Drift and finite population simulations

General comment:

In questions where you are asked about the dependence on a certain parameter, if we didn't learn the mathematical expression in class, a qualitative answer based on the simulations you will run is expected (increase/decrease/no change).

Neutral evolution – compare Moran model to Wright-Fisher:

1. Simulate neutral evolution of a population with N=50 individuals and 2 alleles: A and a. Start with equal frequencies for the two alleles and simulate until one of the two alleles fixes. Register the number of generations until fixation. Repeat this procedure 500 times for each of the two algorithms we studied in class (Wright-Fisher and Moran model). Plot the distributions of fixation times you obtained for each algorithm. Compare the mean time for fixation – is it the same? (20 points)
   (bonus): to better understand the previous result calculate the variance in the number of offspring for the Wright-Fisher and for the Moran model and compare the two results. <u>Guidance</u>: for Wright-Fisher recall that this a Binomially distributed random variable. For Moran – use the probabilities for transition in number of copies that we saw in class. Remember that each Wright-Fisher generation is equivalent to N Moran steps. What did you learn from this comparison? (10 points)
2. Choose one of the algorithms and repeat this neutral simulation for several different population sizes (e.g. N = 10, 30, 50, 100) and again register the time to fixation over 500 repeats for each population size. How does the mean time to fixation depend on the population size (qualitative answer is enough)? (15 points)

Mutation + drift - inifinite alleles (Wright-Fisher)

3. Now add mutations to the simulation, where each mutation event forms a new type of allele that does not yet exist in the population. Use the Wright-Fisher model. In the simulation, you can simply denote each allele type by an integer number. Explain how you implemented this. Use a fixed population size (N = 1000) and test different mutation rates (0.1/N, 1/N, 10/N per individual per generation). Run each of them a constant number of generations. (i) What is the number of different allele types at the end of the run for each mutation rate? (ii) Calculate H (the proportion of heterozygotes, or the probability that two random alleles are different by state) vs. time for each of the mutation rates and plot it H against the time throughout the whole run. Does it equilibrate to a fixed value? If so, to what value? (iii) Plot the distribution of allele frequencies for different mutation rates. How does allelic diversity change with the mutation rate? (25 points)

Drift + selection

4. Simulate a finite population with two alleles A and a having fitness values w_A = 1.1, w_a=1. Start the simulation with the fitter allele A is at frequency p=0.02. Use Wright-Fisher

algorithm. Run the simulation until one of the 2 alleles fixes, for 3 different population sizes, N=50, 500 and 10,000 (5 repeats for each). Compare to the deterministic dynamics $p(t+1)=w_A*p(t)/\bar{w}$ ($\bar{w}$ is the mean population fitness). Which population size behaves more closely to the deterministic dynamics? Does the fitter allele always fix? What is the mean time to fixation for each population size? Does the time to fixation depend on the population size and if so how (qualitative answer)? (20 points)

Quasi-species with and without back mutation

5. Write the quasi-species equations for a model with 2 genotypes A_1 and A_2 with fitness values w_1 and w_2, respectively and one-way mutation rate μ_1 from A_1 to A_2 and a mutation rate μ_2 from A_2 to "dead state" (with fitness zero) – see left-side scheme. Assume $w_1>w_2$.

(i) Write the dynamical equations for the number of individuals, not for their frequencies. Calculate the eigenvalues of the system. What is their meaning? What is the population composition in equilibrium under different parameter regimes? The loss of the fitter genotype A_1 is called "error catastrophe". At which parameter combinations will this occur? (10 points).

(ii) Now add also a back mutation from A2 to A1 at rate $\mu3$ – see right side scheme. Calculate the eigenvalues for the model with back mutation. What is the equilibrium population composition for this model? Is there error catastrophe in this case, too? If so, at which mutation rate(s)? If not, explain why. (10 points).
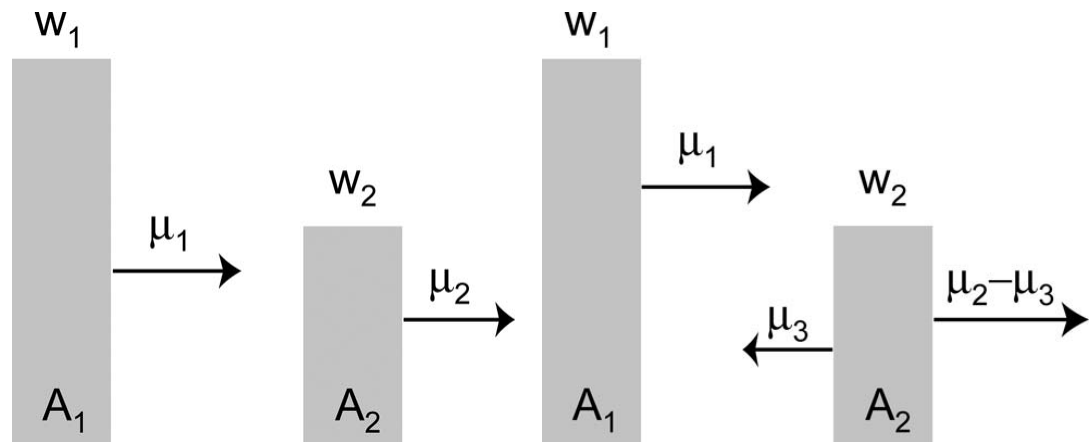
Left – no back mutation. **Right** – model with back mutation.