Stochastic Models in Mathematical Genetics
MSc Problem Sheet 4

Michaelmas Term 2020

1. Consider a Wright-Fisher model with *selection* and *mutation*. There are two types, A and a in a population of constant size $2N$. Population members choose their parents independently at random, with relative probability 1 of choosing each a type parent and $1+s$ of choosing each A type parent. *After* a parent is chosen, mutations from type a to type A occur with probability μ_A and from A to a with probability μ_a.

   (a) If the number of individuals in the population who are type A in generation k is Z_k, explain why Z_{k+1} has a Binomial distribution and find the parameters of this distribution in terms of Z_k.

   (b) Mimicking the argument in the lectures, show that as $N \to \infty$ while $\gamma = 2Ns$, $\theta_A = 4N\mu_A$, $\theta_a = 4N\mu_a$ remain fixed, and setting $X_t = Z_{\lfloor 2Nk \rfloor}/(2N)$, the process X_t will converge to a diffusion process with infinitesimal variance and mean

$$
\begin{aligned}
a(x) &= x(1-x) \\
b(x) &= \gamma x(1-x) + \theta_A(1-x)/2 - \theta_a x/2.
\end{aligned}
$$

2. Consider a general Wright-Fisher model where in a constant size population of $2N$ chromosomes, there are two types A and a, and in the kth generation, the proportion of chromosomes carrying the A type is X_k, where $X_0 = x$. Now suppose that in generation $(k+1)$, each chromosome independently chooses a parent, and that the probability that parent is of type A is given by

$$
P(X_k) = X_k + \frac{b(X_k)}{2N} + o(N^{-1}).
$$

   Show that as $N \to \infty$, the Markov chain describing the proportion of chromosomes carrying the A type converges to a diffusion process limit after an appropriate rescaling of time, and give the diffusion and drift parameters for the limit process.

3. Code up a simulator for the Wright-Fisher model with selection. This should take the form of a function that takes as arguments: (x_0, s, N, t) where x_0 is the initial frequency of allele A, s is the selection coefficient for allele A, N is the population size, and t is time in generations. Plot the frequency of allele A from 0 to t and explore what happens for varying s, using $x_0 = 0.1$, $t = 4,000$ and $N = 10,000$. You should terminate the simulation once loss or fixation occurs (i.e. fix X). Use the simulator to estimate the fixation probabilities for a range of s values $s = \{-0.001, 0, 0.00005, 0.0002, 0.002\}$. Run the simulator for 100,000 generations; what is the average time until fixation?

4. Consider the diffusion limit of the Wright-Fisher process with mutation and selection considered in Question 1 (i.e., consider the diffusion with this infinitesimal mean and variance).

   (a) Write down the generator of the diffusion.

   (b) Consider the problem of estimating the fixation probability $h(x)$ of the A allele, which has initial frequency x in the population at $t = 0$. (Assume that fixation of either the A or a alleles is certain in finite time.) Write down a differential equation, and boundary conditions, satisfied by $h(x)$.

(c) Solve the differential equation to give the appropriate fixation probability in terms of a ratio of integrals involving the parameters γ, θ_A, θ_a.

5. Augment your simulator for the Wright-Fisher model to include mutations as per Question 1. You should terminate the simulation once loss or fixation occurs (i.e. fix X). Take $x_0 = 0.5$, $s = \frac{5}{2N}$, $N = 10,000$, $t = 100N$, $\mu_a = \mu_A = \frac{0.1}{4N}$ and calculate the average time to fixation. What happens as you increase μ_a towards $\frac{1}{4N}$?