Mutations and Fluctuations

For this lab, we will be simulating the emergence of genetic mutations and analyzing how this leads to diversity in populations inspired by the experiment of Luria and Delbruck. Complete the following questions and address any included questions. For your submission, include any requested figures, answers to all included questions, and your MATLAB code.

Problem 1: Simulation of a Single Generation of Cell Division and Mutation (Demo)

Luria and Delbrück's experimental results were used to illustrate that mutations arise independently from selection, i.e., mutations occur randomly and lead to diversity that propagates through cell replication. Here we will use two different methods to simulate a single generation (emergence of mutations and cell division) of the experiment. (Note that a single generation simulation can correspond to the situation when resistance mutations are caused by the exposure to viruses, i.e. only the most recent generation will produce mutants.)

a) Consider *E. coli* colonies with N=2000 cells in each. Initially, no mutant cells are present. At the time of cell division mutant cell can emerge with a mutational probability of $\mu = 10^{-3}$. What is the average number of mutants you expect to emerge? Write a code that simulates a process of cell division for each cell (looping over them one by one) and uses a random number generator to determine if a mutation has occurred during replication. How many mutations occurred in your simulation?

b) Now assume you have 100 identical colonies as in Parts a. Use your code from (a) to determine the number of mutants in each culture. Plot the distribution of the number of mutants in each colony after a single generation. What are the mean and variance of the distribution?

c) Now consider the same experimental design and simulate a single generation by sampling the number of mutants that will occur from a Poisson distribution with $\lambda = \mu N$. Write a code that takes the number of non-mutated (wild-type) and mutated cells at the beginning of the generation and mutation rate μ and outputs the number of wild-type and mutated cells after one round of division. Consider the same situation as in a) and b), with 100 colonies each with N=2000 cells with a mutational probability of $\mu = 10^{-3}$. Compute and plot the distribution of the number of mutants after one round of division in each colony. What is the mean and variance of the distribution? How does this compare to the methods used in parts a) and b)?

Problem 2: Simulation of Many Generations of Cell Division and Mutation

Now we will simulate many generations of cell division and mutation using the methods from Problem 1 Part c).

a) Write a code that simulates cell division and mutation for g=15 generations that keeps track of the number of starting from N=500 wild-type cells with a mutation rate of 10^{-7}. The code should output the number of mutants and wild-type cells in each generation i=1:g. Plot the number of mutant cells in each generation for runs of your code.

b) Use your code to simulate evolution of resistance in 1000 *E. coli* colonies (independent runs of your code) saving the number of mutants in each generation for each of the run. Plot the distribution of the final number of mutants in each colony. What are the mean and variance of the number of mutants in each colony. How do these values compare to a the result from a single colony you generated in part a)?

c) Use the results of b to find the five colonies with the largest number of mutants after fifteen generations. Additionally, randomly select 5 other colonies from the remaining ones. Plot the

number of cells after each generation for the selected colonies. Compare the dynamics of mutants and comment on what you infer about the events necessary for an emergence of a large number of mutants. Hint: see documentation for MATLAB's maxk command. You can use this to obtain the colonies with the largest number of mutants after fifteen generations.

Problem 3: Inference of Mutation Rate from Experimental Data

In their experiments, Luria and Delbrück used their experimental results to infer the mutation rate leading to bacteriophage resistance in *E. coli*. Use the first 100 colonies from your simulation results in Problem 2 as "experimental data" for this problem.

a) Estimate the most likely mutation rate given your experimental data based on the number of cultures with no mutants. What is your estimated value? How does this compare with the known value of mutation rate, what you used in Problem 2 for generating the data?

b) Now estimate the most likely mutation rate based on the average number of mutations in each colony. What is your estimated value? How does this compare to your results from part a) and the known value in Problem 2? Why might one method of estimating mutation rate be advantageous compared to the other? (Hint: For this you need to use fzero to solve the equation described below. To prevent errors instead of using a single initial guess use [eps 1] as the boundaries.)

Hint: Recall from the reading that we can estimate the mutation rate based on the amount of colonies with zero mutants as $\widehat{\mu_a} = -\frac{\log{(f_0)}}{N}$, where f_0 is the fraction of colonies with zero mutants and N is the number of cells in each colony at the time of exposure to selective pressure (in our case, the final number of cells from Problem 2). Our other method involves modeling the mutation and cell division process as spontaneous and discrete, i.e. fractional values cannot occur. In this case, we obtain an expression for the average number of mutants in each colony $\bar{m} = \mu N_f \log{(C \mu N_f)}$. Here, C is the number of colonies and N_f is the final number of cells in each colony after growth, which is equal to N in our first equation.