Problem Set 5
Assigned Nov 9 2023, Due Nov 21 2023

Problem 1 (20 points): In this problem, we will consider a few ways we might pose a single parameter inference problem. We will assume that our problem is to learn mutation rates in a bacterial genome. We assume each site in the genome has its own unique mutation rate μ_i. We will imagine that we have examined many generations of bacterial division and have counted the number of mutations at each site. We therefore know a number of generations g and an observed number of mutations n_i for each site.

a. Pose the problem of inferring the mutation rates μ_i as a least-squares optimization problem. Provide a mathematical formula formally specifying the objective function for this model of the problem.

b. Now pose this as a maximum likelihood problem. Provide a mathematical formula formally specifying the likelihood function for this version of the problem.

Problem 2 (10 points): Find the stationary distribution of a two state Markov chain in which the transition probability to a good state (state 1) is the same wether the system is currently in a good or bad state. Let us denote this probability by p. Provide an intuitive explanation for the results, which should depend on a single parameter p.

Problem 3 (20 points): Consider a patch within a forest subject to slash and burn agriculture. Assume that the vegetation that grows best on the patch depends on whether the patch was occupied by an early or late successional forrest before being cleared. For example, corn (state 1) might grow better when an early successional forest (state 2) is cleared but beans (state 4) might grow better after a late successional forest (state 3) is cleared. Assume that the annual probability that a patch is cleared is e_2 and e_3 for early and late successional forests, that early successional forests become late ones with probability b_2 and that the patch remains agricultural land once cleared. The transition probability matrix then becomes:

$$\begin{pmatrix} 1 & e_2 & 0 & 0 \\ 0 & (1-e_2)(1-b_2) & 0 & 0 \\ 0 & (1-e_2)b_2 & 1-e_3 & 0 \\ 0 & 0 & e_3 & 1 \end{pmatrix}$$

a) What are the four eigenvalues of this matrix?

b) What are the absorbing states of matrix M?

c) Calculate the probability that the patch ultimately becomes a corn field (state 1) starting from either state 2 or state 3.

Problem 4 (30 points): In class we discussed and derived results for the Moran birth-death model with overlapping generations. We answered the question: what is the probability of fixation

of a new mutant A in the population composed of B individuals, given a fixed population size N and the assumption that the new mutant does not confer a fitness benefit (neutrality assumption).

a). Simulate the Moran birth-death model, plot some trajectories over time and see whether you can recapture the result we derived in class. (The probability of fixation is $1/N$, for a given N fixed population size.)

b). If we now assume a selective benefit of s for allele A (such that B individuals have probability 1 to give birth and A individuals have probability $(1+s)$) discuss (using simulations) how this changes the probability of fixation for A.

*Problem 5 (20 points): Suppose that a new strain of influenza has been found in the human population. You are asked to make predictions about a potential pandemic so you decide to model the spread of this new strain using a birth death stochastic model in discrete time. You thus choose the time steps to be small enough so that, in each time step, the number of infections can either increase by 1, decrease by 1 or remain the same. The probability that the number of infections increases by 1 is βSI where I is the current number of infections, S is the number of susceptible people (assumed to be constant) and β is the transmission rate. The probability that the number of infections decreases by 1 is $(c + d)I$ where c and d are rates of recovery and death respectively. There are two outcomes of interest: epidemic dies out or number of infections reaches m at which point pandemic is inevitable.

a. Write out transition probabilities.

b. Suppose that $\beta S > c + d$. (This means that in a deterministic version of this stochastic model the number of infections would increase). What is the probability that global pandemic will occur, starting from one infection? You can use simulations to find this probability (as a function of parameters, that you can plot on the x axis). Use what we know about birth death Markov Chains (including online content or chatGPT) and see if you can derive this probability analytically.

***Recall that the starred problems are only for the 02-712 students.