HST.508 Evolutionary Genomics
Problem Set 2

Due Friday October 13th

Problem 1: Fixation of a beneficial mutation

In the class we computed the probability of fixation for a beneficial mutation and considered the role of the selection coefficient, s. Other important parameters to consider are the initial frequency of a mutation *p* and the final frequency of the polymorphism.

$$\Pi(p) = \frac{1 - e^{-2NSp}}{1 - e^{-2Ns}}$$

(a) Plot what the probability of fixation as a function of initial frequency, p, will look like as a function of s. Consider both the weak selection (s<<1/N) and the strong selection (1/N<<s<<1) cases. You can do this is analytically, or by plotting the probability of fixation as a function of s for a fixed N (e.g. N=300) for different values of p=0.01, p=0.001 etc.

(b) Compare the probability of fixation for a new mutation (e.g. a mutation occurring with frequency of Pnew=1/2N) and the one arising due to migration and finite p=10%, 1% or 0.1%. Do you see different behaviors for weak and strong selection? You can plot the ratio of the probability of fixation for a given p (above) and for Pnew.

Extra credit: Compute the probability of achieving "near-fixation" rather than fixation, i.e. compute the probability of achieving frequency f (e.g. 90% or 99%) starting with some frequency p. The differential equation to find such probability of fixation is the same as derived in the class, but the boundary conditions are different, i.e. the general solution for the probability of fixation is

$$\Pi(p) = A\, e^{-2NspA} + B$$

For the probability of fixation, one uses $\Pi(1) = 1;\ \Pi(0) = 0$ but for the probability of reaching a frequency f, one uses $\Pi(f) = 1;\ \Pi(0) = 0$.

Find the initial p0 that provides 99% probability of fixation in the limit of large N*s. Use this frequency to calculate the actual number of mutated chromosomes in the population (you can set the selection coefficient, s = 0.001). Will the number depend on population size? Interpret your results.

Problem 2: Heritability

Please do the data analysis for this problem in a Python Notebook, and submit your code with your completed set. Use *times.csv* as the dataset.

In this problem we are going estimate heritability for a trait generated from synthetic data. The trait being considered is the time in seconds it takes a candidate astronaut to orient themselves in a balanced standing position after being put into a human centrifuge. We'll call this recovery time. For the purpose of this problem, assume this trait is known to have a genetic component. The goal of the problem set is for you to gain an appreciation for heritability estimation, and how environmental factors can contribute to genotypic variation.

(a) Narrow-sense heritability. Calculate the heritability, h^2 via the regression between parents and offspring. For this problem, calculate the heritability: 1) Using the covariance between parental phenotype and child's phenotype and 2) Using the covariance of mid-parent phenotype and child's phenotype. For each method, plot the scatter of phenotypic values and the regression line that you calculated. The data for this question is attached to your problem set.

(b) Effect of environmental variance. The phenotypes in the previous question were given to you in the absence of any environmental components of variance. Describe/explain in 2 sentences whether heritability would increase, decrease, or stay the same in the presence of environmental sources of variation.

Now add different levels of random environmental noise (as normally distributed random numbers with zero mean and standard deviation s) to the data and recompute h^2.
You can try:

- the same magnitude of noise added to parents and children

- different magnitude of environmental noise parents and children

- different magnitude of environmental noise acting on siblings

How do these effects change your estimates of heritability?