

The Quasispecies model.

Maxime Berger

Ecole normale supérieure

April 27, 2021

Evolution is a balance

Mutation

- ▶ Constantly discover new behaviors.
- ▶ Too much mutation and the fittest species kills itself when reproducing.

Selection

- ▶ Sort good behavior from bad behavior.
- ▶ Too much selection and the population can't adapt to changing environments.

A mathematical population

Genome of individuals: a finite chain of ℓ digits: 0 or 1.

A mathematical population

Genome of individuals: a finite chain of ℓ digits: 0 or 1.

Mutations Every digit of its genome is changed with probability q , independently of others.

A mathematical population

Genome of individuals: a finite chain of ℓ digits: 0 or 1.

Mutations Every digit of its genome is changed with probability q , independently of others. Chain u mutates to chain v with rate

$$M_{uv} = q^{d_H(u,v)}(1 - q)^{\ell - d_H(u,v)}$$

Selection Only one sequence, $0 \cdots 0$ has a fitness $\sigma > 1$, the **master sequence**. All others have fitness 1.

The Eigen model. 1971

We follow molecules evolving in a primitive soup.

to create the chain u , every chain v play their part according to

- ▶ their concentration x_v in the population
- ▶ their fitness $f(v)$
- ▶ the rate of mutation M_{vu}

The Eigen model. 1971

We follow molecules evolving in a primitive soup.

to create the chain u , every chain v play their part according to

- ▶ their concentration x_v in the population
- ▶ their fitness $f(v)$
- ▶ the rate of mutation M_{vu}

The Eigen System:

$$x'_u(t) = \sum_v f(v)x_v(t)M_{vu} - x_u(t)\lambda(t),$$

λ is the mean fitness of the population

A formula for the mean fitness

For all ℓ : size of genomes,

For all q : parameter for mutation,

$$\frac{1}{2^\ell} \sum_{k=0}^{\ell} \binom{\ell}{k} \frac{1}{\frac{\lambda}{(1-2q)^k} - 1} = \frac{1}{\sigma - 1}.$$

The critical parameter

In the asymptotic regime $\ell \rightarrow \infty$, $q \rightarrow 0$, and ℓq stays bounded.

The critical parameter

In the asymptotic regime $\ell \rightarrow \infty$, $q \rightarrow 0$, and ℓq stays bounded.

There exists an error threshold separating two regimes

$$\ell q^* \rightarrow \ln \sigma ,$$

where ℓ is the length of the genome of individuals,
and σ the fitness of master sequences.

The critical parameter

In the asymptotic regime $\ell \rightarrow \infty$, $q \rightarrow 0$, and ℓq stays bounded.

There exists an error threshold separating two regimes

$$\ell q^* \rightarrow \ln \sigma ,$$

where ℓ is the length of the genome of individuals,
and σ the fitness of master sequences.

- ▶ If $\lim \ell q > \ln \sigma$, the population is uniform.
It is a neutral regime.

The critical parameter

In the asymptotic regime $\ell \rightarrow \infty$, $q \rightarrow 0$, and ℓq stays bounded.

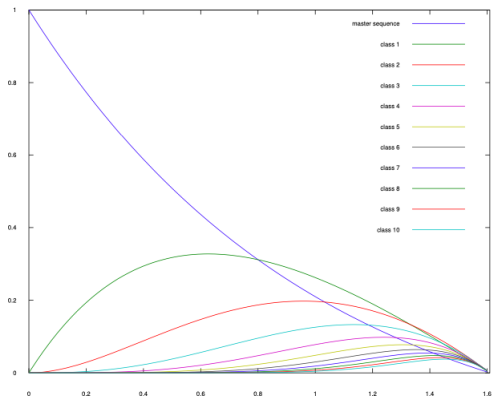
There exists an error threshold separating two regimes

$$\ell q^* \rightarrow \ln \sigma,$$

where ℓ is the length of the genome of individuals,
and σ the fitness of master sequences.

- ▶ If $\lim \ell q > \ln \sigma$, the population is uniform.
It is a neutral regime.
- ▶ If $\lim \ell q < \ln \sigma$, master sequences occupy a positive proportion of the population. It is the **Quasispecies**.

The distribution of the quasispecies



Distribution of hamming classes at equilibrium, for $\sigma = 5$.
 x -axis is the limit of the product lq .
 y -axis is the proportion of individuals

The discrete model

Let us consider now a finite population, composed of m individuals.

The discrete model

Let us consider now a finite population, composed of m individuals.

At each generation, the population follow 3 steps.

- ▶ A parent is chosen in the population, master sequences are chosen preferentially with parameter σ .

The discrete model

Let us consider now a finite population, composed of m individuals.

At each generation, the population follow 3 steps.

- ▶ A parent is chosen in the population, master sequences are chosen preferentially with parameter σ .
- ▶ The genome of the parent is copied, then undergoes mutations.

The discrete model

Let us consider now a finite population, composed of m individuals.

At each generation, the population follow 3 steps.

- ▶ A parent is chosen in the population, master sequences are chosen preferentially with parameter σ .
- ▶ The genome of the parent is copied, then undergoes mutations.
- ▶ The offspring replaces an individual uniformly chosen in the population.

The two phases

Through time, two phases alternate:

The two phases

Through time, two phases alternate:

- ▶ **A Neutral phase**, without any master sequence.
Lasts for 2^ℓ generations. No selection.

The two phases

Through time, two phases alternate:

- ▶ **A Neutral phase**, without any master sequence. Lasts for 2^ℓ generations. No selection.
- ▶ **A Quasispecies phase**, where master sequences occupy a positive proportion of the population. Selection is strong.

The two phases

Through time, two phases alternate:

- ▶ **A Neutral phase**, without any master sequence. Lasts for 2^ℓ generations. No selection.
- ▶ **A Quasispecies phase**, where master sequences occupy a positive proportion of the population. Selection is strong.

A critical point of the model is reached when the duration of both phases are of the same order.

What happen in the population

How to compute the number of master sequences ?

What happen in the population

How to compute the number of master sequences ?

- ▶ Master sequences reproduce with rate r_0 ,

$$r_0 = \frac{\sigma(1 - q)^\ell - 1}{\sigma - 1}$$

What happen in the population

How to compute the number of master sequences ?

- ▶ Master sequences reproduce with rate r_0 ,

$$r_0 = \frac{\sigma(1 - q)^\ell - 1}{\sigma - 1}$$

- ▶ Other individuals can create master sequences with good mutations. Hard to compute !

Estimates on the survival time

If X_n stands for the number of master sequences at generation n ,
What we need to compute is the expectation of

$$\tau_0 = \inf \left\{ n > 0 : X_n = 0 \mid X_0 = 1 \right\}.$$

Estimates on the survival time

If X_n stands for the number of master sequences at generation n ,
What we need to compute is the expectation of

$$\tau_0 = \inf \left\{ n > 0 : X_n = 0 \mid X_0 = 1 \right\}.$$

- ▶ **Lower bound.** No non-master sequence individual can give birth to a master sequence.

Estimates on the survival time

If X_n stands for the number of master sequences at generation n ,
What we need to compute is the expectation of

$$\tau_0 = \inf \left\{ n > 0 : X_n = 0 \mid X_0 = 1 \right\}.$$

- ▶ **Lower bound.** No non-master sequence individual can give birth to a master sequence.
- ▶ **Upper bound.** Every non-master sequence individual can give birth to a master sequence as if it were in the first hamming class.

Comparison of two times

The two bounds give the same estimation

$$E(\tau_0) \sim \exp\left(\frac{\sigma - 1}{2} m r_0^2\right).$$

with

$$r_0 = \frac{\sigma(1 - q)^\ell - 1}{\sigma - 1}$$

Time of the neutral phase was $\sim 2^\ell$,

Comparison of two times

The two bounds give the same estimation

$$E(\tau_0) \sim \exp\left(\frac{\sigma - 1}{2} m r_0^2\right).$$

with

$$r_0 = \frac{\sigma(1 - q)^\ell - 1}{\sigma - 1}$$

Time of the neutral phase was $\sim 2^\ell$,

Comparing the two times leads to a development of a critical parameter

$$q^* = \frac{\ln \sigma}{\ell} - \frac{\sqrt{2(\sigma - 1) \ln 2}}{\sqrt{\ell m}} + \dots$$

Another possible definition

τ_0 is the survival time of master sequences

$$E(\tau_0) \sim \exp\left(\frac{\sigma - 1}{2} m r_0^2\right).$$

with

$$r_0 = \frac{\sigma(1 - q)^\ell - 1}{\sigma - 1}$$

The point at which this time grows smaller than a polynomial (or stays bounded) leads to

$$q^* = \frac{\ln \sigma}{\ell} - \frac{C}{\ell \sqrt{m}} + \dots$$

Thank you for attention.



Manfred Eigen.



M. Berger, *Expansion of error thresholds in the Moran model*, 2019.