An individual-based model for the Lenski experiment, and the deceleration of the relative fitness


SPA Moscow, 25. July 2017
The team

Noemi Kurt (TU Berlin)  
Lenski Experiment  
25 July 2017  

Content of the talk

1. The Lenski experiment, and a bit of biological background
2. An individual based model and first insights
3. Relative fitness, main results, and our interpretation
4. Some aspects of the proofs
The Lenski experiment of evolution

- Population of E. coli bacteria in a glucose medium
- (Asexual) reproduction until glucose is deployed
- Sample from the population at the end of the day
- Repeat with sampled population under identical conditions

\[ N \approx 5 \cdot 10^9 \]

\[ N \approx 5 \cdot 10^7 \]

Noemi Kurt (TU Berlin)
## The Lenski experiment of evolution

### Long term experiment

- So far this has been going on for more than 25 years, or \( \approx 60'000 \) generations.
- During this time, the population has *evolved*, that is, adapted to the glucose environment.
- *Samples* of the population have been stored at regular intervals.
- This allows to compare the unevolved founder strain of the population with the evolved present population.

For details on the experiment see e. g. [Lenski, Rose, Simpson, Tadler, American Naturalist, 1991], or the project website http://myxo.css.msu.edu/ecoli
Relative fitness of two strains

Measuring adaptation

- A population of size $A_0$ of the unevolved strain and a population of size $B_0$ of the evolved strain perform a direct competition in the glucose medium.
- The respective population sizes at the end of the day are denoted by $A_1$ and $B_1$.
- The (empirical) relative fitness $F(B|A)$ of strain $B$ with respect to strain $A$ is

$$F(B|A) = \frac{\log(B_1/B_0)}{\log(A_1/A_0)}.$$
Relative fitness over time

Figure from [Lenski, Travisano, PNAS, 1994]

**Fig. 4.** Trajectory for mean fitness relative to the ancestor in one population of *E. coli* during 10,000 generations of experimental evolution. Each point is the mean of three assays. Curve is the best fit of a hyperbolic model.
Relative fitness over time

Figure from [Barrick, Yu, Yoon, Jeong, Oh, Schneider, Lenski, Kim, Nature 2009]
Relative fitness over time

Figure from [Wiser, Ribeck, Lenski, Science express 2013]

Fitted curve: Parabola \( w(t) = (1 + ct)^{1/2g} \), where \( c, g \) are parameters
Challenge

Goal
Understand the shape of the relative fitness curve, in particular the deceleration. Which mechanisms are involved?

Mathematical approach
Define an *individual based* (microscopic) model for the evolution of the bacterial population, and study the (macroscopic) *relative fitness* of the population over time. Show that in the limit of large populations, under a suitable time-rescaling and for a suitable choice of the parameters, the relative fitness process converges to a deterministic function.
Basic mechanisms at work

Mutation and selection

- **Beneficial mutations** add to the reproductive success of an individual
- Beneficial mutations may or may not *fixate* in the population
- Fixation of beneficial mutations lead to an *increase* in the relative fitness of the population

Observations in the Lenski experiment: The relative fitness increases over time, in line with the elementary principles of Darwinian evolution. However, the increase gets slower and slower.

*Why the slowdown?*
Mechanisms of evolution in the Lenski experiment

Possible explanations for the deceleration

- **Clonal interference**: Several mutations interfere with each other, changing their respective probabilities of fixation
- **Epistasis**: Beneficial effects of different mutations depend on each other, “diminishing returns”
- **The design of the experiment**: Daily cycles, limited supply of resources, sampling procedure

In large parts of the literature, clonal interference and epistasis are considered to be the predominant reasons for the observed slowdown.

Our simple model will focus mainly on the design of the experiment.
Content of the talk

1. The Lenski experiment, and a bit of biological background
2. An individual based model and first insights
3. Relative fitness, main results, and our interpretation
4. Some aspects of the proofs
An individual-based mathematical model

Information about the experiment

- At the beginning of each day there are $N$ individuals.
- Within each day, individuals reproduce by binary splitting at a constant rate $r > 0$.
- The reproduction process will stop when the glucose has been consumed, which happens when there are $\approx \gamma N$ individuals, for some $\gamma > 1$.
- $N$ individuals out of the $\approx \gamma N$ are sampled uniformly without replacement, to form the initial population at the next day.

Intraday and interday

- The dynamics has two parts: (Continuous) growth of the population within a day, and (discrete) sampling between days.
An individual-based mathematical model

Figure: The intraday/interday model (Adrián González Casanova)
An individual-based mathematical model

Figure: The intraday/interday model (Adrián González Casanova)
An individual-based mathematical model

**Figure:** The intraday/interday model (Adrián González Casanova)
An individual-based mathematical model

Figure: The intraday/interday model (Adrián González Casanova)
An individual-based mathematical model

One daily cycle, homogeneous population

- Fix parameters $r > 0$ and $\gamma > 1$.
- Let $(Y_t)_{t \geq 0}$ be a pure birth process with rate $r$ started at $Y_0 = N$, that is, a Yule processes with parameter $r$.
- Define a (deterministic) stopping time

$$\sigma = \sigma(\gamma, r) := \inf\{t > 0 : \mathbb{E}[Y_t] = \gamma N\} = \frac{\log \gamma}{r}.$$

- The intraday process is then $(Y_t \wedge \sigma)_{t \geq 0}$.

Sampling rule

At the end of each day, we sample uniformly at random $N$ individuals (out of the $Y_\sigma \approx \gamma N$) to start the population at the beginning of the next day.
Two types of individuals

- Assume that $0 < k < N$ individuals (the mutants) reproduce at rate $r + \varphi_N$, while the other $N - k$ individuals reproduce at rate $r$.
- $\varphi_N > 0$, we assume $\varphi_N \to 0$ as $N \to \infty$.
- Offspring have the reproduction rate of their parent
- Stop the population growth at time $\sigma_k = \sigma_k(r, \gamma)$ when the expected total population size is $\gamma N$
- Sample uniformly at random $N$ individuals for the next day
- We are interested in the interday process

$$(K_i)_{i \in \mathbb{N}_0},$$

where $K_i$ denotes the number of mutants in the population at the beginning of day $i \in \mathbb{N}_0$.

Note: $\sigma_k$ is decreasing in $k$. 
Selective advantage

**Expected number of offspring**

- If every individual has the same reproduction rate, every one of the $N$ individuals at the beginning of day 0 has in expectation one offspring in the population at the beginning of day 1.
- In the two-types model of the previous slide, we have

$$\mathbb{E}[K_1|K_0 = 1] = 1 + \rho_N \frac{\log \gamma}{r} + o(\rho_N).$$

Hence $\rho_N$ is connected to the *selective advantage* of a (mutant) individual. $(K_i)_{i \in \mathbb{N}_0}$ may be thought of as a *slightly supercritical branching process*. 
Selective advantage

More generally,

\[ \frac{1}{k} \mathbb{E}[K_1 | K_0 = k] = 1 + \left(1 - \frac{k}{N}\right) \varrho_N \frac{\log \gamma}{r} + o(\varrho_N). \]

The selective advantage decreases in both \( k \) and \( r \). This reflects the design of the experiment: For fitter populations, the “Lenski days” are shorter.

**Figure:** The biker effect (Adrián González Casanova)
Probability and speed of fixation

- Define the probability of fixation

\[ \pi_N := \mathbb{P}(\exists i \in \mathbb{N} : K_i = N \mid K_0 = 1) \]

- If \( K_0 = 1 \), let

\[ \tau^N := \inf\{i \in \mathbb{N} : K_i \in \{0, N\}\} \]

**Theorem 1 (Probability and speed of fixation)**

*Under the assumptions of our model, as \( N \to \infty \),*

\[ \pi_N \sim \frac{\gamma}{\gamma - 1} \frac{\log \gamma}{r} Q_N. \]

*Moreover, for any \( \delta > 0 \) there exists \( N_\delta \in \mathbb{N} \) such that for all \( N \geq N_\delta \)*

\[ \mathbb{P}(\tau^N > Q_N^{-1 - 3\delta}) \leq (7/8) Q_N^{-\delta}. \]

*morally, \( \tau^N \approx \frac{1}{Q_N} \) with very high probability, as \( N \to \infty \).*
Content of the talk

1. The Lenski experiment, and a bit of biological background
2. An individual based model and first insights
3. Relative fitness, main results, and our interpretation
4. Some aspects of the proofs
The weak mutation - moderate selection model

Now look at the basic model over long time scales, where many mutations may occur over time, and go to fixation/extinction.

<table>
<thead>
<tr>
<th>Strength of mutation and selection (Assumption A)</th>
</tr>
</thead>
<tbody>
<tr>
<td>i) Beneficial mutations add $\varrho_N$ to the reproduction rate of the individual that suffers the mutation.</td>
</tr>
<tr>
<td>ii) In each generation, with probability $\mu_N$ there occurs a beneficial mutation. The mutation affects only one (uniformly chosen) individual, and every offspring of this individual also carries the mutation.</td>
</tr>
<tr>
<td>iii) There exists $0 &lt; b &lt; 1/2$, and $a &gt; 3b$, such that $\mu_N \sim N^{-a}$ and $\varrho_N \sim N^{-b}$ as $N \to \infty$.</td>
</tr>
</tbody>
</table>

This implies

$$\mu_N \ll \varrho_N, \quad \text{resp.} \quad \mu_N^{-1} \gg \varrho_N^{-1},$$

which allows us to exclude *clonal interference* with high probability.
The process of relative fitness

Let $R_{i,j}, j = 1, \ldots, N$ denote the reproduction rates of the individuals present at the beginning of day $i$, and assume $R_{0,j} \equiv r_0$.

We define the fitness of the population at the beginning of day $i$ relative to the initial population of day 0 as

$$F_i := \frac{\log \frac{1}{N} \sum_{j=1}^{N} e^{R_{i,j}u}}{\log e^{r_0u}}$$

where $u$ is a given time for which the two populations are allowed to grow together.

Relative fitness in homogeneous populations

If at day $i$ the reproduction rate within the population is constant and equal to $R_i$, then

$$F_i = \frac{R_i}{r_0}.$$
The fitness process under Assumption A

Assumption A and Theorem 1 show that the fitness process looks like this:

In particular, we can treat the mutations successively, they don’t interact:

*No clonal interference*
The limiting fitness process

Theorem 2 (Convergence of the relative fitness process)

Assume \( R_{0,j} = r_0 \) for \( j = 1, \ldots, N \), and let \((F_i)_{i \in \mathbb{N}_0}\) be the process of relative fitness. Then under Assumption A, the sequence of processes \((F_{\lfloor (Q_N^2 \mu_N)^{-1} t \rfloor})_{t \geq 0}\) converges in distribution as \( N \to \infty \) locally uniformly to the deterministic function

\[
f(t) = \sqrt{1 + \frac{\gamma \log \gamma}{\gamma - 1} \frac{2t}{r_0^2}}, \quad t \geq 0.
\]

The time scale \( Q_N^{-2} \mu_N^{-1} \) arises naturally, since mutations arrive at rate \( \mu_N \), and fixate with probability \( Q_N \), increasing the reproduction rate by \( Q_N \).
Discussion

- The limiting fitness curve in Theorem 2 is a parabola.
- Our model does not a priori include epistasis.
- Due to the design of the experiment (shorter days due to increasing fitness) there is a resulting slowdown effect in the model.

Figure: Limiting curve, data points, and simulations (Sebastian Probst)
Including epistasis

Assume that the *benefit* of a successful mutation that goes to fixation *depends on the current fitness*. For example, motivated by [Wiser et al. 2013], we assume that a mutation that goes to fixation when the relative fitness is \( x \geq 1 \) provides an increment of the form

\[
\mathcal{Q}_N(x) = x^{-q} \mathcal{Q}_N
\]

to the reproduction rate, for some \( q > -1 \).

**Corollary 3 (Relative fitness with epistasis)**

*In this case, under Assumption A, the sequence of processes \( (F_{t\leq t|\mathcal{Q}_N^{-1}N^\muN^{-1}}} t \geq 0 \) converges in distribution as \( N \to \infty \) locally uniformly to the deterministic function*

\[
h(t) = \left(1 + \frac{2(1 + q)C(\gamma)}{r_0^2} t\right)^{\frac{1}{2(1+q)}}, \ t \geq 0.
\]
Fitness curve with epistasis

Figure: Epistasis included (Simulations and figure by Sebastian Probst)

Comparison of the theoretical results with data and simulations will be provided by [Baake, González Casanova, Probst, Wakolbinger, in preparation]
Content of the talk

1. The Lenski experiment, and a bit of biological background
2. An individual based model and first insights
3. Relative fitness, main results, and our interpretation
4. Some aspects of the proofs
Some aspects of the proof

The mathematical work lies in the proof of Theorem 1:

**Theorem 1 (Probability and speed of fixation)**

Under the assumptions of our model, as $N \to \infty$,

$$\pi_N \sim \frac{\gamma}{\gamma - 1} \frac{\log \gamma}{r} \rho_N.$$ 

Moreover, for any $\delta > 0$ there exists $N_\delta \in \mathbb{N}$ such that for all $N \geq N_\delta$

$$\mathbb{P}(\tau^N > \rho_N^{1-3\delta}) \leq (7/8)^{\frac{1}{\rho_N} \delta}.$$
Proof of Theorem 1

Consider the situation of a successful mutation, i.e. eventually $K_i = N$. Starting with $K_0 = 1$, the process $(K_i/N)_{i \in \mathbb{N}}$ undergoes three phases:
Lemma

For any $0 < \varepsilon < 1/2$, for any $k \geq \varepsilon N$ we have

$$\lim_{N \to \infty} P_k(\exists i : K_i \geq (1 - \varepsilon)N) = 1,$$

and the time it takes to reach $(1 - \varepsilon)N$ is at most of order $\varrho_N^{-1-\delta}$ (with probability 1 as $N \to \infty$, for any $\delta > 0$).

Proof: Show by direct calculations that the generator of $(K_{[\varrho_N^{-1}t]/N})_{t \geq 0}$ converges to the generator of a deterministic increasing function on $(0, 1)$. 
Phase 1

- Assume that a Lenski-day is started with $k$ mutants.
- Assume at the end of the day there are $M = M(k)$ mutants and $Z = Z(k)$ non-mutants.
- By construction, the number of offspring of one mutant at the end of the day is $\sim \text{Geo}(e^{-(r+\rho N)\sigma_k})$
- Let $\Gamma := \frac{M+Z}{N}$ ($\approx \gamma$).
- Given $\Gamma$, a given individual will be sampled with probability $1/\Gamma$.

### Difficulty

*Exchangeable but not independent sampling!*

Independent sampling + independent reproduction = Galton-Watson process.
Phase 1: Coupling to Galton-Watson processes

- Let $\overline{M} \sim \text{Geo}(e^{-(r+\varepsilon N)\sigma_0})$, and $\underline{M} \sim \text{Geo}(e^{-(r+\varepsilon N)\sigma_{\varepsilon N}})$.
- For $1 \leq k \leq \varepsilon N$, this implies $\underline{M} \leq M \leq \overline{M}$.
- Fix $\alpha > 0$. Let $(\overline{K}_i)_{i \in \mathbb{N}}$ be the Galton-Watson process whose offspring distribution is mixed binomial with parameters $\overline{M}$ and $\frac{1}{\gamma} + N^{-\alpha}$, and let $(\underline{K}_i)_{i \in \mathbb{N}}$ be the Galton-Watson process whose offspring distribution is mixed binomial with parameters $\underline{M}$ and $\frac{1}{\gamma} - N^{-\alpha}$.

Lemma (Coupling)

Let $K_0 = 1$. We can couple $(K_i)_{i \in \mathbb{N}_0}, (\overline{K}_i)_{i \in \mathbb{N}_0}$ and $(\underline{K}_i)_{i \in \mathbb{N}_0}$ such that

$$K_i \leq \underline{K}_i \leq \overline{K}_i$$

for all $i \leq \inf\{j : K_j \geq \varepsilon N\}$, with probability exponentially close to 1 as $N \to \infty$.

This means that throughout phase 1, our process $(K_i)_{i \in \mathbb{N}_0}$ can be “sandwiched” between two Galton-Watson processes.
Phase 1: Extinction of near-critical Galton-Watson processes

\((\overline{K}_i)_{i \in \mathbb{N}_0}\) and \((K_i)_{i \in \mathbb{N}_0}\) are near-critical Galton-Watson processes:

- \(E_1[\overline{K}_1] = 1 + \frac{\log \gamma}{r} \varrho_N + o(\varrho_N)\)
- \(E_1[K_1] = 1 + \frac{\log \gamma}{r} (1 - \varepsilon) \varrho_N + o(\varrho_N)\)
- The offspring variances are \(\frac{2(\gamma-1)}{\gamma} (1 + O(\varrho_N))\).


As \(N \to \infty\), the survival probability of \((\overline{K}_i)_{i \in \mathbb{N}_0}\) is

\[
\frac{\gamma}{\gamma - 1} \frac{\log \gamma}{r} \varrho_N + o(\varrho_N),
\]

and the time to either extinction or reaching \(\varepsilon N\) is of order at most \(\varrho_N^{-1-\delta}\). Similarly for \((K_i)_{i \in \mathbb{N}_0}\).
Proof of Theorem 1: Recap

- Phase 2 is dealt with by a straightforward ODE approximation.
- Phase 1 and Phase 3 can be taken care of by a coupling with suitable near-critical Galton-Watson processes.
- For the Galton-Watson processes in question, the probability of fixation and the time until fixation can be calculated.
- The difficult part is the construction of the coupling: Need to take dependence due to the sampling rule (without replacement) and the stopping rule (shorter days) into account.
Summary

- An individual based (microscopic) model for an evolving population in the set up of the Lenski experiment
- Macroscopic quantity: Relative fitness
- Convergence to a power function, qualitative behaviour in agreement with data
- Slowdown effect due to the design of the experiment contributes to the observed power law behaviour
Thank you!