EE 511 – Project # 0 Muthulakshmi Chandrasekaran

Coin Tossing Experiment USC ID 4486180242

September 6, 2017

Problem Statement:

To simulate Coin Tossing experiment for the given number of trials using Matlab.

Part A: A fair coin has to be tossed for 50 trials, and the outcome of each toss is recorded.

i. To find the number of heads N_H

ii. To find the length of the longest run of heads L_H

Part B: Assuming the same fair coin is used, to find the number of trials needed to obtain the first occurrences of the following sequences

i. Sequence S_1, H

ii. Sequence S_2, HH

iii. Sequence S_3, HHH

iv. Sequence S_4, HHHH

Both Part A and Part B are to be done for 5000 trials, and the results are to be analyzed.

Simulation using Matlab:

Part A:

Matlab Code :

```
% Program to simulate the Tossing of a Fair coin
clc;clear all;close all;
disp('----Experiment : Coin Tossing----');
disp('Input : ');
N=input('Enter the no.of samples:');     % N indicates total number of
                                                               samples
trials=input('Enter the no.of trials :');% trials indicates the total
                             no.of times the experiment has to be done
NH=[];LH=[];                 %arrays for storing values for all trials
for p=1:trials
```

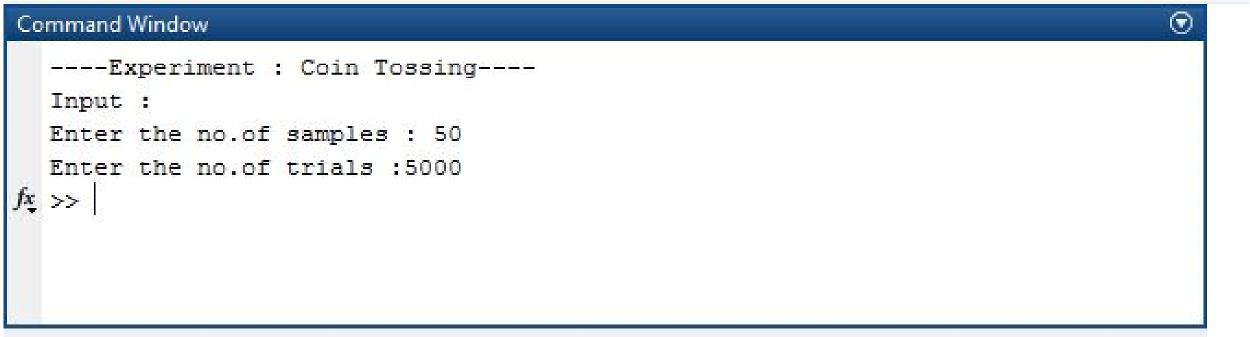
```
     P=zeros(1,N);                      %P indicates the fraction of heads
                                                occuring in the experiment
     H=rand(1,N);                     % H indicates the values of outcomes
                                     of experiment, uniformly distributed
                                     random numbers in range [0,1]
     nh=0;                       % nh denotes the number of heads obtained
     max_lh=0;                 % lh denotes length of longest run of heads
     lh=0;
     for i=1:N
          if(H(i)<=0.5)               % asssume the range for getting the
                                        probability of head as 0 to 0.5.
               nh=nh+1;
               lh=lh+1;
               if(lh>max_lh)         % Checking  for longest run of heads
                    max_lh=lh;
               end
         else
               lh=0;
         end
     P(1,i)=nh/i;
     end
     NH=[NH nh];
     LH=[LH max_lh];
end
hist(NH,1:1:50);                             % Plot the results
figure
hist(LH,1:1:50);
```

<u>Description of the Code :</u>

The number of samples for each trial, N and the total number of trials, $trials$ is obtained from the user. Each outcome of the toss is generated as a random sequence using the command randn. The command $rand\ (1, N)$ generates a 1D array of N elements, with each element in the range [0, 1]. The number of heads in each trial is given by nh. The length of the longest run of heads is given by max_lh. Tossing of a coin has two outcomes, Head and Tail, each with a probability of 0.5. Hence, the range [0, 1] is divided into two intervals – [0, 0.5] and (0.5, 1]. The decision logic used here takes a value in the range [0, 0.5] as Head. Then by simple mathematical calculations, the length of the longest sequence of Heads is found. The values for all trials are stored in an array and then the frequency distribution curve is plotted using the $hist$ command.

<u>Sample Command Window:</u>

The initial inputs are the number of samples and the number of trials. A sample command window is shown in Figure 1, the screen where the number of tosses and the total number of trials for the experiment are given as input.



```
----Experiment : Coin Tossing----
Input :
Enter the no.of samples : 50
Enter the no.of trials :5000
>>
```

Figure 1 Command Window

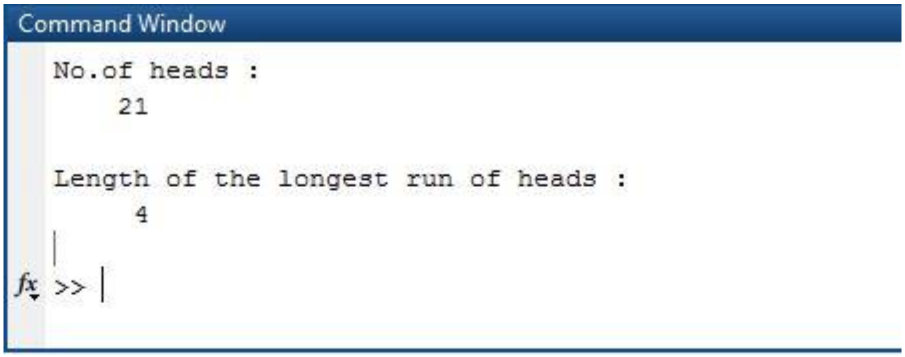
<u>Sample Data:</u>

Using $randn$ function, a set of sample input data for 50 tosses is

Row 1	0.922465	0.996823	0.266027	0.40588	0.841036	0.897987	0.479962	0.083218	0.599403	0.192149
Row 2	0.8608689	0.016781	0.833114	0.527116	0.831325	0.211209	0.844146	0.805952	0.781065	0.228199
Row 3	0.8847168	0.286851	0.893228	0.780943	0.65058	0.775233	0.18971	0.237172	0.720732	0.132715
Row 4	0.404555	0.198507	0.794863	0.781944	0.714748	0.766373	0.833542	0.668718	0.928957	0.955271
Row 5	0.7986401	0.692793	0.856049	0.728139	0.234923	0.174646	0.431165	0.881319	0.184294	0.368547

<u>Probability Measure for One trial:</u>

For one set of input data, the actual values of probability are plotted. The command window for the output is shown in Figure 2(a). The probability graph is shown in Figure 2(b).



```
No.of heads :
    21

Length of the longest run of heads :
     4

>>
```

Figure 2(a) Command window of output for one trial

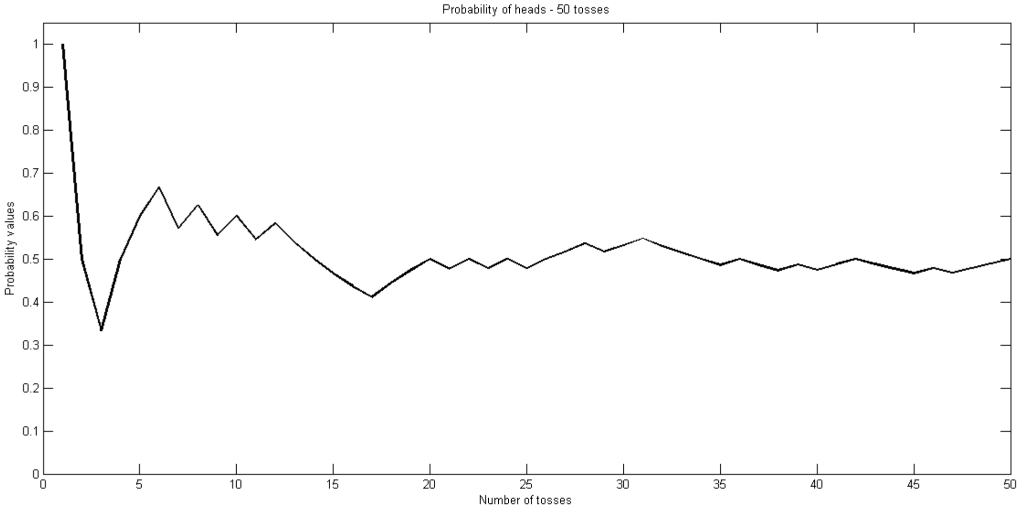


Figure 2(b) Probability values for one trial

Results and Analysis:

It is found that the probability of getting a head is close to 0.5 as the number of trials become very large. The experiment is tried with 500 tosses, 1000 tosses and 10000 tosses. The corresponding graphs are shown in Figure 3(a), Figure 3(b) and Figure 3(c) respectively.

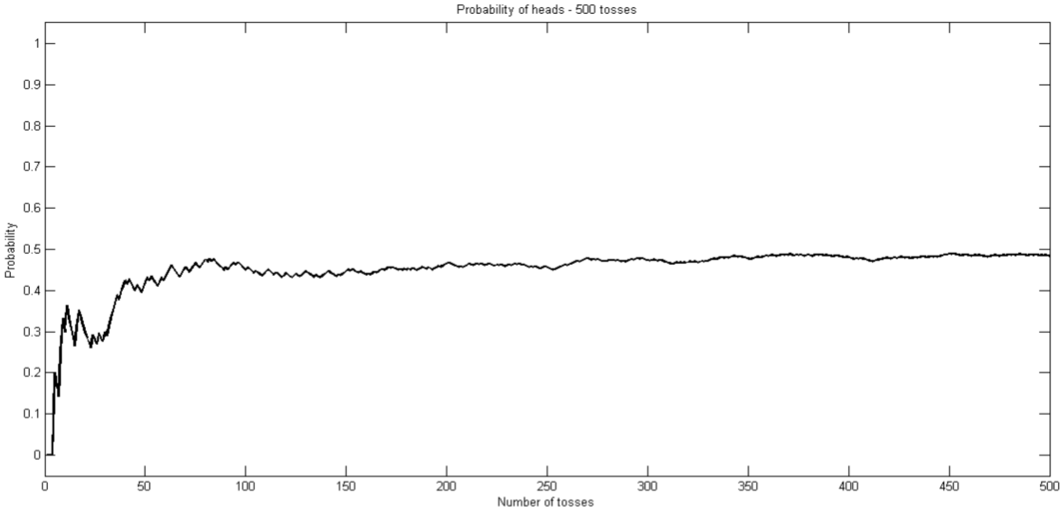


Figure 3(a) Probability of getting heads in 500 tosses

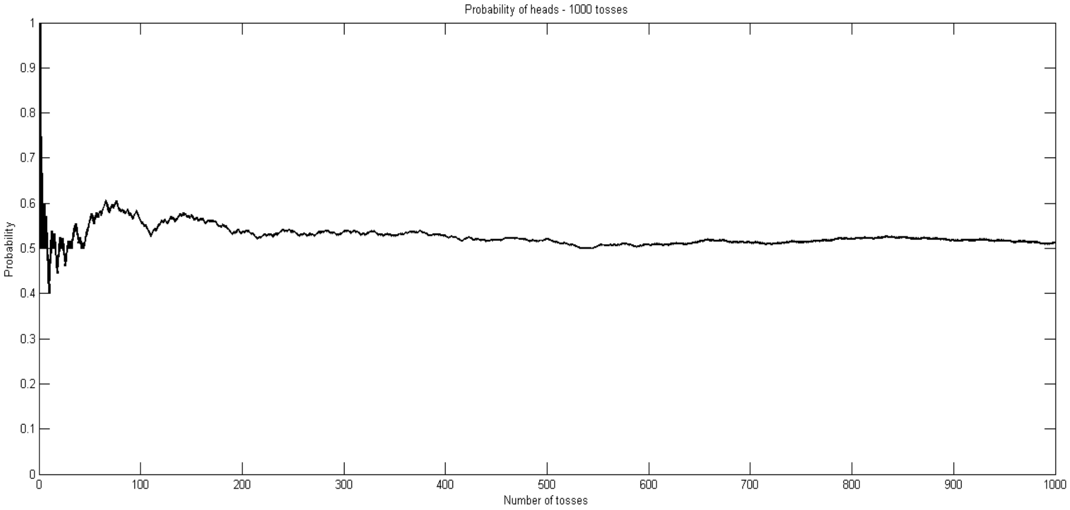


Figure 3(b) Probability of getting heads in 1000 tosses

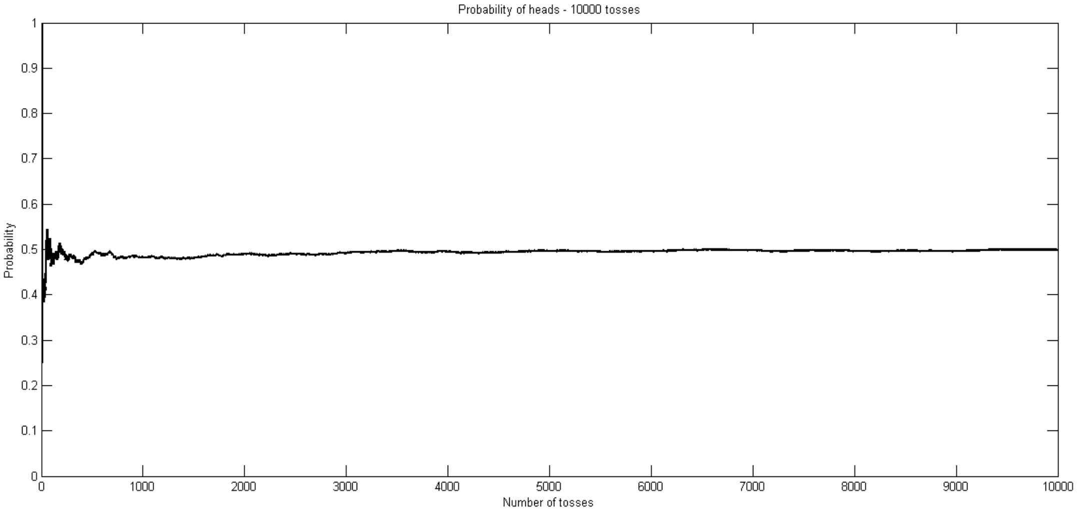


Figure 3(c) Probability of getting heads in 10000 tosses

The number of heads obtained in 5000 trials is plotted, as shown in Figure 4. This is similar to the normal curve. The x axis gives the value of the number of heads occurred in each trial. The y axis gives the number of occurrences of that value of x in the total of 5000 trials. For example, x value of 20 corresponds to the value of 220(appr.) on the y axis. This means that approximately 220 trials out of 5000 trials in the experiment have 20 heads in each trial.

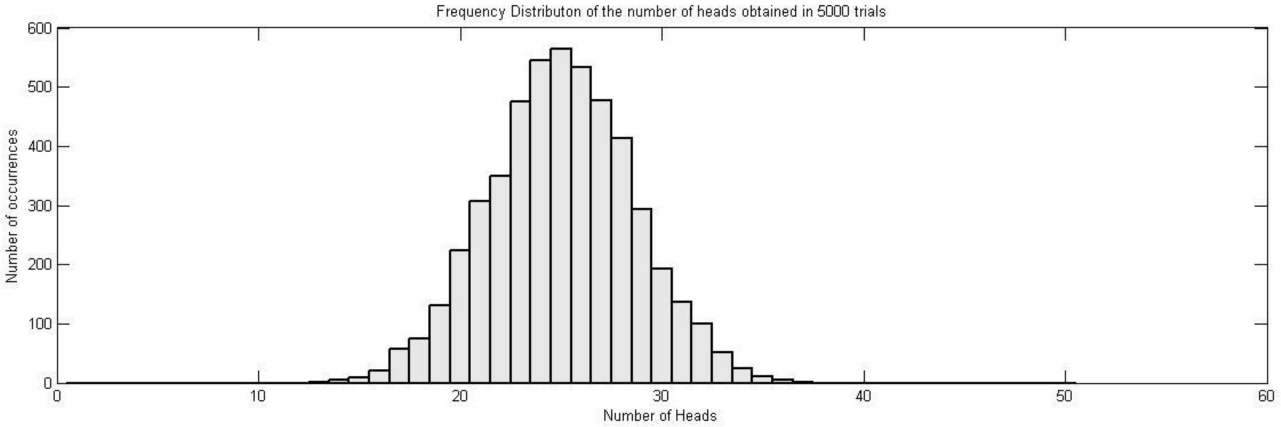


Figure 4 Plot of frequency distribution of the number of heads obtained in 5000 trials

The longest run of heads is plotted as shown in Figure 5. The x axis gives the value of the longest number of heads occurring in each trial. The y axis gives the number of occurrences of that value of x in the total of 5000 trials. For example, x value of 5 corresponds to the value of 1200(appr.) on the y axis. This means that approximately 1200 trials out of 5000 trials in the experiment have a maximum of 5 consecutive heads occurring in their outcomes.

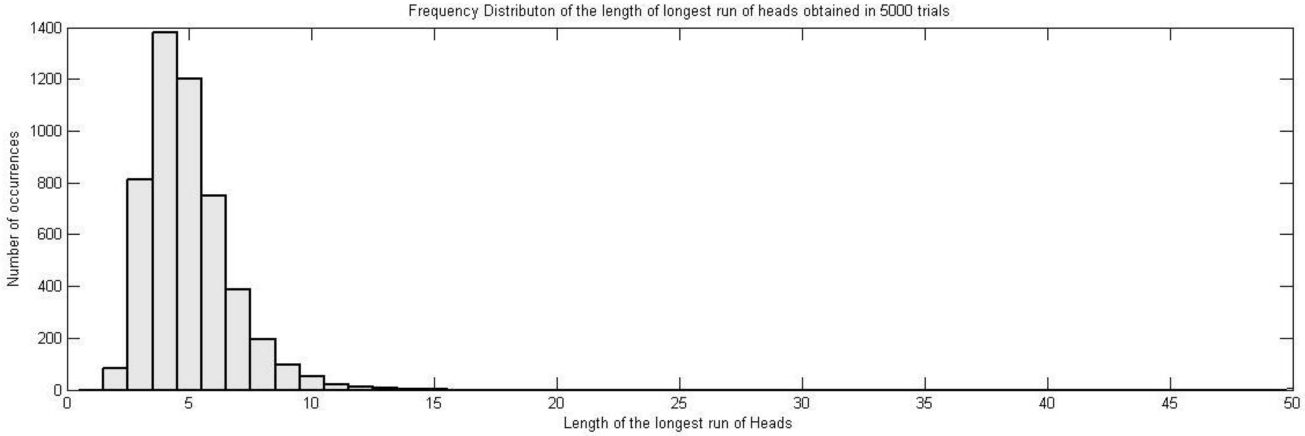


Figure 5 Plot of frequency distribution of the length of longest number of heads obtained in 5000 trials

Part B:

Matlab Code :

```
% Program to simulate the Tossing of a Fair coin
clc;clear all;close all;
%disp('----Experiment : Coin Tossing----');
N=50;                            % N indicates total number of tosses
                                                             per trial
h1_arr=[];h2_arr=[];h3_arr=[];h4_arr=[]; %arrays for storing sequence
                                               values for 5000 trials
for trials=1:5000

     P=zeros(1,N);                % P indicates the fraction of heads
                                            occuring in the experiment
     H=rand(1,N);
     lh=0;
     h1=0;h2=0;h3=0;h4=0;                     % Variables for calculation
     i=1;
     while(i<N) % Logic for finding the sequence S1
          if(H(i)<=0.5)           % Find the first occurrence of Head
                                    i.e first occurrence of an element
                                   in [0,0.5] & then break the loop.
               h1=i;
               break;
          end
          i=i+1;
     end
     while(i<N-1)
          if(H(i)<=0.5 && H(i+1)<=0.5)   % Find first occurrence of HH
                                      i.e two consecutive elements in
                                         [0,0.5]& then break the loop.
               h2=i+1;
               break;
          end
          i=i+1;
     end
     x=0;
     while(i<N-2)
          if(H(i)<=0.5 && H(i+1)<=0.5&&H(i+2)<=0.5)
                                     % Find first occurrence of HHH
                                     i.e three consecutive elements in
                                         [0,0.5]& then break the loop.
               h3=i+2;
               x=1;
               break;
          end
          i=i+1;
     end
     if(x==0)
         h3=N;
     end
     x=0;
     while(i<N-3)
          if(H(i)<=0.5 && H(i+1)<=0.5&&H(i+2)<=0.5&&H(i+3)<=0.5)
                                      % Find first occurrence of HHHH
                                      i.e four consecutive elements in
                                         [0,0.5]& then break the loop.

               h4=i+3;
               x=1;
               break;
          end
          i=i+1;
     end
     if(x==0)
          h4=N;
     end
     h1_arr=[h1_arr h1];                  %Store the values for S1 in
                                         array h1_arr for 5000 trials
     h2_arr=[h2_arr h2];                  %Store the values for S2 in
                                         array h2_arr for 5000 trials
     h3_arr=[h3_arr h3];                  %Store the values for S3 in
                                         array h3_arr for 5000 trials
     h4_arr=[h4_arr h4];                  %Store the values for S4 in
                                         array h4_arr for 5000 trials

end

% Plot frequency distribution for each of sequences S1,S2,S3 and S4
hist(h1_arr,1:1:50);title('h1');figure
hist(h2_arr,1:1:50);title('h2');figure
hist(h3_arr,1:1:50);title('h3');figure
hist(h4_arr,1:1:50);title('h4');
```

Description of the Code :

The number of tosses for each trial is taken as 50 and 5000 trials are done. Similar conditions as in Part A are considered. The command $rand(1, N)$ is used to generate a 1D array of N random elements, with each element in the range [0,1], and any element within the range [0,0.5] is considered Head. Then by simple mathematical calculations, Sequence S1, which gives the occurrence of first Head is obtained. Similar calculations are done to obtain Sequence S2, S3 and S4 that give the minimum number of trials required to obtain 2, 3 and 4 consecutive Heads respectively. The values for all trials are stored in an array and then the frequency distribution curve is plotted using the $hist$ command.

Results and Analysis :

The number of tosses required to get the first Head is the Sequence S1. For 5000 trials, the frequency distribution is as in Figure 6. From the plot, it is observed that in 2050(appr.) trials, the first toss yields a head.

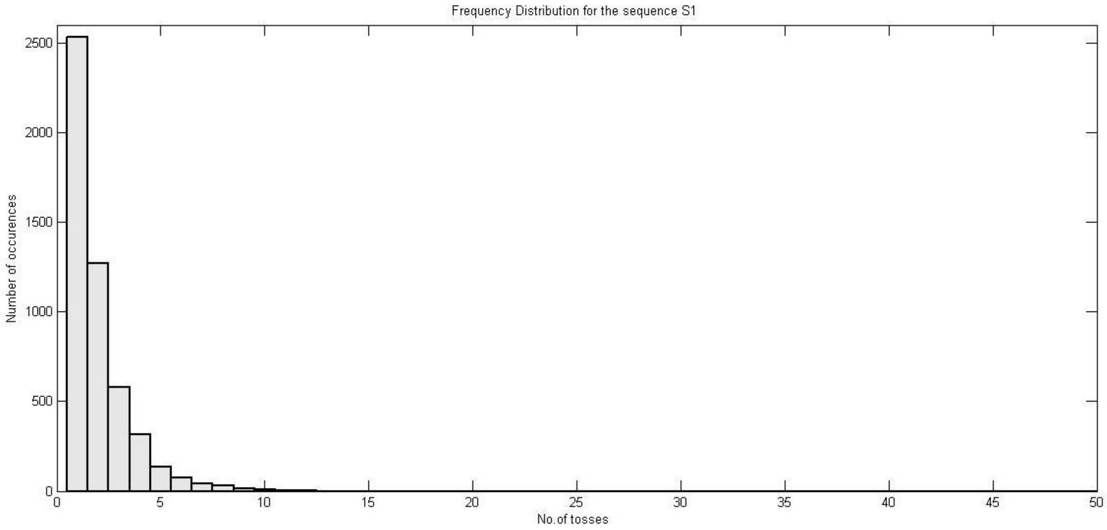


Figure 6 Frequency distribution for Sequence S1

The number of tosses required to get the first two consecutive Heads is the Sequence S2. For 5000 trials, the frequency distribution is as in Figure 7. From the plot, it is observed that in 1250(appr.) trials, two tosses are needed to get the sequence S2. The logic is right, because the value for one toss is zero.

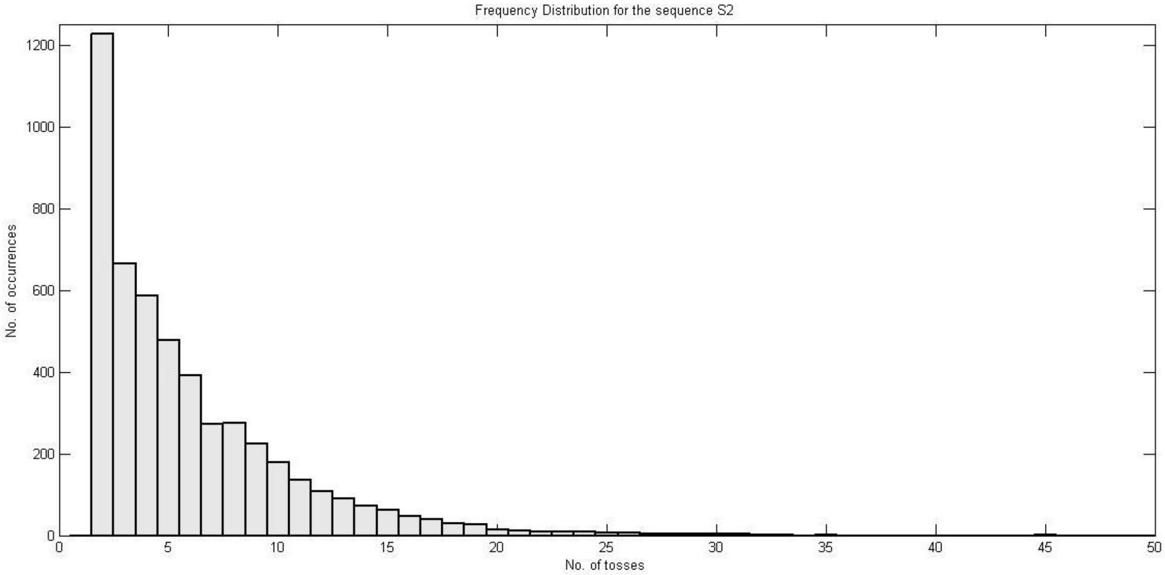


Figure 7 Frequency distribution for Sequence S2

The number of tosses required to get the first three consecutive Heads is the Sequence S3. For 5000 trials, the frequency distribution is as in Figure 8. From the plot, it is observed that in 600(appr.) trials, three tosses are needed to get the sequence S3. The logic is right, because the value for the first two tosses is zero.

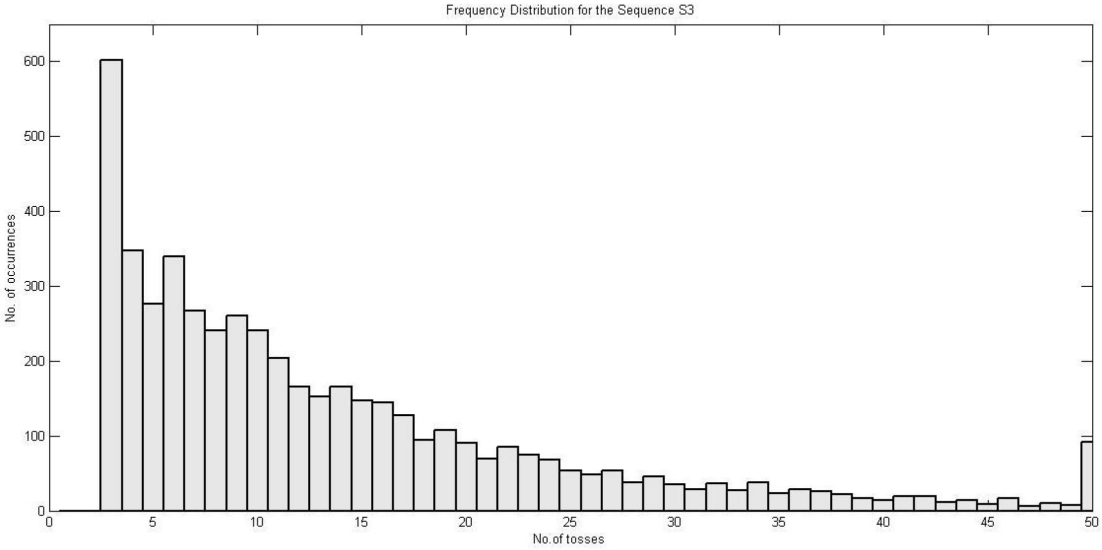


Figure 8 Frequency distribution for Sequence S3

The number of tosses required to get the first four consecutive Heads is the Sequence S4. For 5000 trials, the frequency distribution is as in Figure 9. From the plot, it is observed that in 290(appr.) trials, four tosses are needed to get the sequence S4. The logic is right, because the value for the first three tosses is zero.

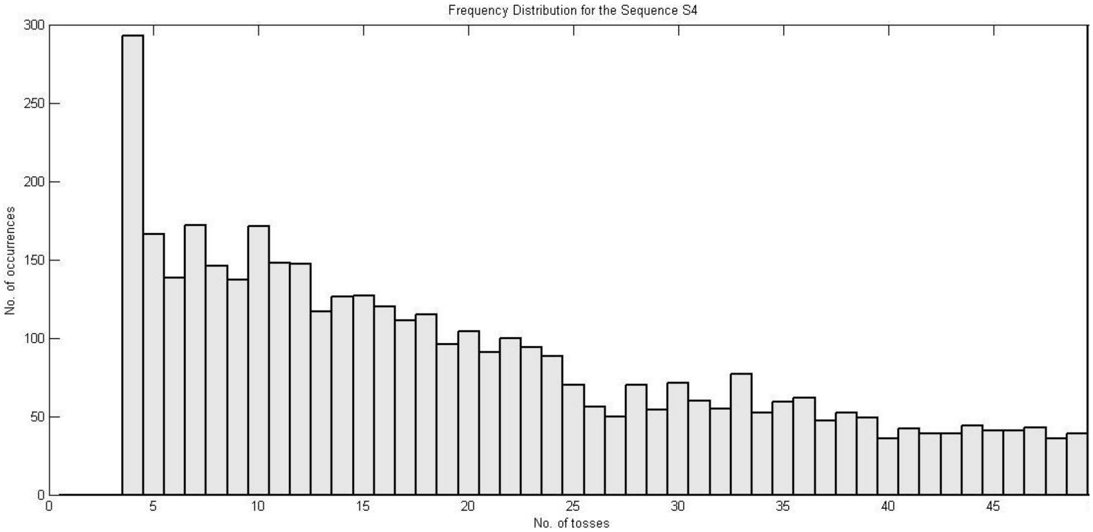


Figure 9 Frequency distribution for Sequence S4

Conclusion and Summary of the Experiment :

The simulation is done for 5000 trials, and in each trial 50 tosses are done. The outcome of each toss is taken as a random number in the range [0,1] using the $randn$ command. In the first part, number of heads and the length of the longest run of heads in one trial are found. The probability of finding a head in any toss is graphed. On doing the experiment for 5000 trials, the frequency distribution plots of number of heads and length of longest run of heads is plotted. The number of tosses required to get a sequence follows geometric distribution. In the second part, the number of tosses required for the first occurrence of the sequences S1(H), S2(HH), S3(HHH) and S4(HHHH) are found. For 5000 trials, their frequency distributions are plotted.