Time scales of evolutionary trajectories w.r.t sequence length

Andreas Pavlogiannis

January 4, 2014
General setting

- Evolution of populations by means of mutation and selection.
  - Q: How long until some target region is found?
- Existing:
  - Single mutation: fixation time of a single mutant.
  - Multiple mutations: beneficial mutations arising at a constant rate.
- Interplay between selection and mutation.
  - Beneficial mutations become rarer along the way.
- Static fitness landscape.
In a population of $N$ type A individuals, a mutant B is introduced with selective advantage $r$. At each time step:

1. An individual is chosen uniformly at random to die.
2. An individual is chosen proportionally to its fitness to divide.

A single mutant is fixed with probability $\rho = \frac{1-1/r}{1-1/r^N}$.
### Population

- **L**: sequence length (from some alphabet \( \{0, \ldots, \kappa\} \))
- **N**: population size
- **u**: point mutation probability \((N \cdot u < 1)\)
- **t = \vec{0}**: ideal sequence
- **c**: determines target set of sequences \(\tau\): \(|t - \tau|_H \leq c \cdot L\)
- **s**: fitness slope extends to sequences \(\tau\): \(|t - \tau|_H \leq s \cdot L\)
- **r**: fitness factor \(f_{i-1} = r \cdot f_i\), where \(f_i\) is the fitness of all sequences \(\tau\): \(|t - \tau|_H = i\)
All together

Hamming Distance

- Target set
- Fitness landscape

0  cL  sL  L
Homogeneous population on a hypercube. At each time point:
- A point mutation might occur (either beneficial or deleterious)
- Moran dynamics determine the fixation probability of the new mutant
- Mutant either takes over the population, or swept out

**Hitting Time**

For Markov Chain $M$ and states $n_1$, $n_2$, denote with $H_M(n_1, n_2)$ the expected hitting time of $n_1$ from $n_2$.

$$H_M(n_1, i) = 1 + \sum_j \delta(i, j) \cdot H_M(n_1, j)$$

**Q:** What is the expected hitting time of the target set, as a function of the genome length $L$?
From $L$ dimensions to Markov Chain on a line

\[ \delta(i, i - 1) \delta(i, i) \delta(i + 1, i + 1) \]

\[ \delta(i - 1, i - 2) \delta(i - 1, i) \delta(i + 1, i + 2) \]

\[ \forall i \in \mathbb{N} \]

\[ \delta(i, i + 1) = \frac{L - i}{i} \kappa \cdot r^{-(N-1)} \]  

\[ \delta(i, i) \leq k \text{ (constant - indp of } L) \]
The unloop variant $\overline{M}$ of a Markov Chain on a line $M$ ignores self-loops.

$1 - a - b$

Hitting times on the unloop variant

1. $H_{\overline{M}}(n_1, n_2) \leq H_M(n_1, n_2)$.
2. $H_M(n_1, n_2) \leq z^* \cdot H_{\overline{M}}(n_1, n_2)$, where $z^* = \max_{0 < i \leq L} \frac{1}{1 - \delta(i, i)}$.

For $z^*$ constant, $H_{\overline{M}}(n_1, n_2) = \Theta(H_M(n_1, n_2))$
Hitting times on unloop variants

Lower & upper bound

1. For states \( n_1 \leq x < n_2 \leq y \) and \( y = x + k \), if \( \delta(i, i+1) \geq A < \frac{1}{2} \) for all \( x < i < y \), then \( H_{\overline{M}}(n_1, n_2) = 2^k \cdot \Omega(L) \).

2. For states \( n_1 \leq n_2 \), if \( \delta(i, i-1) \geq \frac{1}{2} \) for all \( n_1 \leq i \leq L \), then \( H_{\overline{M}} = O(L^2) \).
(b): Dichotomy, $c \cdot (1 + \frac{r^{N-1}}{\kappa}) < 1$
(b): Dichotomy, \[ c \cdot \left(1 + \frac{r^{N-1}}{\kappa}\right) < 1 \]

Remains exponential for polynomially many population replicates
Dichotomy in numbers

Hitting times for $N=100$ and $\kappa=1$

- $c \cdot \left(1 + \frac{r^{N-1}}{\kappa}\right) = 0.67$
- $c \cdot \left(1 + \frac{r^{N-1}}{\kappa}\right) = 0.73$
- $c \cdot \left(1 + \frac{r^{N-1}}{\kappa}\right) = 0.8$
- $c \cdot \left(1 + \frac{r^{N-1}}{\kappa}\right) = 1.0$
For large $N$, $s$ behaves as $c$.

$$s = \frac{1}{2}$$

![Graph showing discovery time in generations vs. sequence length $L$. Legend indicates different values of $N$ and $c$.](image-url)
Bounded selection in the Wright-Fisher model

\[ N=10^4, \ r=1.01, \ s=\frac{1}{2} \]
Bounded selection in the Moran model

\[ s = \frac{1}{2}, \ r = 1.01 \]
Randomly distributed targets

- \( m << (\kappa + 1)^L \) targets distributed uniformly at random on the \( L \)-dimensional space
- Each surrounded by a fitness slope extending at most to \( s \cdot L, s < \frac{\kappa}{\kappa + 1} \)
The Hamming distance of the origin from a target follows $\text{Binomial}(L, \frac{\kappa}{\kappa + 1})$.

$P[|\tau - t|_H \leq sL] = 2^{-O(L)}$.

- **Hoeffding’s inequality**: In a Binomial process, the probability of deviation from the expectation drops exponentially.

By union, the probability to fall in any of the $m$ Hamming spheres is $p < m \cdot 2^{-O(L)}$.

The process repeats and iterations follow $\text{Geometric}(p)$.

Expectation: $\frac{2^{\Omega(L)}}{m}$ (with high probability)
Randomly distributed targets

\[ \kappa = 3 \]
$L^{k+1}$ regenerations suffice to hit the target set in $O(L^{k+1}/u)$ expected time, with probability at least $1 - e^{-L}$.

- Every regeneration hits the target in $k$ steps with probability at least $L^{-k}$. 
In all cases of exponential lower bounds, the hitting time is also exponentially upper bounded.

Then for $\propto \frac{L}{\log L}$ independently evolving loci, hitting times are polynomial in $L$.

- Assuming once target is hit it remains fixed.