

Open-Ended Evolution
Experiments - Mathematical models - Computational simulations
LTEE - Power Law - Avida

by Alexander Stein

Mid Term Report
Course: CL717: Evolutionary Dynamics
Instructor: Professor Supreet Saini

Abstract

Fitness is one of the central concepts in evolution. Nevertheless, the investigation of the long-term dynamics of fitness has proved to be complex and remains unclear. In the past years, researchers have been working on experiments and explanations for the boundedness of fitness. They took different approaches that reach from biological and computational experiments to mathematical explanations. Within this course project, we will explore the different approaches that try to answer the question above, investigate their advantages and disadvantages and their relationship to each other. Eventually, we will give an outlook to open questions.

Introduction

Open-end evolution has been studied from two sides. On the one hand, there is biological evolution that has started with Darwin and deals with the explanation of the diversity and development of biological organisms. Experimental difficulties arise since typical time scales of evolution are too large and there are many effects in an ecological system that make research really complicated or almost impossible. The most promising experiment nowadays is the Long-Term Evolution Experiment (LTEE) from Richard Lenski [Michael J. Wiser et.al.] that started in 1988 and has produced more than 50'000s generations of E.Coli in controlled laboratory environment. On the other hand, we have research of open-ended evolution that is coming from digital evolution experiments started in the 1950s. The ultimate goal of this branch is to find a system where novelty and complexity can accumulate indefinitely. The currently most widely used digital evolution system is called Avida, and is used to study a wide range of evolutionary and ecological dynamics in populations of self-replicating computer programs [T.Taylor et.al.].

The focus of this project is on the boundedness of the fitness curve. Theoretically, we expect that the number of beneficial mutations that can occur in an organism is limited, thus the fitness has a maximum value where all possible beneficial mutations are realized. However, experimentally we find results that seem to contradict this thought. The fitness curve seems to be unbounded. In the following, we will provide two models that try to explain the fitness curve over time. [2] We will discuss the methods and results of the LTEE, which is a unique experiment [1] and compare these with the computer simulation done by Michael J. Wiser et.al. [3]. Note that we will use the concept of fitness landscapes of S.Wright as they prove to be a helpful tool for explanations.

Theoretical mathematical model

We study two different mathematical models, namely the Hyperbolic model and the Power Law. They have in common that both of them increase fast quickly in the beginning, but flat off quickly as well. But, they have a crucial difference as well: The Hyperbolic model has a maximum value, a asymptote. It approaches this value as time goes to infinity. In contrast, the Power Law unbounded and goes to infinity as the time goes to infinity. Both models have 2 parameters, such that we can say that both are from the same complexity and a better fit is not due to overfitting. [1] [3]

Hyperbolic Model	Power Law
$\bar{w} = 1 + \frac{at}{t+b}$	$\bar{w} = (bt + 1)^a$
<ul style="list-style-type: none"> - \bar{w} is the mean fitness, t is time in generations and at $t = 0$, we have $\bar{w} = 1$ - a and b are parameter models - $\bar{w} \rightarrow 1 + a$ as $t \rightarrow \infty$ 	<ul style="list-style-type: none"> - \bar{w} is the mean fitness, t is time in generations and at $t = 0$, we have $\bar{w} = 1$ - a and b are parameter models - $\bar{w} \rightarrow \infty$ as $t \rightarrow \infty$

Michael J. Wiser et. al. provide a theoretical explanation for the Power Law. We do not perform the full derivation as this is done really clearly in the [2], but discuss the most important steps and assumptions here. The derivation starts with the definition of three important functions:

$f(x)$ = distribution of available beneficial effects

$\pi(s)$ = probability that a beneficial mutation of effect s survives drift

$\lambda(s)$ = number of other beneficial mutations interfering with a mutation of effect s

$$f(s) = \alpha e^{-\alpha s}$$

$$\pi(s) = 4s$$

$$\lambda(s) = \frac{\mu}{s} N \ln N \int_s^{\infty} \pi(s') f(s') ds'$$

The parameter s is the selection strength, μ is the mutation rate and $1/\alpha$ is the mean advantage. The distribution of $f(s)$ is assumed to be given by the exponential distribution for mathematical convenience as the theory of clonal interference is robust to the exponential distribution. Within the lecture and the discussion of Moran processes, we already derived the expression of $\pi(s)$. Note, that we have an additional factor 4. For the expression of $\lambda(s)$, we refer to Gerish and Lenski [7]

In the following steps of the derivation, averages of the selection strength $\langle s \rangle$ and the time $\langle t \rangle$ are constructed. Particularly important will be the expected values for the advantage and time to fixation for the first fixed beneficial mutation $\langle s_1 \rangle$ and $\langle t_1 \rangle$. We assume that we deal with small selection strength $\langle s \rangle \ll 1$ and further long time approximations are made $\langle t \rangle \gg 1$, which will help us to approximate α .

Next, we assume that the distribution of available benefits declines after a mutation with advantage $\langle s \rangle$ fixes, such that α increases linearly with $\langle s \rangle$:

$$\alpha_{n+1} = \alpha_n(1 + g \langle s_{n+1} \rangle)$$

This g will provide a parameter for the effect of diminishing-returns epistasis. Diminishing-return epistasis is the effect that occurs when a marginal improvement from a beneficial mutation is declined because of increasing fitness. α_n and s_n are α and s of the n 'th fixed mutation.

Following some approximations, such as a Taylor expansion of the exponential function and assuming that $\bar{w}(\alpha)$ can be approximated as a constant after long times, we result the Power Law of the form:

$$\bar{w} = (2g \langle s_1 \rangle e^{g \langle s_1 \rangle \frac{t}{\langle t_1 \rangle}} + 1)^{1/(2g)}$$

To get this expression, we used the effect of diminishing-return epistasis, which is represented by the factor g and clonal interference, which depends on $\langle s_1 \rangle$ and $\langle t_1 \rangle$, which in turn are functions of the population size N , beneficial mutation rate μ and initial mean beneficial effect $1/\alpha_0$. Recall; clonal interference was discussed in class. It refers to effect due to competition among organisms with different beneficial mutations, which impedes their spread in asexual populations.

Note, that we assumed that deleterious mutations do not appreciably affect the dynamics. This assumption is reasonable since deleterious mutations occur at a higher rate than beneficial mutations, but the resulting load is very small relative to the fitness increase.

Biological experiments

The LTEE is a unique experiment that enables us to explore long-term evolutionary dynamics. Fitness experiments with 12 *E. Coli* populations from the same ancestor have been performed over 50'000 generations. The results of the fitness curve are fitted to the two mathematical models. Lenski also compared the predictions the two models make after 20'000 generations. The curves as given in [1] are shown in Figure 1.

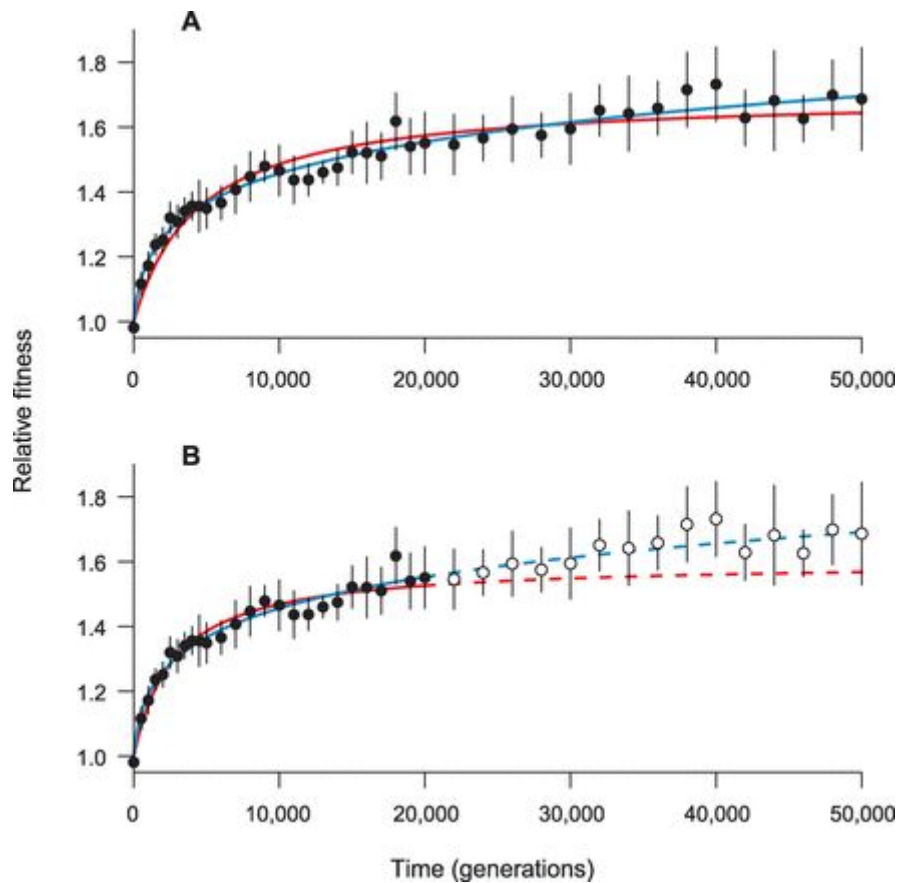


Fig. 1 Comparison of hyperbolic and power-law models.

(A) Hyperbolic (red) and power-law (blue) models fit to the set of mean fitness values (black symbols) from all 12 populations. (B) Fit of hyperbolic (solid red) and power-law (solid blue) models to data from first 20,000 generations only (filled symbols), with model predictions (dashed lines) and later data (open symbols). Error bars are 95% confidence limits based on the replicate populations.

Both fits are good fits. The hyperbolic model has correlation value 0.969 whereas the power law has correlation value 0.986. This leads to a significant BIC difference such that one can say that the power law fit is the better one. Further, if one compares the predictions made by the two models after 20'000 generations, one notes that the hyperbolic model as bounded model underestimates the future, or more precisely the next 30'000 generations. In contrast, the power law predicts the future values of the fitness well.

A further investigation made by Michael J. Wiser et. al. are hypermutation. 6 out of 12 populations evolved hypermutator phenotypes that increased their point-mutation rates approximately 100-fold. 3 of them have been found to become hypermutable early, $t \in (2500, 8500)$ with $t = \text{"number of generations"}$ and had measurable fitness trajectories through at least 30'000 generations. The power law predicts hypermutators to adapt faster. This prediction matches with experimental results. This is one more reasoning that the power law is the better model and that the fitness curve is unbounded.

Computational experiments

In this part, I want to discuss the open-ended evolution research as it is done by side of Artificial Intelligence researchers that approach biological process. The section is based on [3] and [4] and discussed experiments done with Avida (see more in [5] and [6]).

Avida is a digital evolution software platform. It creates digital organisms that reproduce themselves by execution of their instructions. One defines the mutation rate that will replaces, add or delete instructions and leading to variation within the population. Further, one can define the environment, which is done by rewarding certain behaviour of instructions. The environment chosen in [3] is called logic-77, which is a more complex systems and hence leads more likely to a rugged fitness landscape (see discussion in homework 2 and class). Further, logic-77 shows substantial variation across replicate runs. A reward in this environment can be received by performing any of 77 distinct input tasks. Organisms can perform any combination of these tasks, repeating any individual task up to 10 times.

The experiments that is done in this environment is the following. We let 10 populations of digital organisms evolve for 200'000 generations. Then, we extract the most common genotype and let each of these 10 genotypes evolve for another 200'000 generations. The results are fitted to the 2 mathematical models discussed earlier.

As one study the observed plot, one finds quickly that unlike in natural systems, the fitness changes rapidly followed by stagnation. The fitness was fitted to the power law and hyperbolic model for the first 200'000 generations. By comparing the correlations, the hyperbolic law gives the better fit for 10 out of the 10 cases. But, if one continues the experiment for another 200'000 generations, one finds that some of the fitness trajectories exceed the theoretical asymptote obtained by the fit of the first 200'000 generations. This happens for a few trajectories from 5 out of the 10 cases. This indicates that the hyperbolic model as bounded model is invalid. The authors talk about the Boundedness Illusion.

Discussion

In the introduction, we said that there must be theoretical maximal fitness, considering the fact that DNA has a limited complexity it can takes. This holds true particularly for computer models, that deal with limited memory. However, the results in the experiments are show different results. The explanation is that we are far away of saturate the beneficial mutations in our experiments. Digital evolution has some fundamental differences to biological evolution. Nonetheless, it can be an helpful tool to study certain aspects such as open-end evolution.

Up to now, I studied mathematical, computational and experimental methods that try explain open-end evolution. In my further studies, I want to continue this work. I want to deepen my knowledge of Avida (see [5], [6]) and investigate the connections of digital evolution with biological evolution in the aspect of open-end evolution. I plan to estimate and compare typical time scales, figure out the advantages and disadvantages of both approaches and how they extent each other. How close are digital and biological evolution? - If one can

answer this question, one may be enabled to investigate dynamical landscapes with computer simulations, which will be beyond my project, however, it will be the ultimate goal.

References:

1. Michael J. Wiser, Noah Ribeck and Richard E. Lenski: Long-Term Dynamics of Adaptation in Asexual Populations (Science, 2013)
2. Michael J. Wiser, Noah Ribeck and Richard E. Lenski: Supplementary Materials for “Long-Term Dynamics of Adaptation in Asexual Populations” (Science, 2013)
3. Michael J. Wiser, Emily L. Dolseon, Anya Vostinar, Richard E. Lenski and Charles Ofria: The Boundedness Illusion: Projections from early evolution can underestimate evolutionary potential (Third Workshop on Open-Ended Evolution, 2018)
4. Tim Taylor et.al.: Open-Ended Evolution: Perspectives from the OEE Workshop in York (2016)
5. Ofria, C. and Wilke, C.O. Avida; A software Platform for Research in Computational Evolutionary Biology (2004)
6. Ofria, C., Bryson, D.M. and Wilke, C.O. Avida. In Artificial Life Models in Software, pages 3-35 (Springer London, 2009)
7. P.J. Gerish, R.E. Lenski: The fate of competing beneficial mutations in an asexual population (Genetica, 1998)