Problem Set 9 (70 points)

1 Quasispecies Equation (12 points)

This problem will introduce you to the quasispecies equation and teach you how to set up the equation for a few different scenarios. The word quasispecies refers to a group of individuals that form a single species but have slightly different genotypes. The quasispecies equation, first formulated by Manfred Eigen and Peter Schuster, is a mathematical model that describes how the frequencies of different genotypes change within a population over time. Let's take a look at this equation[1]:

$$\dot{x_i} = \sum_{j=1}^{n} x_j f_j Q_{ji} - \phi(\vec{x}) x_i$$

Here, x_i and f_i are the frequency and fitness of genotype i. $\overleftrightarrow{Q}$ is the mutation matrix. Q_{ji} denotes the fraction of individuals of genotype j that mutate to genotype i. We note that $\overleftrightarrow{Q}$ is a right stochastic matrix (transition matrix), i.e., $\overleftrightarrow{Q}$ is a square matrix of nonnegative real numbers, with each row summing up to 1. $\phi(\vec{x}) = \Sigma_i f_i x_i$ is the average fitness of the population.

a. [6 points] Now that we know what the various terms mean, it's time to perform a sanity check on the quasispecies equation.

   1. Explain in words what it means when $\overleftrightarrow{Q}$ is equal to the identity matrix.
   2. Set the mutation matrix $\overleftrightarrow{Q}$ equal to the identity matrix and simplify the quasispecies equation to obtain an equation for $\dot{x_i}$.
   3. The frequencies of all the genotypes must add up to 1 ($\Sigma_i x_i = 1$). So what is $\Sigma_i \dot{x_i}$?
   4. Verify your result in part a3 by summing up the simplified quasispecies expression you found for $\dot{x_i}$ in part a2.

b. [6 points] Let's return to the original quasispecies equation. ($\overleftrightarrow{Q}$ is an arbitrary right stochastic matrix.) $\phi(\vec{x}) \equiv \phi(\vec{x}(t))$ denotes the average fitness of the population at time t. Define $\psi(t) = \int_0^t \phi(\vec{x}(s))ds$ and $N_i(t) = N_0 x_i(t) e^{\psi(t)}$, where $x_i(t)$ is the fraction of individuals of type i at time t.

   1. Find out the equation for$\dot{N_i}(t)$.
   2. Define$N_T(t)$=$\sum N_i(t)$. How does $\dot{N_T}(t)$ depend on $\phi(\vec{x})$?
   3. Show that we can interpret N_T as the total population size and N_i as the number of individuals of type i.

[1] Chapter 3 of "Evolutionary Dynamics" by Martin Nowak.

Note that the fitness of a given individual can depend on the actual population size and on the relative frequencies of other individuals. The population size may be important when resources are limiting. On the other hand, frequency dependent selection can arise when there is a cooperative behavior in the population that can be exploited by cheaters. To incorporate these features into the quasispecies equation, the fitness function can be modified to include the relevant dependencies $f_i = f_i(N_T, \vec{x})$.

2 Adaptation in a Sharply Peaked Fitness Landscape (10 points)

You will use the quasispecies equation to explore adaptation in a sharply peaked fitness landscape. In this simple model, you will find that there exists a critical mutation rate u_c such that if the mutation rate μ of the population is larger than μ_c then the fittest sequence cannot be maintained in the population.

Consider a species with a genome of length L and the mutation rate per nucleotide μ. The probability that a sequence replicates itself without any mutations is $q = (1-\mu)^L$. Our sharply peaked fitness landscape will be constructed as follows: there is a single fit sequence (the master sequence), x_M, that has fitness $r > 1$, and all other sequences have fitness 1. We will group all the other sequences together into a single variable called x_O. Assume that at time $t = 0$ the population only consists of fit individuals ($x_M = 1$). The approximate quasispecies equation reads:

$$\begin{aligned} \dot{x_M} &\approx x_M(rq - \phi) \\ \dot{x_O} &\approx x_M r(1-q) + x_O - \phi x_O \end{aligned}$$

a. Some terms have been neglected to acquire the approximate quasispecies equation. Explain what these terms are and why it is reasonable to neglect them.

b. Use the equations above to show that in order to maintain the fittest sequence, rq should be larger than 1.

c. Assume that the fitness advantage of the master sequence is neither too large nor too small, so that $log(r) \approx 1$. Show that this means that $\mu_c \approx \frac{1}{L}$.

3 Repeated Prisoner's Dilemma (12 points)

The Prisoner's Dilemma is a two-player, two-strategy game, with the payoff matrix

	Cooperate	Defect
Cooperate	R	S
Defect	T	P

where $T > R > P > S$ and $R > (T+S)/2$.

Let us consider a repeated Prisoner's Dilemma[2]. Suppose that after each round of game there is a probability w that another round will be played between the two players. Players have a

[2] Chapter 5 of "Evolutionary Dynamics" by Martin Nowak.

good memory and can remember the actions of their opponents in all previous rounds. While defection is an evolutionarily stable strategy (ESS) when individuals know they will never meet again, the finite probability w of repeated interaction may make other strategies evolutionarily stable. Axelrod and Hamilton[3] argued that a strategy called "tit-for-tat" is evolutionarily stable if w is large enough. The tit-for-tat (TFT) strategy is defined as follows:

- Cooperate during the first interaction,
- Do whatever your opponent did at the previous step of the game.

If two players both play TFT again each other, their payoff is

$$R + Rw + Rw^2 + \ldots = \frac{R}{1-w}. \tag{1}$$

a. [6 points] Let's follow Axelrod's and Hamilton's derivation of TFT being evolutionarily stable for a large enough w. In order to prove it, you need to know the that if TFT cannot be invaded by either the always defect (ALLD) strategy or the alternation of defection and cooperation (DC), it is evolutionarily stable.

   1. Find a condition for which TFT cannot be invaded by ALLD. For this, compare the payoff of two players playing TFT (given by Equation 1) and the payoff of ALLD playing against TFT.
   2. Find a condition for which TFT cannot be invaded by DC. (Note: The first play of DC strategy is to defect.)

   We know that TFT is ESS if and only if it is invasible neither by ALLD nor by DC. Thus if both of the above conditions are satisfied by w, TFT is ESS.

b. [6 points] Boyd and Lorberbaum questioned the evolutionary stability of TFT. They used a different definition of evolutionary stability. In particular, they showed that TFT can be invaded by a pair of mutants playing different strategies. In their discussion, they considered the following two strategies: 1) the tit-for-two-tats strategy (TFTT), which allows two consecutive defections before retaliating; and 2) the suspicious-tit-for-tat strategy (STFT), which defects on the first encounter but thereafter plays tit-for-tat.

   Compare how TFT and TFTT behave against STFT. Show that TFT is not evolutionarily stable (if two mutants invade it simultaneously, one playing TFTT, the other playing STFT) for large w. What is the critical w_c?

4 Stochastic Simulations of the Error Threshold (17 points)

In this problem, you will perform stochastic simulations to demonstrate the error threshold concept in finite populations. Let's consider a population of N individuals, each having a genome of length L. Each position of the genome can be in either of the two states: 0 or 1. Every time an individual reproduces, at each position a mutation from 0 to 1 or from 1 to 0 can occur with probability μ. The fitness of a particular sequence is

[3] R Axelrod and WD Hamilton. The evolution of cooperation. Science 211: 1390-1396 (1981)

$$f = L + s_0 \sum_{l=1}^{L} g_l,$$

where s_0 is a constant and g_l is the state of the position l in the genome (either 0 or 1).

a. [1 point] For $s_0 = 0$, what is the steady state distribution of the number of mutations in a given genome? What is the mean?

b. [2 points] What is the condition on N, L and s_0 that makes all the mutations non-neutral? Why is it important that mutations are not neutral?

c. [4 points] **COMPUTATION** For $N = 50$, $L = 10$ and $s_0 = 1$, perform stochastic simulations (Moran process) with various mutation rates (0.005, 0.05, 0.2, 0.5 mutations per position per generation). Remember that after a beneficial mutation has been acquired, the reverse mutation may occur. Plot several trajectories of the mean fitness of the population. Make sure that the trajectories approach a steady state.

d. [2 points] Is the error threshold located in the regime of clonal interference?

e. [3 points] For the mutation rates higher than the error threshold, what is the steady state fitness mean? With the decrease of the mutation rate, how do you expect the steady state mean fitness to change?

Now, consider another fitness function

$$f = \exp(-0.2 \sum_{l} g_l) + 1.3 \exp(- \sum_{l} (1 - g_l)),$$

f. [2 points] Where are the two fitness peaks located?

g. [3 points] **COMPUTATION** Start the simulations at the highest peak. Use the mutation rates of 0.05, 0.075 and 0.1 per position per generation and run the simulation for 10000 replications. Plot the average amount of the mutations in the population as a function of time. Interpret the results.

5 Conditioned Response vs. Direct Response: Anticipation of Sugars in E. coli (19 points)

E. coli can anticipate the future availability of maltose based on the appearance of lactose. Now consider the two signals S_1 (lactose) and S_2 (maltose), separated by a time Δt where S_2 gives rise to a response R_2, in this case R_2 is some protein necessary for the use of maltose as a nutrient. We have two strains, one only responds directly to S_2 to achieve R_2 (direct response, DR), the other one anticipates R_2 by upregulating genes associated with R_2 after the appearance of S_1 (conditioned response, CR). We will identify the conditions when conditioned response is superior to direct response. The first panel of the following figure shows the response of both systems. The solid line corresponds to CR and the dashed line corresponds to the DR.

a. [2 points] For both types of responses, what is the response function $Y(t)$ (i.e. response level) of each system from time $t = 0$ until the disappearance of S_2? Normalize the functions to the steady state of the system.

b. [4 points] Expression of protein necessary to process S_2 carries a cost c (a decrease in growth rate), and a benefit b (an increase in growth rate) due to the advantage given by responding to S_2.

   1. Sketch in the panels the cost and benefit functions of each system. Set each to one if no benefits or cost exists at certain times. Assume that the cost is proportional to the production level and the benefit is proportional to the amount of the protein in a cell.
   2. Calculate the benefit functions $b(t)$ and cost functions $c(t)$ for the systems given that maximum benefit (i.e. the maximum growth rate without any costs) is $1 + \kappa$ and maximum cost (i.e. the minimum growth rate) is $1 - \eta$.

c. [1 point] The fitness of each species can be written as

$$F = \int_0^\infty b(t)c(t)dt,$$

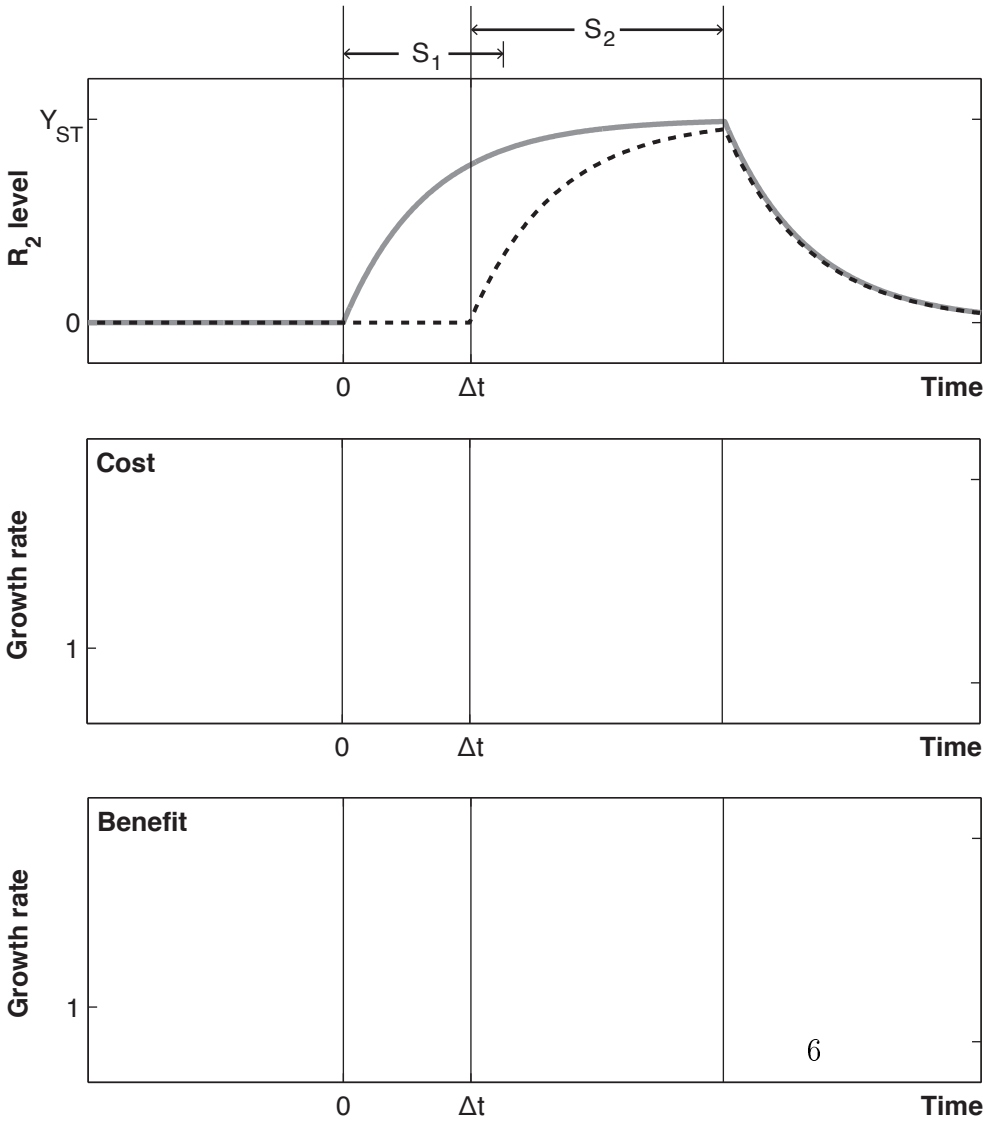
   where $b(t)$ and $c(t)$ are benefit and cost functions. Write down the fitness of CR and DR. Do not evaluate the integrals.

d. [3 points] Write down the relative fitness function $\Delta F_{CR-DR} = F_{CR} - F_{DR}$ for both coupled and uncoupled appearance of signals (i.e. when S_1 and S_2 appear sequentially vs. when only S_1 appears). The duration of the signal S_1 is T_{S_1}. Evaluate the integrals for ΔF_{CR-DR}.

e. [3 points] Let's define p as the probability that S_2 will occur given S_1. Combine the functions derived above into a single equation to give the relative fitness ΔF_{CR-DR} in an environment where both coupled and uncoupled appearances may occur.

f. [3 points] **COMPUTATION** Set $\kappa = 0.17$ and $\eta = 0.045$. The generation time is set to 1 and the duration of the signal S_1 (T_{S_1}) is 0.25. Explore the phase space of Δt and p for parameter values that maximize and minimize the relative fitness ΔF_{CR-DR} .

g. [3 points] **COMPUTATION** Find Δt that gives the maximum overall fitness. Does it make sense?


S1
S2
YST
R2 level
0
0
Δt
Time
Cost
Growth rate
1
0
Δt
Time
Benefit
Growth rate
1
0
Δt
Time

8.591J / 7.81J / 7.32 Systems Biology
Fall 2014