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 \( 2N = 4 \) 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 \( P(\text{fixation}) = 2s \) that was discussed in class?

   (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 \( 2N = 4 \).

   (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?

2. (Requires a little background in linear algebra.) Probabilities of fixation can be computed exactly from the binomial transition probabilities, \( g_{ij} \) 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 for \( i = 0, 1, \ldots, 2N \). Note that \( u_0 = 0 \) (a population with no A alleles has no chance of fixation) and \( u_{2N} = 1 \) (A already is fixed). If \( i \neq 0 \) or \( 2N \), 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 one-step transitions of a population with \( i \) A alleles leads to

\[
(*) \quad u_i = g_{i0}u_0 + g_{i1}u_1 + \cdots + g_{i,2N-1}u_{2N-1} + g_{i,2N}u_{2N} = \sum_{k=1}^{2N-1} g_{ik}u_k + g_{i,2N}.
\]

The boundary conditions \( u_0 = 0 \) and \( u_{2N} = 1 \) are used in the last expression.

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

$$\text{(**)} \quad u = (I - Q)^{-1}r,$$

where $I$ is the identity matrix of size $2N - 1$ and $(I - Q)^{-1}$ is the inverse of the matrix $I - Q$. 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. Be sure to compare your calculations for scenarios with selection ($s = 0.1, 0.01, -0.01$) and the “neutral” case ($s = 0$).

3. Using a branching process approach, we showed in class that the probability $P$ of fixation of a new advantageous mutation (i.e., one with selection coefficient $s > 0$) satisfies $1 - P = e^{-(1+s)P}$ and claimed that $P \approx 2s$. In this problem, you will develop an even 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 $A, B, C, \ldots$ by equating successive coefficients of powers of $x$ on both sides of the equation up to $x^3$.

(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 scenarios $s = 0.1$ and $s = 0.01$ (described in exercise 1). In each case, compare this to the “rough” approximation $P = 2s$. Discuss assumptions made by this approximation approach.

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