Evolution through Programming

Lecturer: Prof. Yitzhak Pilpel
TA: Omer Kerner – omer.kerner@weizmann.ac.il

Assignment 2 – Moran Process

Choose ONE of the following questions. You may create a simulation in any programming language (e.g., Python, MATLAB, Java). Attach your code with your submission and be sure to include clear explanations. Enjoy THE Process :)

Question 1 – Moran Process with a twist

The Moran Process is a fundamental stochastic model of evolutionary dynamics in finite populations. In this assignment, you will implement a computational simulation of the Moran Process, verify its behavior against theoretical predictions, and then design and conduct your own creative investigation into the process's dynamics.

Objective:

Explore the dynamics of selection and drift by simulating the Moran Process. This assignment invites you to think about evolution in a fixed sized population and to test how different rules and parameters balance the forces of selection and drift in a diverse population.

For additional background refer to "Evolutionary Dynamics" by Martin Nowak, Chapter 6 – Finite Populations

Tasks:

- **Simulation:**
    1. Create a function to simulate the basic Moran Process:
        - Initialize a population of N individuals with two types: A (advantageous) and B (baseline)
        - Type A has relative fitness 1+s, where s is the selection coefficient
        - Type B has relative fitness 1
        - In each time step:
            - Select an individual for reproduction with probability proportional to fitness
            - Create an identical offspring
            - Select a random individual (independent of fitness) for death
            - Replace the dead individual with the offspring
            - Continue until either type A or type B goes extinct
    2. Implement functions to calculate:

   - Fixation probability (probability that type A eventually takes over the population)
   - Average time to fixation (number of steps until the population becomes homogeneous)
3. Include options to adjust:
   - Population size N
   - Initial frequency of type A
   - Selection coefficient s
   - Number of independent simulations runs
4. Validate your simulation by comparing its results to theoretical predictions:
   - The theoretical fixation probability of a single advantageous mutant (type A):

$$\rho = \frac{1 - \frac{1}{1+s}}{1 - \frac{1}{(1+s)^N}}$$

   - For neutral evolution (s=0), the fixation probability should be 1/N
5. Plot several individual simulation trajectories showing the frequency of type A over time.
6. Compare the empirical fixation probability from your simulations to theoretical predictions across different values of N and s.
7. Design and conduct your own investigation into the Moran Process by implementing **at least 2** extensions or modifications from the following list, or, better yet - propose your own:
   - Multiple Types: Extend the model to include three or more competing types with different fitness values.
   - Environmental Fluctuations: Introduce temporal variations in the selection coefficient to simulate changing environments.
   - Mutation: Add the possibility of mutation during reproduction, allowing type A to become type B and vice versa.
   - Demographic Stochasticity: Modify the process to include variation in birth and death rates independent of type.

- **Explanation:**
  1. Write a brief summary discussing how the modification you chose effects the simulation dynamics. Why and how does it change the fixation probability and fixation rate.

Question 2 – Moran Process of RNA secondary structures

Objective:

Explore how biological molecules are shaped by real life selection forces using the Moran process.

RNA molecules are inherently short-lived and volatile, but their secondary structures help stabilize them against rapid degradation. Some RNAs are naturally selected for greater stability, leading to the evolution of more stable secondary structures. Using an algorithm estimating the stability of RNA structures, you will estimate the fitness of RNA molecules and simulate their evolution using the Moran process.

Tasks:

- **Simulation:**
    1. Choose a random RNA of 10 bp (e.g. 'GACCGGGTAG')
        - Find the 30 different 1-mutation neighbors of this RNA ('GACCGGGTAC', 'GACCGGGTAA', 'GACCGGGTAT', etc.)
        - For each RNA, use ViennaRNA RNAFold to estimate the secondary structure of the RNA. The lower the free energy, the better the stability of the secondary structure.
    2. Implement a Moran Process function enabling simulation of multiple types (for the 30 different RNAs):
        - Create a population comprised of all the RNAs equally distributed.
        - Run the Moran process for multiple generations.
        - Visualize the distribution of fitness values at the beginning of the simulation vs. the end of the simulation. Show that stable RNAs were positively selected.
    3. Examine how parameters such as population size affects selection probabilities of different variants, given their fitness effect
    4. Postulate a new question and tackle it with the simulation.