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

Andreas Pavlogiannis

January 4, 2014

General setting

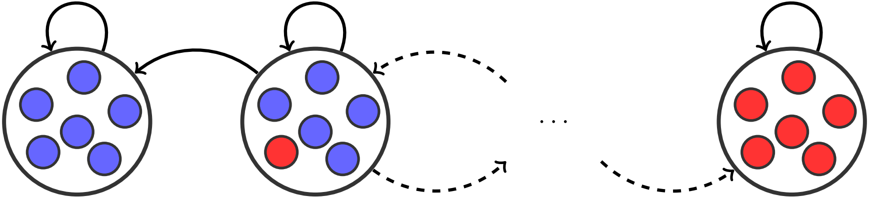
- Evolution of populations by means of mutation and selection.
    - **Q:** How long until some target region is found?
- Existing:
    - Single mutation: fixation time of a single mutant.
    - Multiple mutations: beneficial mutations arising at a constant rate.
- Interplay between selection and mutation.
    - Beneficial mutations become rarer along the way.
- Static fitness landscape.

Moran process

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 $\rho = \frac{1-1/r}{1-1/r^N}$

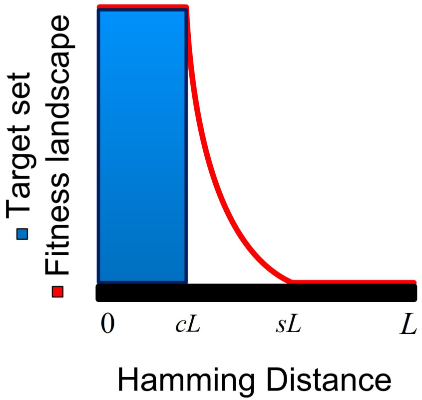


Model

Population

- **L** : sequence length (from some alphabet $\{0, \ldots \kappa\}$)
- **N** : population size
- **u** : point mutation probability ($N \cdot u < 1$)
- **t** $= \overrightarrow{0}$: ideal sequence
- **c** : determines target set of sequences τ: $|t - \tau|_H \leq c \cdot L$
- **s** : fitness slope extends to sequences τ: $|t - \tau|_H \leq s \cdot L$
- **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



Process

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

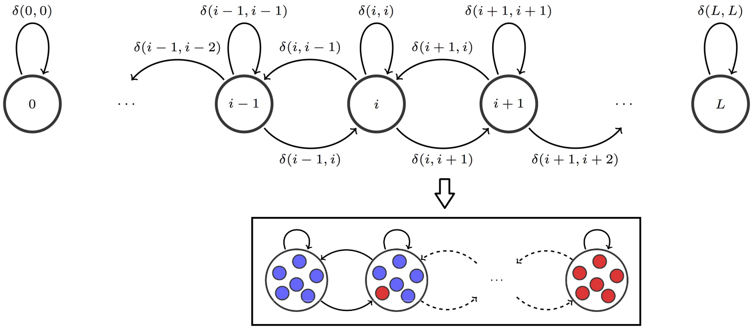
Hitting Time

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?

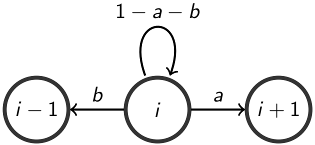
From L dimensions to Markov Chain on a line

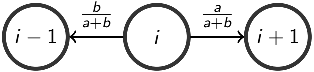


(1) $\dfrac{\delta(i,i+1)}{\delta(i,i-1)} = \dfrac{L-i}{i}\kappa \cdot r^{-(N-1)}$ (2) $\delta(i,i) \leq k$ (constant - indp of L)

Unlooping

The unloop variant $\overline{M}$ of a Markov Chain on a line M ignores self-loops.





Hitting times on the unloop variant

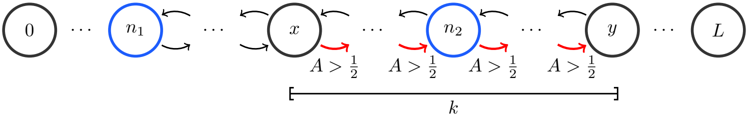
1. $H_{\overline{M}}(n_1, n_2) \leq H_M(n_1, n_2)$.
2. $H_M(n_1, n_2) \leq z^* \cdot H_{\overline{M}}(n_1, n_2)$, where $z^* = \max_{0<i\leq L} \frac{1}{1-\delta(i,i)}$.

$$\text{For } z^* \text{ constant, } H_{\overline{M}}(n_1, n_2) = \Theta\left(H_M\left(n_1, n_2\right)\right)$$

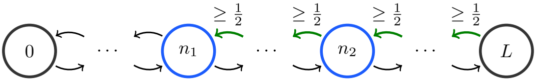
Hitting times on unloop variants

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 $H_{\overline{M}} = O(L^2)$.

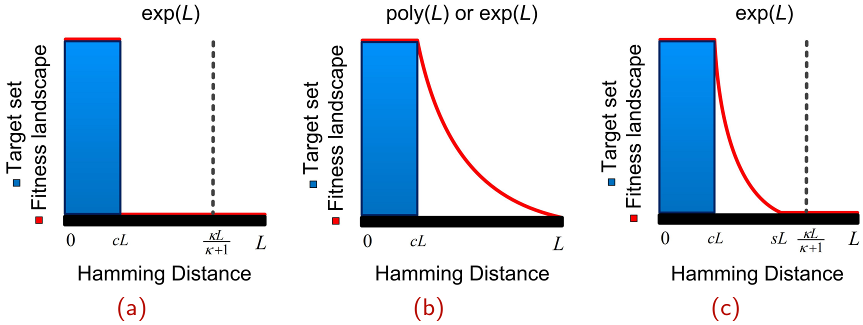


(a)



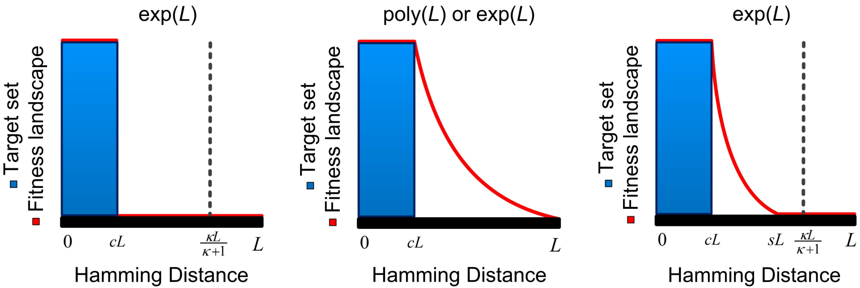
(b)

Broad peak



(b): Dichotomy, $c \cdot (1 + \frac{r^{N-1}}{\kappa}) < 1$

Broad peak

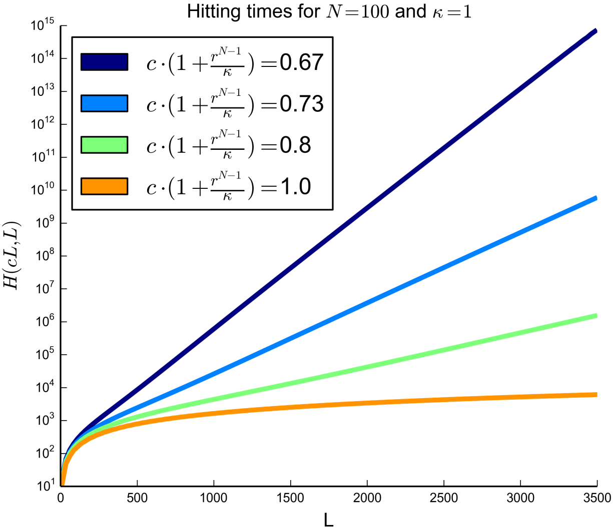


(a) (b) (c)

(b): Dichotomy, $c \cdot (1 + \frac{r^{N-1}}{\kappa}) < 1$

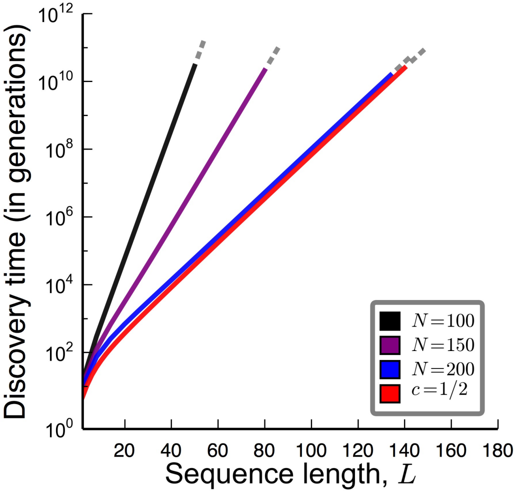
Remains exponential for polynomially many population replicates

Dichotomy in numbers

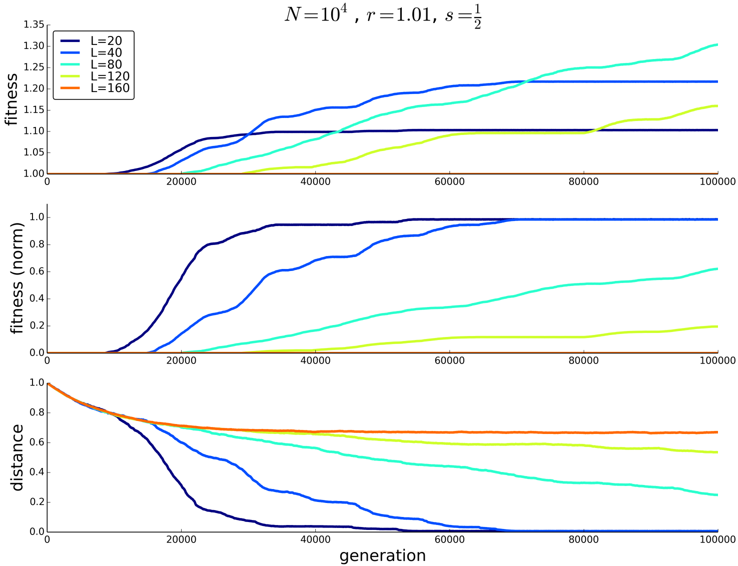


For large N, s behaves as c

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

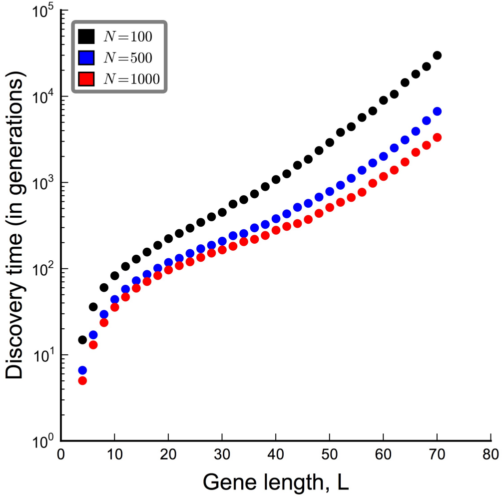


Bounded selection in the Wright-Fisher model



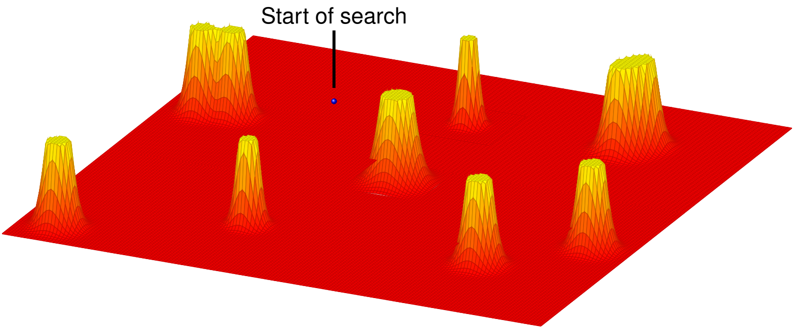
Bounded selection in the Moran model

$$s = \frac{1}{2},\ r = 1.01$$



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 $s \cdot L$, $s < \frac{\kappa}{\kappa+1}$

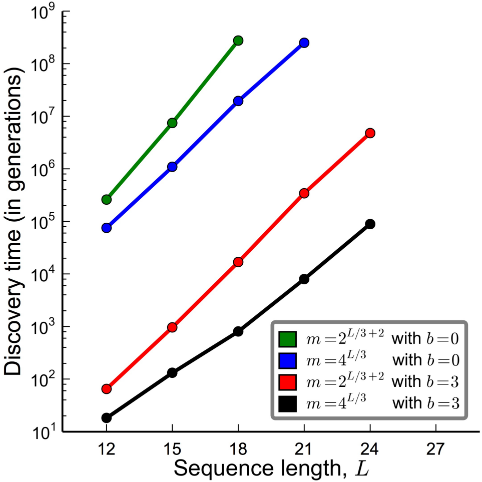


Randomly distributed targets

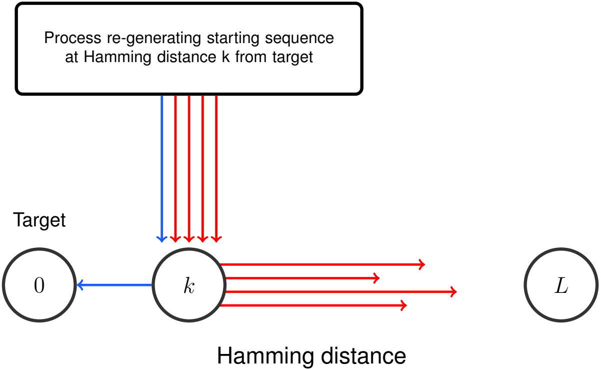
1. The Hamming distance of the origin from a target follows $\textit{Binomial}\left(L, \frac{\kappa}{\kappa+1}\right)$.
2. $P\left[|\tau - t|_H \leq sL\right] = 2^{-O(L)}$.
   - Hoeffding's inequality: In a Binomial process, the probability of deviation from the expectation drops exponentially.
3. By union, the probability to fall in any of the m Hamming spheres is $p < m \cdot 2^{-O(L)}$.
4. The process repeats and iterations follow $\textit{Geometric}(p)$.
5. Expectation: $\frac{2^{\Omega(L)}}{m}$ (with high probability)

Randomly distributed targets

$\kappa = 3$



Poly: Regeneration process



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

- Every regeneration hits the target in k steps with probability at least L^{-k}.

Poly: Independently evolving loci

- 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.