Problem Set 11

Due: Friday 5pm, Dec 7, 2017, in the homework drop box located between the 3rd floor of building 8 and 4th floor of building 16.

1 Fixation probability in the Moran process (12 points)

Consider a population of N individuals in which a mutation of fitness r appears. Assume the population is well-mixed and evolution proceeds according to the Moran process. Write F_n for the probability the mutation will eventually become fixed if there are n mutants in the population. Explain carefully why F_n satisfies the recurrence relation:

$$F_n = \left(\frac{1}{1+r}\right) F_{n-1} + \left(\frac{r}{1+r}\right) F_{n+1}.$$

What are the values of F_0 and F_N?

By considering solutions to the recurrence relation of the form $F_n = \lambda^n$ where λ is a constant, or otherwise, show that

$$F_n = A + \frac{B}{r^n}$$

where A and B are constants you should determine.

2 COMPUTATION Simulation of clonal interference (22 points)

The simulations in this problem can take a long time to run. Please plan accordingly.

Consider a population of size N subject to beneficial mutations that are distributed exponentially with a characteristic magnitude of $s_0 = 0.02$.

a. Simulate the Moran process with a single mutant starting in the population. In each run, the selective advantage of this mutant should be sampled from an exponential distribution. For $N = 10$ and $N = 500$, plot histograms and cumulative distribution functions of the selective advantage of mutants that eventually take over the population, and compare it with what you expect analytically.

When there are multiple mutants competing at the same time, we can encounter clonal interference. Introduce in your simulations a probability μ (the mutation rate) of generating a new beneficial mutation. For simplicity, we do not allow mutants to acquire secondary beneficial mutations. Each time when a non-mutant is chosen to divide, there is a probability μ that the daughter cell will gain a beneficial mutation sampled from the exponential distribution.

b. For $N = 10$ and $N = 500$, at what value of μ do you expect the distribution of fixed beneficial mutations to change significantly relative to your results in part a? How do you expect it to change?

c. For $N = 10$ and $N = 500$, plot histograms of the selective advantage of mutants that eventually take over the population with $\mu = 2 \times 10^{-5}$ and 10^{-2}. $>$500 trials should be enough to give a nice trend (where one trial = a successful fixation of a mutant in the population). Compare with what you predicted above.

3 Bet hedging: To germinate or not to germinate? (18 points)

The following is based on the classical paper of Cohen, Optimizing Reproduction in a Randomly Varying Environment, *J. Theo. Biol.* (1966). While it will not be covered in class, a lecture is not required to complete the problem.

Consider a species of plant whose seeds can germinate once a year. Every year, the environment's harshness fluctuates, introducing uncertainty in the number of successful sprouts. Imagine that a fraction H of the seeds directly germinate, and the rest face a risk of not surviving the season if the environment is arid. The seeds that germinate will grow (or not) and produce Y_t (t being an index for time) seeds for the next season. Note that because the environment fluctuates, Y_t is a random variable (unspecified for now). The remaining seeds "hibernate", and survive to the next season with a probability s independent of the environment's state. We will try to determine in which situations the probability to germinate H is less than 1, which corresponds to the seeds hedging their bets (wait for next year to germinate in case the environment is too harsh).

a. Given that one starts with $n(t)$ seeds, what will be the number of seeds at $t+1$ (we have discrete time here since germination can happen once a year)? Given that we start with $n(0)$ seeds, write down the number of seeds at time t in terms of the $\ell(k) := HY_k + (1-H)s$, with $k < t$.

b. Assuming the environment's fluctuations are independent from year to year (iid random variables), what is $\langle n(t) \rangle$? By taking the logarithm of your expression for $n(t)$, make an argument that $n(t)$ should be log-normal as t becomes large. What is the median of the distribution in $n(t)$ in terms of $\langle \log \ell \rangle$? Is the mean necessarily a representative statistic here?

c. **COMPUTATION** As an example of the above, assume that ℓ can be either 0.7 or 1.35 with equal probabilities. What is $\langle \ell \rangle$? What is $\langle \log \ell \rangle$? Simulate this process for 300 steps, for 100000 trials. How do the mean and median of n(t) evolve as a function of time? How is that possible?

d. The above should have convinced you that the what needs to be maximized is $\rho := \langle \log \ell \rangle$. Note that ρ is a function of H. What is ρ at $H = 0$? Compute the first and second derivative of ρ with respect to H. What can we say about the second derivative? What are then the three qualitatively different scenarios for how ρ changes with H? What is $\rho'(H=0)$? When is $\rho'(H=0) < 0$? What does that mean and what is the optimal H in this case? This case is unrealistic for a real species. Now, argue that the optimum strategy arises for $0 < H < 1$ iff $\rho'(H=1) < 0$ and $\rho'(H=0) > 0$. Thus, obtain a condition involving Y_t (or some form of average thereof) and s for which it is advantageous for some non-zero fraction of the seeds to hibernate.

e. Finally, consider a "good year/bad year" environment, where Y_t equals either M or 0 with probability p and $1-p$ respectively. Compute the optimum H as a function of s, M and p. If $pM \gg s$, what does your expression converges to? Explain in words what that means.