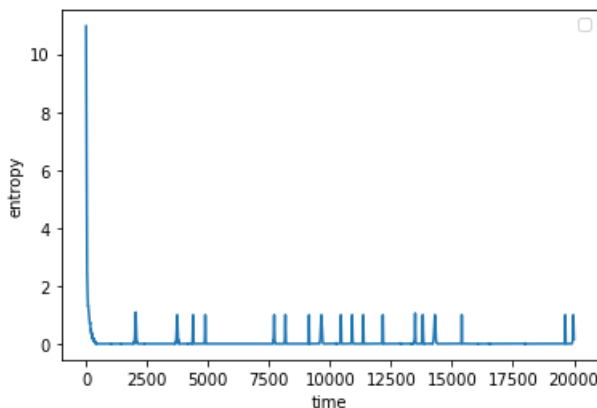


Figure 10: Entropy of the distribution of fitnesses,  $p = 10^{-6}$

### 3.3 Modifying the model by suppressing correlations during the mutation step

During the whole report, readers that are used to error treshold studies should have noticed that I fully avoided the classic stationary fitness plots as a function of the mutation rate, as shown for exemple in figure 3. These plots are the usual way of showing error treshold, but I switched during the project to growth rate as the relevant order parameter. There was 2 reasons for this, the first one being that growth rate was very effective to show population growth versus extinction. But there is also a modelling issue : with this version of the model, there are no stationary properties of the fitness distribution, especially no relaxation towards a fixed mean fitness for the population. The growth in fitness is actually close to linear in time for all range of the mutation range. The rate of change in fitness is shown in figure 11.

This curve quantifies the exploration part of the dynamics : we see that the exploration is indeed very low for low mutation rate and grows when increasing  $p$ . This curve actually has a well-pronounced power law shape with exponent close to 0.3 : the fitness rate increases very fast at first but getting over 0.2 does not drastically change its value. This explains that exploration is the critical phenomena at low mutation rate, and increasing it favors population growth. But above a certain treshold ( $p = 0.01$ ), the growth in fitness rate slows down and it's actually genetic fixation that becomes relevant, leaning to a decrease in growth rate for high values of  $p$ .

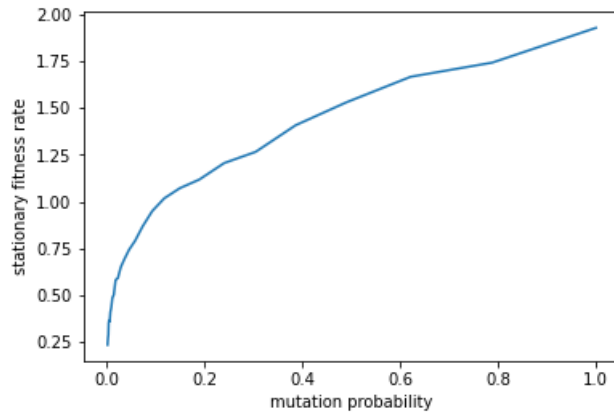


Figure 11: Rate of increase in fitness as a function of the mutation rate

So, with this version of the model, there is no equilibrium fitnesses distribution, or even equilibrium mean fitness. This isn't really an issue when studying population growth but it makes the first version of the model hard to compare with previous works on error threshold, and doesn't really answer questions about transition from order to disorder. Moreover, the way we implemented the mutation step can be criticized in many ways. Indeed, the probability of a random mutation increasing the fitness doesn't depend on the previous value of the fitness. It's always 0.5, because the distribution from which is taken the new value of the fitness is always a gaussian with the mean being the previous fitness. This doesn't match with real genotype, where there is a finite amount of possible configurations and improving the genotype through random mutations gets harder and harder when the general fitness rises. This is also why we see a linear increase in mean fitness during the evolution of the population : the mean fitness actually doesn't play any role in the model and we have translation invariance for the fitness distribution.

To adress these issues, we switched for the mutation step from a model with correlations between the previous and next level of fitness to a model without correlations. The new fitness is now taken from a gaussian always centered at 0 instead of the previous fitness value. This suppresses the never-ending rise in fitness and allows us to actually compute stationary properties. We are still taking the mutation rate as a parameter and working with an initial population size of  $N = 2000$ . Figure 12 shows this stationary fitness value (mean fitness at infinite time).

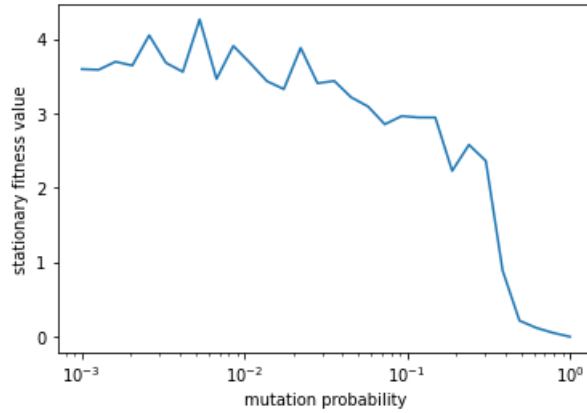


Figure 12: Stationary mean fitness as a function of the mutation rate

This plot allows us to visualize the influence of genetic fixation. Because we changed the model, we cannot compare quantitatively the values of the mutation rate where we see a decrease in fitness. However, using stationary distributions, we suppress the exploration limitation for low mutation rate and we can see the influence in fitness value of genetic fixation. This decreasing behaviour, when combined with the increasing curve for exploration (Figure 11), explains the existence of an optimum in mutation rate for population growth.

To complete this overview of stationary properties, we also take look at the entropy of the stationary fitness distribution. We can clearly see on Figure 13 the transition from order (low entropy) for low mutation rate to disorder (high entropy) at higher mutation rate. On this plot, the entropy is normalized by the maximum possible value for a fully heterogenous population, which is the logarithm of the population size. According to this plot, the error threshold for this version of the model is  $p_c \simeq 0.3$ .

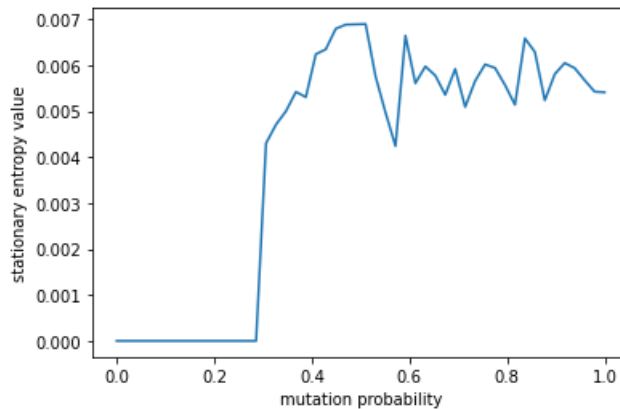


Figure 13: Normalized stationary entropy as a function of the mutation rate

## 4 Conclusion

The personal goal of this project was to build a simple model for mutation-selection with time-varying population size, to be able to see population grow and go extinct. This was achieved with a purely fitness based model. This model has two main advantages :

- It is very simple to run simulations : individuals are only represented by one number, their fitness, and this allows for efficient data processing.
- It shows non trivial behaviour and an optimum in mutation rate for population growth, which relates to real world data.

However, the model can be criticized because it allows for infinite genetic improvement, which led to the second version of the model. The intermediate solution could be to keep some degree of correlations while also reducing the probability of improving the fitness through random mutations if the fitness level is already too high.

Indeed, the choice we make for the mutation step has some influence on the population dynamics. For the no-correlations version, the effective growth rate as a function of the mutation rate, shown in figure 14, has a very different shape than the one with correlations (figure 8).

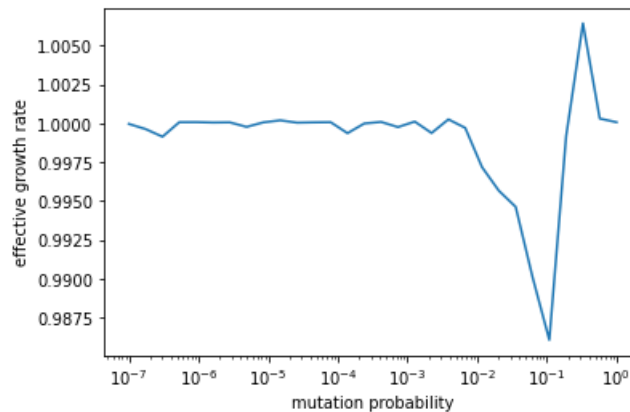


Figure 14: Growth rate as a function of mutation rate

This could lead to further work, along with some analytic calculations for both versions of the model, if I feel motivated in the future.

## 5 References

- [1] Fisher, R. A. 1930. The genetical theory of natural selection. Oxford University press .
- [2] Demetrius, L. (1992). Growth rate, population entropy, and evolutionary dynamics. *Theoretical Population Biology*, 41 (2), 208-236.
- [3] Wangersky, P. J. (1978). Lotka-Volterra population models. *Annual Review of Ecology and Systematics*, 9, 189-218.

- [4] Erik van Nimwegen, James P. Crutchfield, and Melanie Mitchell, "Finite Populations Induce Metastability in Evolutionary Search", *Physics Letters A*, 229:3 (1997) 144-150.
- [5] Baake, E., and Wagner, H. (2001). Mutation–selection models solved exactly with methods of statistical mechanics. *Genetics Research*, 78(1), 93-117.
- [6] Hermisson, J., Redner, O., Wagner, H., and Baake, E. (2002). Mutation–selection balance: ancestry, load, and maximum principle. *Theoretical population biology*, 62(1), 9-46.
- [7] Manrubia, S. C., Domingo, E., and Lázaro, E. (2010). Pathways to extinction: beyond the error threshold. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences*, 365(1548), 1943–1952.