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

Noemi Kurt

Joint work with

Adrián González Casanova, Anton Wakolbinger and Linglong Yuan

$H_\rho \chi \kappa \lambda \varepsilon \nu$, 21 July 2015
The Lenski experiment of long term 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

$N \approx 5 \cdot 10^9$

$N \approx 5 \cdot 10^7$

$N \approx 5 \cdot 10^8$

$N \approx 5 \cdot 10^6$

Noemi Kurt (TU Berlin) Lenski Experiment 21 July 2015
The Lenski experiment of long term evolution

Long term experiment

- So far this has been going on for more than 25 years, that is $\approx 60'000$ generations.
- Due to natural selection, the population has evolved, that is, adapted to the 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!
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

[Lenksi, 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

[**Wiser, Ribeck, Lenski, Science express 2013**]

Fig. 2. 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.

\[
w(t) = (1 + ct)^{1/2g}
\]
Mathematical challenge

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

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 of evolution

**Darwin: 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 the biological literature, clonal interference and (in particular) **epistasis** are considered to be the predominant reasons for the observed slowdown.
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 uniformly sampled 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
An individual-based mathematical model
An individual-based mathematical model
An individual-based mathematical model
An individual-based mathematical model

One daily cycle, homogeneous population

- Fix parameters $r > 0$ and $\gamma > 1$.
- For $i \in \mathbb{N}$, let $(Y_t)_{t \geq 0}$ be the pure birth process with rate $r$ started at $Y_0 = N$, that is, a Yule process 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 $r + \varrho_N$, while the other $N - k$ individuals reproduce at rate $r$.
- $\varrho_N > 0$, we assume $\varrho_N \to 0$ 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

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

Hence \( \varrho_N \) is connected to the \textit{selective advantage} of a (mutant) individual. 
\( (K_i)_{i \in \mathbb{N}_0} \) should be thought of as a \textit{slightly supercritical branching process}. 

Noemi Kurt (TU Berlin)  
Lenski Experiment  
21 July 2015  
14 / 26
Selective advantage

Expected number of offspring

In the two-types model we have

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

More generally,

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

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

- Define the *probability of fixation*\
  \[
  \pi_N := \mathbb{P}\left( \exists i \in \mathbb{N} : K_i = N \mid K_0 = 1 \right)
  \]

- 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 \( N \to \infty \),*

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

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

\[
\mathbb{P}(\tau^N > \varrho_N^{1-3\delta}) \leq \left(\frac{7}{8}\right)^{\varrho_N^{-\delta}}.
\]
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.

Strength of mutation and selection (Assumption A)

i) Beneficial mutations add $\varrho_N$ to the reproduction rate of the individual that suffers the mutation.

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.

iii) There exists $0 < b < 1/2$, and $a > 3b$, such that $\mu_N \sim N^{-a}$ and $\varrho_N \sim N^{-b}$ as $N \to \infty$.

This implies

$$\mu_N \ll \varrho_N,$$

resp.

$$\mu_N^{-1} \gg \varrho_N^{-1},$$

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

$$F_i := \frac{\log \frac{1}{N} \sum_{j=1}^{N} e^{R_{i,j} t}}{\log e^{r_0 t}}$$

where $t$ 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:

\[ x + \frac{q_N}{r_0} \]

In particular, we can treat the mutations successively, they don’t interact.
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 $(F_{\lfloor (\varrho_N^2 \mu_N)^{-1} t \rfloor})_{t \geq 0}$ converges in distribution as $N \to \infty$ locally uniformly to the deterministic function

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

The time scale $\varrho_N^{-2} \mu_N^{-1}$ arises naturally, since mutations arrive at rate $\mu_N$, and fixate with probability $\varrho_N$, increasing the reproduction rate by $\varrho_N$. 
Qualitatively, the curve looks like in [Wiser et al. 2013]

- Our model does not include epistasis in the intraday part
- Due to the design of the experiment (shorter days due to increasing fitness) there is a resulting epistatic effect in the interday part of the model
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:

![Diagram showing three phases of a process](image)
Three phases

- Phase 2 is easy to take care of 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
- No epistasis in the intraday population model, epistatic effect due to the design of the experiment, leads to the observed power law behaviour
Eυχαριστώ πολύ!