Selection & Fixation

1. **Mutation-selection balance.** Consider a population of a fixed number $N$ cells. In each generation, a cell randomly acquires $j$ additional mutations, where $j$ is Poisson distributed, $p(j) = e^{-\mu} \mu^j / j!$, with average $\mu$. These mutations are mildly deleterious, such that a cell with $j$ mutations has a relative fitness of $f(j) = (1 - s)^j$, with (multiplicative) selection coefficient $0 < s \ll 1$. Consider a steady state of mutations and selection in the system.

   (a) Write the recursion relation for the fraction of cells with $k$ mutations, $x'_k$, after one generation. Consider that cells first survive with a probability proportional to their relative fitness, and then acquire new mutations. Remember to normalize by the mean fitness of the population, $\overline{f} = \sum_{i=0}^{\infty} x_i (1 - s)^i$.

   (b) Find the mean fitness of the population by considering the steady state for cells with zero mutations ($x'_0 = x_0$).

   (c) Solve for the steady state distribution, such that $x'_k = x_k$. (Hint: try a Poisson distribution.)

2. **Selection and mutation:** Consider a very large population of individuals characterized by a fitness parameter $f$, which is assumed to be Gaussian distributed with a mean $m$ and variance $\sigma$. The population undergoes cyclic evolution, such that at each cycle:

   (i) one half of the population with lower fitness $f$ is removed without creating progeny; (ii) the remaining half (with $f$ values in the upper half) reproduces before dying; (iii) because of mutations that are on average neutral the $f$ values of the new generation is again Gaussian distributed, with mean value and variance reflecting the parents (i.e. coming from the upper half of the original Gaussian distribution).

   (a) Relate the mean $m_n$ and variance $\sigma_n$ of fitness values of the $n$-th generation to those of the previous ones ($m_{n-1}$ and $\sigma_{n-1}$).

   (b) What happens to the distribution of fitness after many generations?

   (c) Most mutations are deleterious, while at the same time increasing the diversity of the population. To study these effects, assume that at each generation the distribution obtained in (a) above is convoluted with a Gaussian of mean $-\mu$ (thus reducing the mean fitness) and variance $s^2$ (acting to increase the variance in fitness). Find the recursion relations for $m_n$ and $\sigma_n$ in this case.

   (d) What happens to the fitness distribution at long times in this case?

3. **The Moran process,** named after Patrick Moran, is a simple method for modeling a population of constant size. At each step one individual from the population is randomly selected for duplication/reproduction, and another for elimination/death, thus maintaining a fixed size.
(a) For a haploid population of size $N$, with one locus and two alleles $A_1$ and $A_2$, compute the changes $\langle \Delta N_1 \rangle$ and $\langle \Delta N_2 \rangle$ in number of individuals with allele $N_1$ after one step.
(b) Construct the drift-diffusion equation for this model, assuming that $N/2$ steps of the Moran process correspond to one generation time.
(c) How would you modify the process to implement differing fitness values for the two alleles?

*****

4. Multi–allele model: Consider a locus occurring in one of $s + 1$ states ($A_1, A_2, \cdots A_s, A_{s+1}$) with frequencies $x_1, x_2, \cdots, x_s, x_{s+1} = 1 - \sum_{i=1}^s x_i$ in a (haploid) population of fixed size $N$. After 1 generation, random reproduction leads to new population in which the allele numbers $\{m_i\}$ occur with the multinomial probability

$$p(\{m_i\}) = \frac{N!}{s+1!} \prod_{i=1}^{s+1} \frac{x_i^{m_i}}{m_i!}.$$

(a) Show that the changes in frequency after one generation satisfy (for $i = 1, \cdots, s$)

$$\langle \Delta x_i \rangle = 0, \quad \langle \Delta x_i^2 \rangle = \frac{x_i(1-x_i)}{N}, \quad \text{and} \quad \langle \Delta x_i \Delta x_j \rangle = -\frac{x_ix_j}{N}, \quad \text{for} \quad i \neq j.$$

(b) Obtain the forward Kolmogorov equation for the probability $p(\vec{x}, t)$, where $\vec{x} = (x_1, \cdots, x_s)$ is a vector with $s$ components.
(c) Obtain the backward Kolmogorov equation for the probability $p(\vec{x}, \vec{y}, t)$, given an initial state $\vec{y}$.
(d) The probability that an initial polymorphism characterized by $\vec{y}$ disappears (by either loss or mutation) at time $t$ is given by

$$p_x(t|\vec{y}) = -\int d^n\vec{x} \frac{dp(\vec{x}, t|\vec{y})}{dt}.$$  

Find an equation (similar to the backward Kolmogorov equation) satisfied by the mean time for fixation $\langle \tau(\vec{y}) \rangle_x$, and show that a simple generalization of the result for two alleles satisfies this equation.

*****

5. Splitting probability and mean first passage time:
(a) Consider a particle diffusing without a drift ($v = 0, D = \text{const}$) on an interval $[0, 1]$ that has adsorbing boundaries. The particle starts at $0 < x < 1$ and diffuses until it gets adsorbed by either of the boundaries. Using backward Kolmogorov equations calculate the probabilities of adsorption (splitting probabilities) by either boundary $\Pi_0(y)$ and $\Pi_1(y)$.
(b) Obtain the expression for the mean first passage time $\bar{T}(y)$ through either boundary (survival time). Find the location $y_{\text{max}}$ that provides the longest mean life-time for the particle.
(c) **Optional** Consider a generalization of the same problem to two dimensions, with two concentric circles of radii $R_0 < R_1$ centered at the origin. A particle starts a 2D diffusion (without a drift) at distance $R$ ($R_0 < R < R_1$) from the origin. Use the general form of the equations for splitting probabilities, $\nabla^2 \Pi = 0$, and mean exit time, $D \nabla^2 \tau = -1$, to calculate $\Pi_0(R)$, $\Pi_1(R)$ and $\bar{t}(R)$. Find the initial distance $R_{max}$ that provides the longest life-time. Compare to result obtained for one dimension.

(d) **Optional** Consider the same problem in three dimensions with two concentric spheres. Compare results in 1D, 2D and 3D. How do the split of probabilities between the inner and the outer boundaries, and $R_{max}$ change as the dimensionality of the system increases from 1 to 3?