Bi/Ge105: Evolution
Homework 3
Due Date: Thursday, March 01, 2018

1 Problem 1: Genetic Drift as a Force of Evolution

1.1 Simulating the processes of evolution

In class we learned about the mathematical formalism behind population genetics, one of the centerpieces of evolutionary theory. The ideas described in class will provide a quantitative backdrop for understanding the different evolutionary forces that shape life on our planet. It is both profound and amusing how much we can learn about evolution by thinking about coin flips and similar games of chance. Indeed, the broad reach of the mathematics of coin flips is an example of what former Caltech undergrad and now Harvard professor Joe Blitzstein likes to say: "The nouns change, but the verbs remain the same."

In this problem we want you to explore different evolutionary forces by means of simulations. You will use what you learned in the *stochastic simulation* tutorial to explore the interplay of different evolutionary forces such as genetic drift and mutation. By using simulations we will sidestep more advanced mathematics of stochastic differential equations needed to study these concepts analytically while still getting clear insights into how these forces may affect the course of evolution.

1.2 The Buri genetic drift experiment

In 1956 Peter Buri, a student of Sewall Wright published the now classic paper "*Gene Frequency in Small Populations of Mutant Drosophila*" in which he experimentally demonstrated the concept of genetic drift. The idea for this beautiful experiment is depicted in Fig. 1.2. Briefly, Buri began with eight female and eight male flies, all heterozygotes of the *bw* locus. This means that all of the flies had 1 copy of the gene associated with white eyes, and one copy of the gene associated with red eyes. The phenotype that this combination of alleles gives is flies with orange eyes. He then allowed the

flies to reproduce, and after removing the adults, he randomly chose 8 males and females from the next generation of offspring without looking at the eye color. These new 8 males and 8 females were transferred to a new flask and the procedure was repeated for 19 generations.

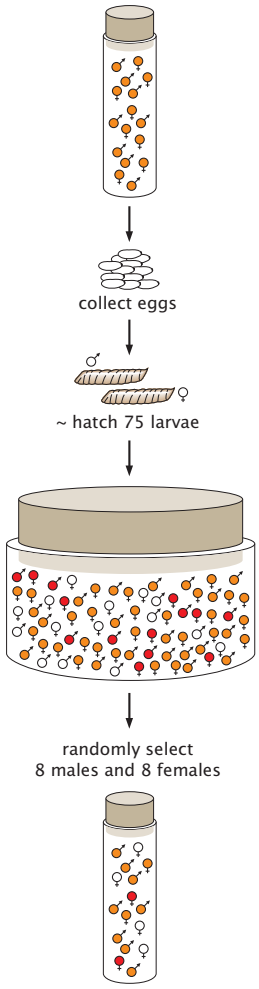

Figure 1: **Buri's experimental setup.** At time $t = 0$ eight heterozygote females and eight heterozygote males were allowed to reproduce. From their offspring, eight males and eight females were chosen at random and transferred into a new flask.

Question 1a: Work out what is the expected genotype frequency of

red-eyed flies, white-eyed flies and orange-eyed flies after the first generation. (Hint: Recall that each allele is drawn from the parent's pool **at random with replacement**. This means that to compute the frequency of red-eyed flies you should calculate $f_{rr} = P(\text{red allele first draw}) \cdot P(\text{red allele second draw})$.

Since the offspring that made it to the next generation were chosen at random, Buri knew that the outcome would be different if he repeated an identical experiment in different vials. As a result, for statistical power he simultaneously tracked 107 flasks as shown in Fig. 1.2. Each generation, he counted the number of red-eyed, white-eyed and orange-eyed flies he had randomly chosen. Fig. 2 shows the outcomes for these different vials after 19 generations. Because the flies are allegedly mating at random, with each generation there is an accumulation of fluctuations. As a result, after 19 generations, many vials contained only white-eyed or red-eyed flies, though some vials still contained a mixture of eye colors.

Having quantified the number of red-eyed, white-eyed and orange-eyed flies Buri was able to quantify the frequency of alleles in the population. Since none of the alleles were dominant, he could infer the genotype by looking at the phenotype of the flies.

Question 1b: Write down the formula for the genotype frequencies in terms of the eye color count. Use the notation N_{red} for the number of red-eyed flies in a given vial, N_{white} for the number of white-eyed flies in that same vial and finally, N_{orange} for the number of orange-eyed flies in that same vial. Your task is to figure out the frequency of red (f_r) and white (f_w) alleles in a given vial given the counts of the number of red-, white- and orange-eyed flies.

Fig. 3 summarizes the results of the experiment. By tracking alleles over time with these 107 populations exposed to the same conditions, Buri was able to observe evolution driven entirely by genetic drift! He saw how in some of the populations one of the alleles went extinct, arising from nothing more than the fluctuations inherent in small populations.

It is now time for us to use our computational prowess to simulate and explore the Buri experiment.

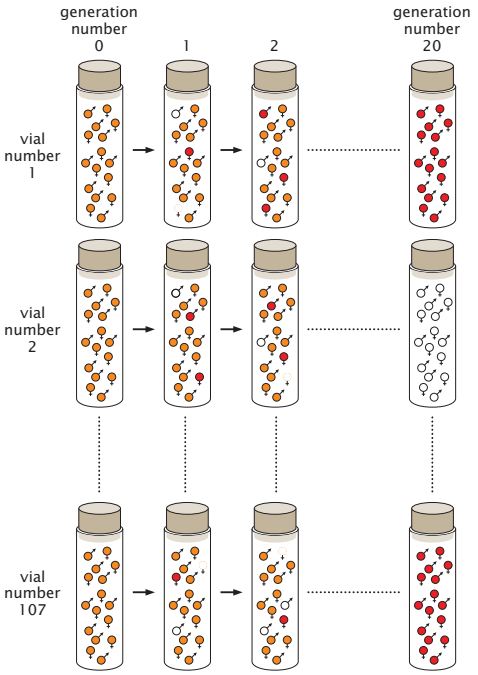

Figure 2: **Multiple replicates of the Buri experiment.** Buri repeated his experiment in 107 separate vials, with the evolutionary trajectory different each time as a result of genetic drift. Note that in the long time limit, many of the vials have gone to fixation with all flies having either white or red eyes.

1.3 Reproducing the Buri experiment *in-silico*

Your first task will be to reproduce Fig. 3 by means of stochastic simulations. The key elements of the code you need to do this analysis you already worked out in the stochastic simulation tutorial.

Question 1c: Perform stochastic simulations of genetic drift for 107 populations over 19 generations using the same population size as Buri, i.e. 16 flies total (32 alleles). Plot histograms of the allele frequency for generation numbers 0, 1, 10, and 19.

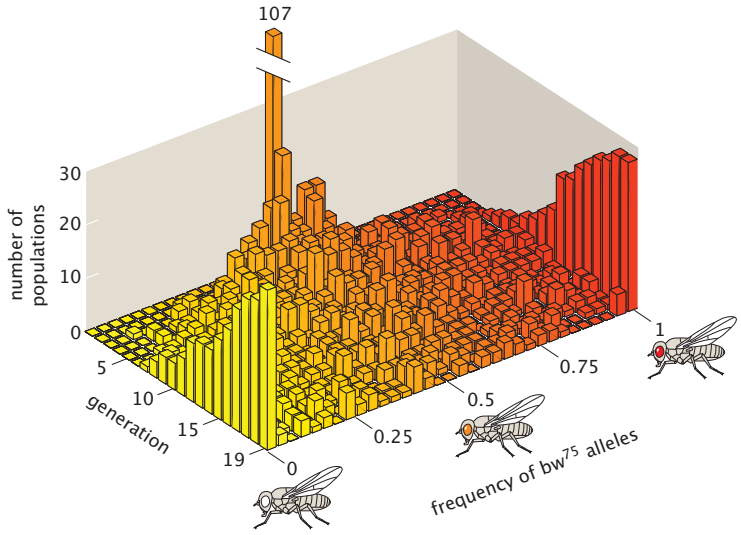

Figure 3: **Results of the Buri experiment.** By tracking the phenotypes of the flies, Buri was able to infer the allele frequencies for each population. The allele frequencies change as a result of genetic drift and after 19 generations, many of the vials contain flies all with the same eye color, implying fixation of alleles and evolution due to genetic drift.

1.4 The effect of the population size

Using these exact same tools we will now explore the effect of the population size.

Question 1d: Repeat the stochastic simulations for 107 populations during 1000 generations using the same population size as Buri. Quantify the time it takes for each of these populations to have one of the alleles fixed, i.e. find the time point for each population at which the allele frequency becomes either zero or one, and save the generation number at which this happened. Now repeat the simulation for varying population size ($N = 4$, 8, 16, 32, and 64). Plot the mean time to fixation as a function of the population size and comment on how this average time to fixation scales as the population size changes. What do these results mean for the role genetic drift plays in different populations? (Hint: to find which generation one of the alleles was fixed in the population, the function `numpy.where` might become handy. Basically you just need to find a way for Python to tell you

at which entry of the array the frequency became `f == 1` **or** `f == 0`. You might also want to check the `numpy.logical_or` function that allows you to perform boolean `or` operations over numpy arrays.)

1.5 The effect of mutations

Let's now explore the effect of another evolutionary force – mutation. In our toy model, rather than thinking about tracking the complexity of single base pair mutations, we will think of a "reaction" of the following form

$$A \underset{\mu_2}{\overset{\mu_1}{\rightleftharpoons}} a \tag{1}$$

where A and a are the two versions of the allele (for example red and white), and μ_1 and μ_2 are the mutation rates that take you from one allele to the other. To simplify things even further we will assume $\mu_1 = \mu_2 \equiv \mu$.

Question 1e: Implement a stochastic simulation to include the effect of mutation for a single population and plot the allele frequency over time. Comment on the differences with respect to the case without mutation. (Hint: The mating still happens at random in this scenario, but now each allele after being selected for the next generation must flip a second coin to decide if it remains as the same allele, or it mutates into the other allele). Use the value $\mu \approx 0.001$ for your simulations.

Question 1f: Extend the algorithm you just wrote and simulate 100 populations. Plot 10 of these trajectories, as well as histograms of allele frequency at representative time points such as $t = 0, 5, 10, 50, 100, 500$ generations. Compare this to the null model where the mutation rate is equal to zero and comment on the differences if any between the distributions over time.

Question 1g: You will now explore the effect of the magnitude of the mutation rate. Run the simulation for 100 generations for μ =0, 0.001, 0.01, 0.1 and plot the histogram of allele frequencies of the final time point for

each of these mutation rates. Comment on how the distribution changes as the mutation rate increases.

2 Problem 2: Symbiosis as a driving force of evolution.

Question 2a: In class we discussed the arms race between the poisonous rough skinned newt and garter snakes and the observed increased resistance to the toxin by the predator. How does the newt (or another toxin producing organism) avoid poisoning itself? Find an example in the literature, describe the mechanism in 1-2 paragraphs and cite the source.

Question 2b: How would you go about characterizing the taxonomy a new organism whose genome was 1/3 eucaryotic, 1/3 archaeal, and 1/3 bacterial?