Robert Heeter
BIOE 446
6 October 2023

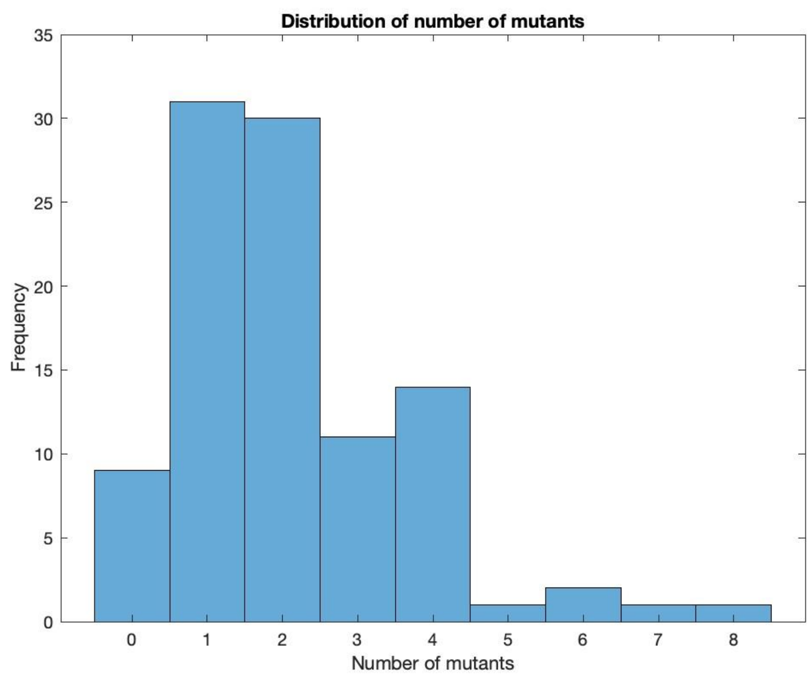
Lab 6: Mutations and Fluctuations

ANSWERS

1. **Simulation of a single generation of cell division and mutation (DEMO)**
    a. The expected number of mutants is 2.
       The simulated number of mutants is 2.

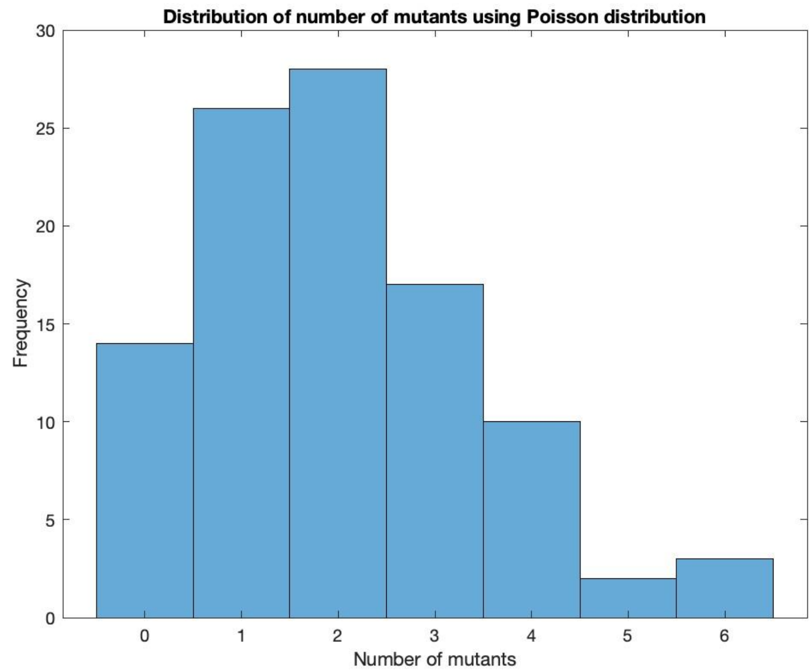
    b. Mean of distribution = 2.1200
       Variance of distribution = 2.3693

       Figure



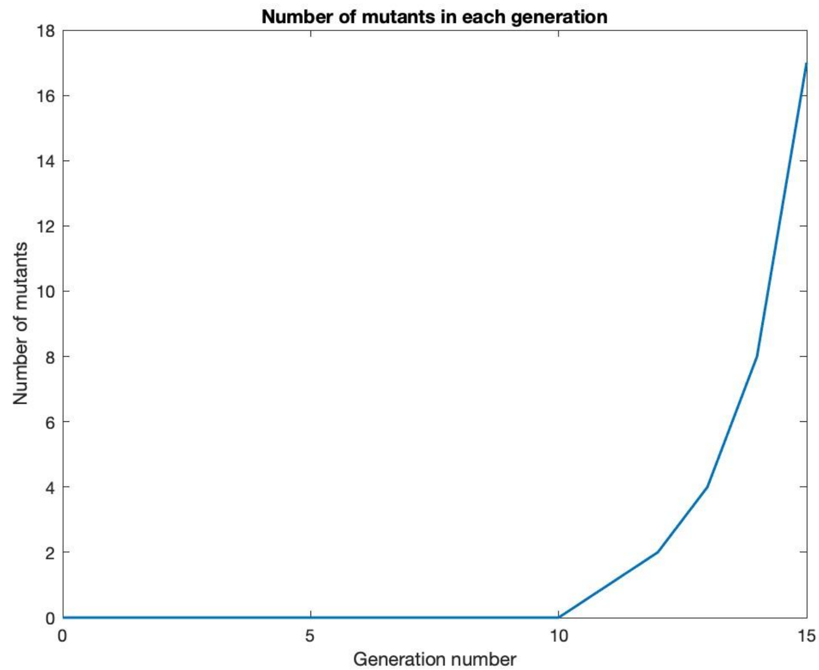
    c. Mean of distribution = 2.0100
       Variance of distribution = 2.0706

       Figure



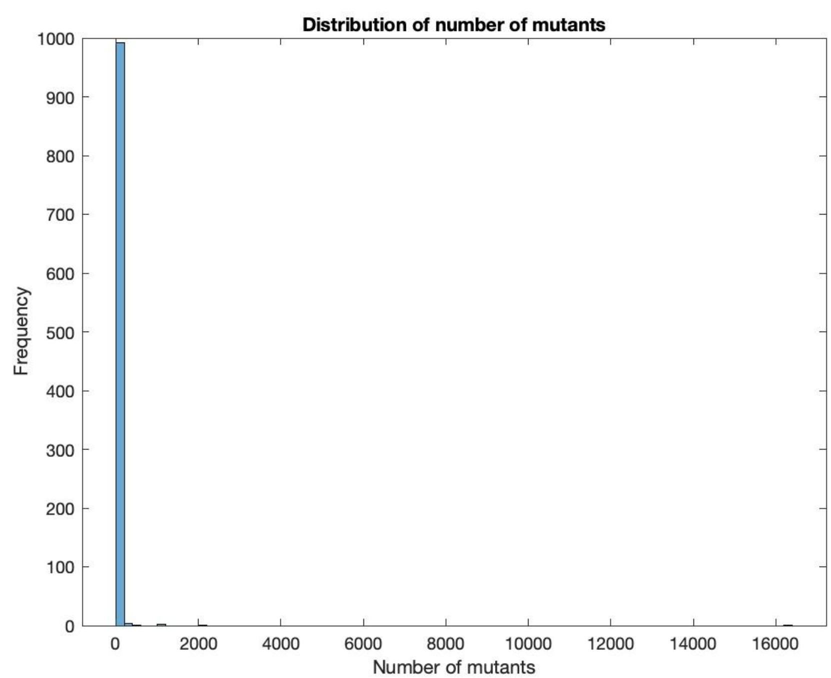
The mean and variance of this distribution are very similar to that in part B (without using the Poisson distribution), though it is clear that this distribution more closely resembles a Poisson curve and that the random distribution has more 'noise'. There are fewer large outliers using the Poisson distribution using only 100 colonies. With enough colonies, the distributions from both should be the same (same mean and same variance).

2. **Simulation of many generations of cell division and mutation**
   a. Figure



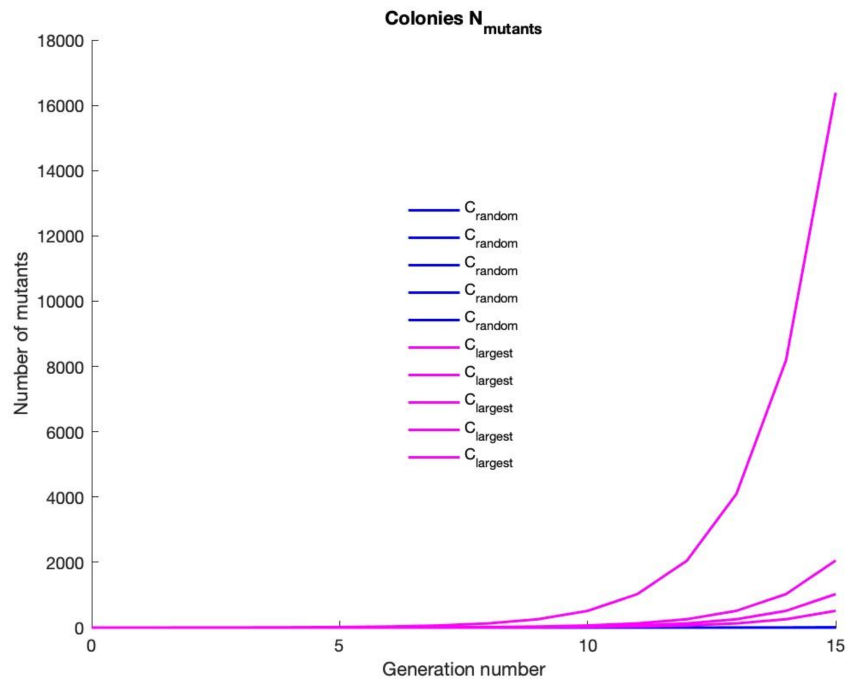
b. Mean of the number of mutants = 16.2490
Variance of the number of mutants = 2.3712e+04

Figure



The vast majority of colonies end with a very small number of mutations after 15 generations (likely less than 100, though the histogram bins *make it hard to determine). The very few colonies that produce mutations in the first few generations end up with a large number of mutations (more than 1000), since the number of mutants increases exponentially with each generation.*

c. Figure



In order for a large number of mutants to emerge, it appears that a mutant must arise at an early generation (low generation number), since the large number of mutants is produced due to exponential growth over a significantly large number of generations.

3. **Inference of mutation rate from experimental data**
   a. Mutation rate as determined based on the number of cultures with no mutants = 5.0109e-08. This value is similar to the mutation rate 1e-7 (10e-8) used in Problem 2 for generating the data. I am a bit confused because I thought that this methodology for determining the mutation rate would result in a much higher value, given that the number of cultures with no mutants after 15 generations would likely be very small (hence giving the impression that the mutation rate is high).

   b. Mutation rate as determined based on the average number of mutations in each colony = 3.9633e-08. This value is also similar to the mutation rate 1-e7 (10e-8) used in Problem 2 and the result from part A. This method is advantageous compared to the method in part A since it accounts for the fact that given enough time (enough generations), mutations will inevitably arise, even with a very small mutation rate.

CODE

```
close all
clc

%% Problem 1: Simulation of a single generation of cell division and mutation (DEMO)
disp('PROBLEM 1')
clear

%% Problem 1, Part A
% defining variables
N = 100;
mu = 1e-3;
N_mutants = 0;

% calculating expected number of mutants
expected_mutants = N*mu;
fprintf('PROBLEM 1A: The expected number of mutants is %i.\n', expected_mutants);

% simulating whether or not mutation occurs during replication
for ii = 1:N
    if rand < mu
        N_mutants = N_mutants + 1;
    end
end

fprintf('PROBLEM 1A: The simulated number of mutants is %i.\n', N_mutants)

%% Problem 1, Part B
% defining variables
C = 100;
N = 2000;
mu = 1e-3;
N_mutants = zeros(1, C);

% iterating through each colony
for ii = 1:C
    % iterating through each cell to see if a mutation occurs during
    % division
    for jj = 1:N
        if rand < mu
            N_mutants(ii) = N_mutants(ii) + 1;
        end
    end
end

figure(1)
histogram(N_mutants);
title('Distribution of number of mutants')
xlabel('Number of mutants')
ylabel('Frequency')
```

```
avg = mean(N_mutants)
variance = var(N_mutants)

%% Problem 1, Part C
% defining variables
C = 4000;
N = 2000;
mu = 1e-3;
N_mutants = zeros(1, C);

% random probability from Poisson distribution
for ii = 1:C
    n_mutations = poissrnd(N*mu);
    N_mutants(ii) = n_mutations;
end

figure(2)
histogram(N_mutants);
title('Distribution of number of mutants using Poisson distribution')
xlabel('Number of mutants')
ylabel('Frequency')

avg = mean(N_mutants)
variance = var(N_mutants)


%% Problem 2: Simulation of many generations of cell division and mutation
disp('PROBLEM 2')
clear

%% Problem 2, Part A
g = 15;
N = 500;
mu = 1e-7;
N_mutants = zeros(1, g+1);

for ii = 2:g+1
    n_mutations = poissrnd(N*mu);
    N_mutants(ii) = N_mutants(ii-1)*2 + n_mutations;
    N = N*2;
end

disp(N_mutants)

figure(3)
plot(0:g, N_mutants, LineWidth=1.5)
title('Number of mutants in each generation')
xlabel('Generation number')
ylabel('Number of mutants')

%% Problem 2, Part B
```

```
g = 15;
C = 1000;
mu = 1e-7;
N_mutants = zeros(C, g+1);

for ii = 1:C
    N = 500;
    for jj = 2:g+1
        n_mutations = poissrnd(N*mu);
        N_mutants(ii, jj) = N_mutants(ii, jj-1)*2 + n_mutations;
        N = N*2;
    end
end

N_mutants_end = N_mutants(:, end);

figure(4)
histogram(N_mutants_end);
title('Distribution of number of mutants')
xlabel('Number of mutants')
ylabel('Frequency')

avg = mean(N_mutants_end)
variance = var(N_mutants_end)

%% Problem 2, Part C
[B, I] = maxk(N_mutants_end, 5);
C_largest_N_mutants = N_mutants(I, :);

N_mutants_other = N_mutants;
N_mutants_other(I, :) = [];
C_random = N_mutants_other(randi([1 (C-5)], 1, 5), :);

figure(5)
hold on
plot(0:g, C_random, '-b', LineWidth=1.5, DisplayName='C_{random}')
plot(0:g, C_largest_N_mutants, '-m', LineWidth=1.5, DisplayName = 'C_{largest}')
title('Colonies N_{mutants}')
xlabel('Generation number')
ylabel('Number of mutants')
legend(Location="best")
legend box off

%% Problem 3: Inference of mutation rate from experimental data
disp('PROBLEM 3')
% clear

%% Problem 3, Part A
[f,~] = size(find(N_mutants(1:100,15)==0));
f0 = f/100;
u = -log(f0)/N
```

```
%% Problem 3, Part B
C = 100;
m = mean(N_mutants(1:100,15));
fsolve(@(mu) mu*N*log(C*mu*N) - m, 1e-7)


%% Functions

% None
```