Exercise Set 1

Do at least one of the following three exercises:

1. Write a program that simulates the haploid Wright-Fisher model with selection, as described in class. Note that, ultimately, the *A* allele must sweep to fixation or be lost (i.e., achieve a frequency of $q = 0$ or 1) in every population.

   (a) Use your program to estimate the probability (= fraction of replicate simulations) that a new advantageous mutation (i.e., a single copy of the *A*-type with fitness $1 + s$) becomes fixed in a population of size $N = 4$ when $s = 0.1$, which represents a substantial selective advantage, and in a population of the same size but for a new mutation with a smaller advantage $s = 0.01$. How do these compare with the branching process-based approximation P(fixation) = $2s$?

   (b) Can a deleterious mutation sweep to fixation? Use your program to estimate the probability of a new mutation with $s = -0.01$ sweeping to fixation in a population of size $N = 4$.

   (c) Use your program to estimate the probability that a new "neutral" mutation (i.e., one with $s = 0$) will sweep to fixation. How does this probability compare with those of weakly advantageous ($s = 0.01$) and weakly deleterious ($s = -0.01$) mutations?

   (d) [O*ptional*] Use your program to explore how population size affects fixation probabilities in the different scenarios above. Discuss!

2. (*Requires a little background in linear algebra*.) Probabilities of fixation can be computed *exactly* from the binomial transition probabilities, g_{jk} of the Wright-Fisher model with selection. Recall the argument from class: Let u_i be the probability that a population with *i* copies of *A* ultimately becomes fixed for *A*. Note that $u_0 = 0$ (a population with no *A* alleles has no chance of fixation) and $u_N = 1$ (*A* already is fixed). If $i \neq 0$ or *N*, the population will have *j* copies of *A* in the next generation with probability g_{ij}, in which case the probability of fixation would then be u_j. Considering all possible transitions of a population with *i A* alleles leads to (for $i = 1, 2, \ldots, N-1$)

$$(*)\quad u_i = g_{i0}u_0 + g_{i1}u_1 + \cdots + g_{i,N-1}u_{N-1} + g_{i,N}u_N = \sum_{k=1}^{N-1} g_{ik}u_k + g_{i,N}$$

   where we used $u_0 = 0$ and $u_N = 1$.

   Matrix methods can be used to compute all $N - 1$ of the unknown fixation probabilities. Let the vector of unknown probabilities be $\mathbf{u} = \left(u_1, u_2, \ldots, u_{N-1}\right)^{\mathrm{T}}$. In addition set $\mathbf{r}=\left(g_{1,N},g_{2,N},\ldots,g_{N-1,N}\right)^{\mathrm{T}}$ and let **Q** be a matrix with *ij*th element equal to g_{ij}, with *i* and *j* both ranging from 1 to $N-1$. Equation (*) is equivalent to the matrix equation $\mathbf{u}=\mathbf{Qu}+\mathbf{r}$. This can easily be solved for **u**:

$$(**)\ \mathbf{u}=\left(\mathbf{I}-\mathbf{Q}\right)^{-1}\mathbf{r},$$

where **I** is the identity matrix of size N–1 and $(\mathbf{I}-\mathbf{Q})^{-1}$ is the inverse of the matrix $\mathbf{I}-\mathbf{Q}$. The fixation probability of a new mutant *A* is the first element, u_1, of **u**. Use (**) to compute the exact probability of fixation for the scenarios described in exercise 1. Discuss!

3. Using a branching process approach, we showed in class that the probability *P* of fixation of a new advantageous mutation satisfies $1-P=e^{-(1+s)P}$ and argued that $P\approx 2s$. In this problem, you will develop a better approximation for *P*.

   (a) Taking the natural log of both sides, the equation defining *P* is equivalent to $\ln(1-P)=-Pe^{x}$ where $x=\ln(1+s)$. Replace $\ln(1-P)$ and e^{x} in this equation with their Taylor series in *P* and *x*, respectively.

   (b) Substitute $P=Ax+Bx^2+Cx^3+\cdots$ in the left-hand side of the resulting equation and solve for the coefficients *A*, *B, C, …* by equating coefficients of powers of *x* on both sides of the equation.

   (c) Finally, expand $\ln(1+s)$ in a Taylor series and collect terms to find an approximation for *P* that is at least cubic in *s*.

   (d) Use this approximation to compute *P* for the scenario described in exercise 1. (Compare this to the “rough” approximation $P = 2s$.) Discuss!

   (e) [*Optional*] Compare your answer to (d) with the approximate fixation probability computed via simulation (ex. 1) or the exact probability (ex. 2).