Problem Statement

We need to do following simulations of some interesting discrete random variables.

1. **Sum of Uniform RV's**

   Define:

   $N = \text{Min}\left\{ n : \sum_{i=1}^{n} U_i > 1 \right\}$

   where $\{U_i\}$ are iid Uniform (0,1) RV's. Find (by simulation): $\hat{m} = E[N]$ an estimator for the mean.

2. **Minima of Uniform RV's**

   Define:

   $N = \text{Min}\{ n : U_1 \le U_2 \le \ldots \le U_{n-1} > U_n \}$

   i.e. the *n*th term is the first that is less than its predecessor, where {Ui} are independent identically distributed (iid) Uniform (0,1) RV's. Find (by simulation): $\hat{m} = E[N]$an estimator for the mean.

3. **Maxima of Uniform RV's**

   Consider sequence of iid Uniform RV's $\{U_i\}$. If $U_j > \max_{i=1:j-1}\{U_i\}$ we say U_j is a record.

   Example: the records are underlined.

   $$\{U_i\} = \{\underline{0.2314}, \underline{0.4719}, 0.1133, \underline{0.5676}, 0.4388, \underline{0.9453}, \ldots.\}$$

   (note that the U_i are on the real line and we are just showing 4 digits of precision).

   Let X_i be an RV for the distance from the $i-1^{\text{st}}$ record to the i^{th} record. Clearly $X_1 = 1$ always. In this example, $X_2 = 1, X_3 = 2, X_4 = 2$.

   Distribution of Records: Using simulation, obtain (and graph) a probability histogram for *X*2 and *X*3 and compute the sample means. Can you find an analytical expression for *P*(*X*2=*k*)? What does this say about *E* [X2]?

Problem 1

Theoretical Explanation

Suppose X_1, X_2, …., Xn are iid uniform random variables in range [0,1]. The PDF for $Y = X1 + X2 + \cdots + Xn$ is given by (Reference 1),

$$f_n(x) = \frac{1}{2(n-1)!} \sum_{k=0}^{n} (-1)^k \binom{n}{x} sgn(x-k)(x-k)^{n-1}$$

Where,
n = number of uniform RVs added
sgn = signum function

The event N=n occurs when sum of n-1 RVs is less than 1 and sum of n RVs exceeds 1. The probability of N=n is given by,

$$P(N = n) = \int_1^n f_n(x)dx - \int_1^{n-1} f_{n-1}(x)dx$$

We can simplify the above integral as given, (Reference 1)

$$P(N=n) = \left[1 - \frac{1}{n!}\right] - \left[1 - \frac{1}{(n-1)!}\right]$$

$$\therefore P(N=n) = \frac{1}{n(n-2)!}$$

The expected value for N can be given by,

$$E[N] = \sum_{n=2}^{\infty} \frac{n}{n(n-2)!}$$

$$\therefore E[N] = \sum_{n=2}^{\infty} \frac{1}{(n-2)!}$$

$$\therefore E[N] = e = 2.7183$$

Simulation

Here, we generate a iid uniform RV between 0 and 1, add them until their sum exceeds 1. If sum exceeds 1 then we store the number of iids required for sum to exceed 1 in array N. Then we plot the probability distribution of N and calculate the value of E[N].

Code:

```
clc
clear all
close all
flag=0;
total=0;
count=0;
n_of_runs=100000; % Number of simulations to get more closer value of E[N]
N=zeros(n_of_runs,1);
for i=1:n_of_runs
    flag=0;
    while(flag==0)
        xn=rand(1,1); % Generate i.i.d uniform RV between 0 and 1
        total=total+xn;
        count=count+1;
        if(total>1) % If total exceeds 1 then we store the index of i.i.d
            N(i)=count;
            total=0;
            count=0;
            flag=1;
        end
    end
end
tb=tabulate(N); % Generate histogram results for N
bar(tb(:,1),tb(:,2)/sum(tb(:,2))); % Plot probability distribution of N
xlabel('Value of N')
ylabel('Probability')
title('Probability distribution of N')
hold on
mean(N); % Calculate E[N]
k=(2:10);
stem(k,1./(k.*factorial(k-2)),'red') % Plot theoretical probability
distribution of N
```

Output:

Estimated value of E[N] = 2.7180 for 100000 samples which is very closer to theoretical value of E[N] = e = 2.7183.

The probability distribution of N can be seen in figure 1. We can clearly see that the random variable N has Geometric Distribution. We can also see that the estimated probability distribution is similar as theoretical distribution which is $P(N = n) = \frac{1}{n(n-2)!}$

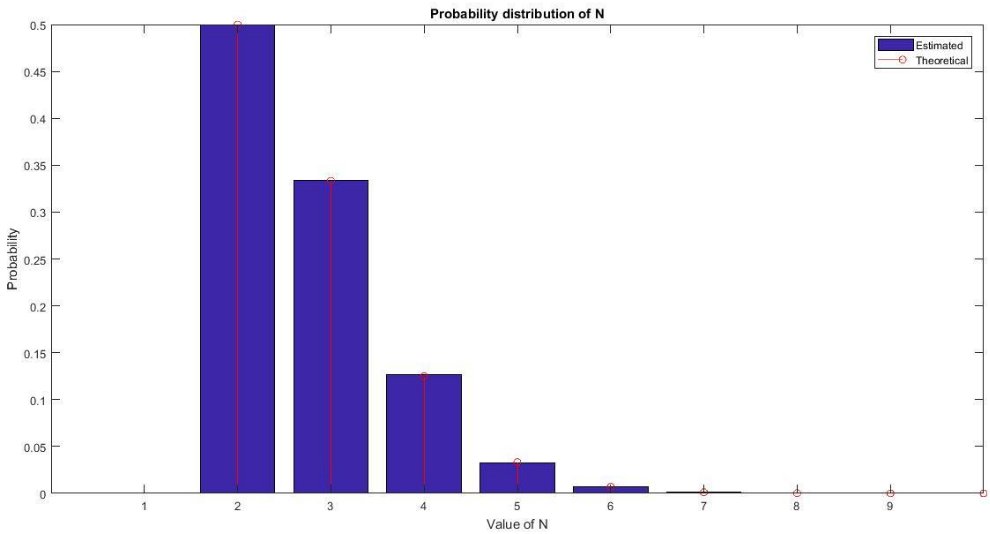

Figure 1 Theoretical V/S Estimated Probability distribution of N

Problem 2

Theoretical Explanation

Consider the event N: $\{U_1 \leq U_2 \leq \ldots \leq U_{n-1} > U_n\}$.

Suppose for n=3, N: {0.1, 0.5, 0.3} then possible arrangements to meet condition for RV are {0.3, 0.5, 0.1} and {0.1, 0.5, 0.3}. So, P(N=3) = 2/3!

Suppose for n=4, N: {0.1, 0.2, 0.6, 0.4} then possible arrangements to meet condition for RV are {0.2, 0.4, 0.6, 0.1}, {0.1, 0.2, 0.6, 0.4} and {0.1, 0.4, 0.6, 0.2}. So, P(N=4) = 3/4!

From the above two examples, we can generalise,

$$P(N = n) = \frac{n-1}{n!} = \frac{1}{n(n-2)!}$$

We can see that probability distribution of “Minima of Uniform RVs” is same as probability distribution of “Sum of Uniform RVs”.

$$\therefore E[N] = e = 2.7183$$

Simulation

Here we generate iid uniform RV between 0 and 1 and continuously compare the next term with the previous one. If the next term is less than the previous term then we store the index of iid in array N.

Code:

```
clc
clear all
```

```
close all
n_of_runs=100000; % Number of simulations to get more closer value of E[N]
N=zeros(n_of_runs,1);
for i=1:n_of_runs
    count=2;
    flag=0;
    xp=rand(1,1); % Generate the first iid uniform RV
    while(flag==0)
        x=rand(1,1); % Generate next iid uniform RV
        if(x>=xp) % If next term is less than previous term then we store
the index of iid
        count=count+1;
        xp=x;
        else
            N(i)=count;
            count=2;
            flag=1;
        end
    end
end
tb=tabulate(N); % Generate histogram results for N
bar(tb(:,1),tb(:,2)/sum(tb(:,2))); % Plot probability distribution of N
xlabel('Value of N')
ylabel('Probability')
title('Probability distribution of N')
hold on
mean(N); % Calculate E[N]
k=(2:9);
stem(k,1./(k.*factorial(k-2)),'red') % Plot theoretical probability
distribution of N
legend('Estimated','Theoretical')
```

Output:

Estimated value of E[N] = 2.7229 for 100000 samples which is very closer to theoretical value of E[N] = e = 2.7183.

The probability distribution of N can be seen in figure 2. We can clearly see that the random variable N has Geometric Distribution as in problem 1. We can also see that the estimated probability distribution is similar as theoretical distribution which is $P(N = n) = \frac{1}{n(n-2)!}$

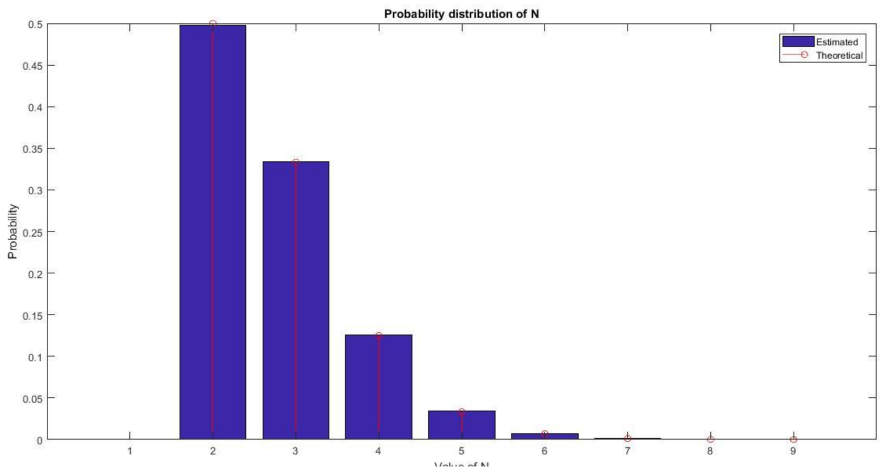

Figure 2 Theoretical V/S Estimated probability distribution of N

Problem 3

Theoretical Explanation

Consider the sequence {0.6, 0.5, 0.2, 0.3, 0.8, 0.1, 0.9}, according to the definition of the random variable X_1=1 (for record 0.6), X_2=4(for record 0.8>0.6 because it's the second maxima) and X_3=2(for record 0.9>0.8 because it's the third maxima). We need to find the probability distributions of X_2 and X_3.

Simulation

Here, we generate a sequence of iid uniform random variables between 0 and 1. We assign first term in sequence as maximum. Then we iterate through the sequence and check whether next term is greater than maximum. If yes then second maxima have occurred and we store the value of term index-1 as X_2 and store the value of index for calculation of X_3. Again, the loop iterates and when third maxima occurs, we store value of X_3. At the end, we plot the probability distribution of X_2 and X_3.

Code:

```
clc
clear all
close all
n_of_runs=100000; % Number of simulations to get more closer value of E[N]
n_of_sample_points=1000; % Number of iid uniform RV between 0 and 1 in a
sequence
N=zeros(n_of_sample_points,1);
for j=1:n_of_runs
    Ui=rand(1,n_of_sample_points); % Generate a sequence of iid uniform RV
    max=Ui(1); % Assign first number of sequence as maximum (first maxima)
    count=0;
    flag=0;
    for i=1:n_of_sample_points
        if(Ui(i)>max) % If next number of sequence is more than max then
maxima occured
            max=Ui(i);
            flag=flag+1;
            if flag==1
                X2(j)=i-1; % When second maxima occurs value of X2 is
stored
                count=i;   % Index of X2 is stored for calculation of X3
            end
            if flag==2
                X3(j)=i-count; % When third maxima occurs value of X2 is
stored
            end
        end
    end
end
figure('Name','X2');
tb1=tabulate(X2); % Generate histogram values of X2
bar(tb1(:,1),tb1(:,2)/sum(tb1(:,2))); % Plot probability distribution of X2
title('Probability distribution of X2');
xlabel('Values of X2');
ylabel('Probability');
xlim([0 25])
figure('Name','X3');
tb2=tabulate(X3); % Generate histogram values of X2
bar(tb2(:,1),tb2(:,2)/sum(tb2(:,2))); % Plot probability distribution of X2
title('Probability distribution of X3');
xlabel('Values of X3');
```

```
ylabel('Probability');
```

Output:

For 100000 runs,
Sample mean of X2 = 6.5012
Sample mean of X3 = 20.3191

The estimated probability distribution of X_2 and X_3 are shown in figure 3 and 4.

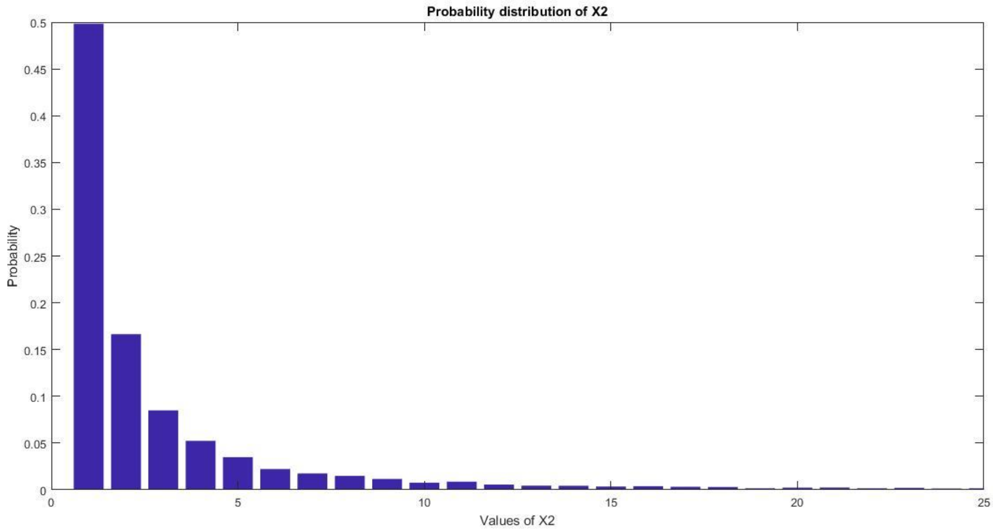

Figure 3 Estimated probability distribution of X2

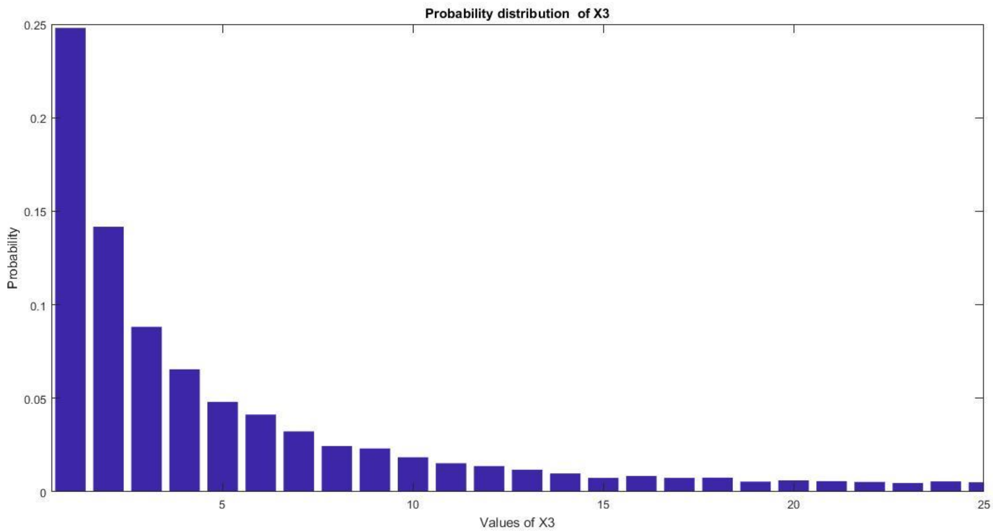

Figure 4 Estimated probability distribution of X3

The distribution looks interesting because they are the geometrical distribution which can be seen from the value of probabilities as shown in Table 1.

Value of X2	Probability (P(X2=k))	Equivalent theoretical value
1	0.4998	1/1*2
2	0.1626	1/2*3

3	0.0826	1/3*4
4	0.0553	1/4*5
5	0.0336	1/5*6

Table 1 Probability distribution of X2

We can generalise,

$$P(X_2 = k) = \frac{1}{k(k+1)}$$

The expected value of X2 is given by,

$$E[X_2] = \sum_{k=1}^{\infty} \frac{k}{k(k+1)} = \sum_{k=1}^{\infty} \frac{1}{(k+1)} = \infty$$

Thus, expected value of X_2 cannot be found because it is divergent in nature.

References

[1] http://www.randomservices.org/random/special/IrwinHall.html