Project 3： Some Interesting DRV's

EE 511 – Section Tuesday 5:00pm—5:50pm
Name: Yiqi Feng
Student ID #: 9057129035

Problem A

1. Problem Statement

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.

Can you guess (or derive) the true value for $E[N]$?

2. Theoretical Analysis

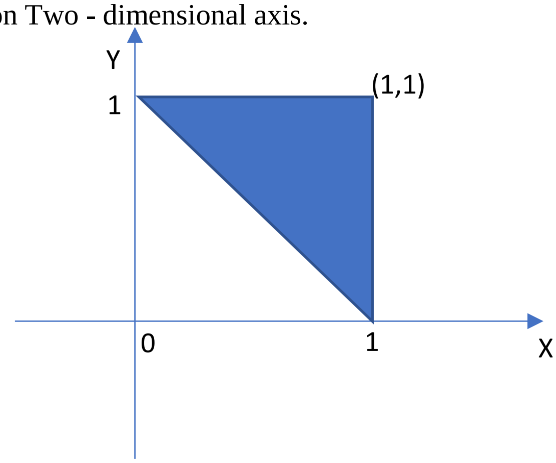
For N = 1, we can assume that X is uniform random variable in (0,1).
Because X < 1, N doesn't exist.
Thus, we can get that:

$$P[N = 1] = 0$$

For N = 2, we can assume that X and Y are uniform random variables in (0,1).
Thus, we can get that:

$$P[N = 2] = P[X + Y > 1]$$

we can shadow the goal area on Two - dimensional axis.



We can get that:

$$P[N = 2] = P[X + Y > 1] = 1 - \frac{1}{1 \times 2} = \frac{1}{2}$$

For N = 3, we can assume that X, Y and Z are uniform random variables in (0,1).
Thus, we can get that:

$$P[N = 3] = P[X + Y + Z > 1]$$

we can map the three random variables in Three - dimensional space.
Thus, we can get that:

$$P[X + Y + Z > 1] = 1 - \frac{1}{1 \times 2 \times 3} = \frac{5}{6}$$

$$P[N = 3] = P[X + Y + Z > 1] - P[X + Y > 1]$$

$$= \left(1 - \frac{1}{1 \times 2}\right) - \left(1 - \frac{1}{1 \times 2 \times 3}\right)$$

Thus, when it comes to n, we can infer that:

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

When $n \to \infty$:

$$E[N] = \sum_{n=2}^{\infty} \frac{n-1}{(n)!} \times n = \sum_{n=0}^{\infty} \frac{1}{(n)!} = e \approx 2.718$$

3. Simulation Methodology

In this problem, I continuously generate random variable U_i in (0,1) until $\sum_{i=1}^{N} U_i > 1$ where the value N is recorded. Repeat simulation 100000 times, and record each N and calculate mean value in every simulation.

4. Experiments and Results

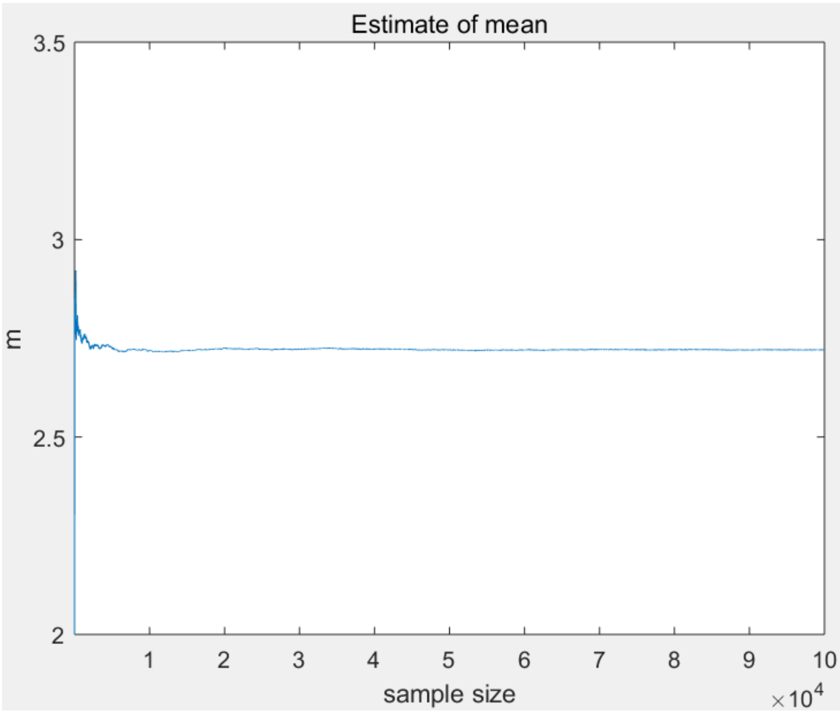


Figure 3.1 Estimate of mean in different sample size

From the figure3.1, when the experiment time start from 1 to 100000, we can see that the larger the sample size is, the closer to e the value of mean is. Also, we can see the value of estimate is floating up and down around e, and the range of the floating is becoming smaller and smaller. Therefore, we can predict that when the sample size is infinite, the estimate will be the true value, e.

When sample size is 100000, the frequency distribution of N is presented below,

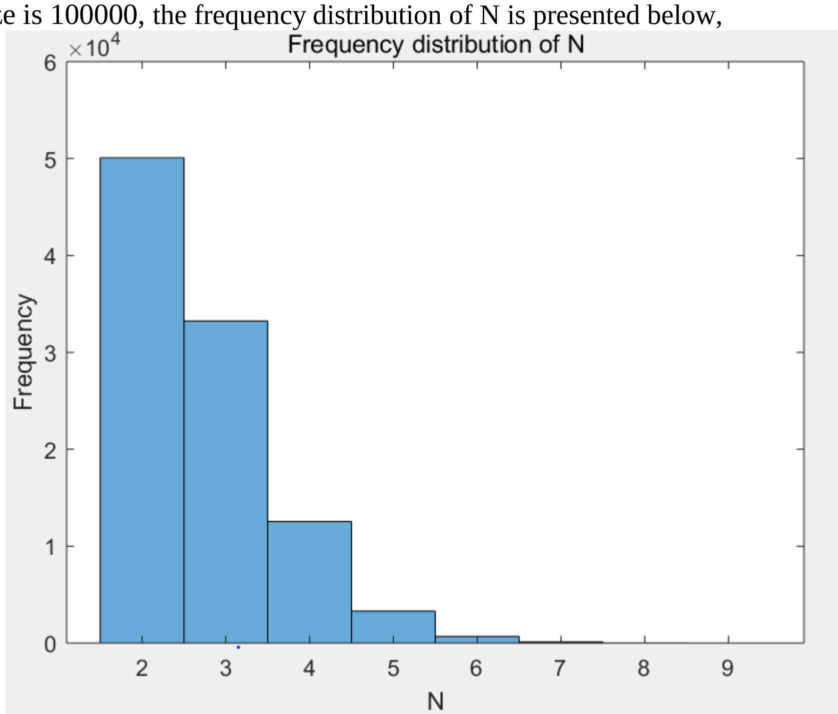


Figure 3.2 Frequency distribution of N

```
m =

    2.7173
```

Figure 3.3 Mean value by simulation(100000 times)

In figure 3.3, the simulation of the mean value is 2.7173, which is close to the true value of E[N]=e≈ 2.718.

5. Source Code

The code for Problem A:

```
N=zeros(1,100000);
m=zeros(1,100000);
for i=1:100000
   sum = 0;
   k=0;
   while sum <= 1
        sum= sum + rand();
        k=k+1;
   end
   N(i)=k;
   m(i) = mean(N(1:i));
end
figure(1);
plot(m(1,1:100000))
axis([1 100000 2 3.5]);
title('Estimate of mean');
ylabel('m');
xlabel('sample size');
figure(2);
histogram(N)
title('Frequency distribution of N');
ylabel('Frequency');
xlabel('N');
```

Problem B

1. Problem Statement

Minima of Uniform RV's

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

i.e. the n^{th} term is the first that is less than its predecessor, where $\{U_i\}$ are independent identically distributed (iid) Uniform(0,1) RV's.

Find (by simulation): $\hat{m} = E[N]$ an estimator for the mean.

Can you guess (or derive) the true value for $E[N]$?

2. Theoretical Analysis

Assume we generate two random numbers U_1 and U_2which are both uniform random variables in (0,1).

$$P[N = 2] = P[U_2 > U_1] = \frac{1}{2}$$

Assume we generate three random numbers U_1 ,U_2 and U_3 which are all from uniform random variables in (0,1). There are 3! possible results:

$$\{U_1 \leq U_2 \leq U_3, U_1 \leq U_3 \leq U_2, U_2 \leq U_1 \leq U_3, U_2 \leq U_3 \leq U_1, U_3 \leq U_1 \leq U_2, U_3 \leq U_2 \leq U_1\}$$

The possibility of each outcome is equal to 1/6. So we can get that:

$$P[n = 3] = P[U_1 \leq U_3 < U_2] + P[U_3 \leq U_1 \leq U_2] = \frac{1}{6} \times 2 = \frac{3-1}{3!}$$

Therefore, we can refer that for n random numbers $\{ U_1, U_2 \cdots U_n\}$, there are $n!$ results. And there are $(n-1)$ results which could meet the condition of $U_1 \leq U_2 \leq U_3 \cdots U_{n-1} > U_n$. So we can get that:

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

When $n \to \infty$:

$$\text{E}[\text{N}] = \sum_{n=2}^{\infty} \frac{n-1}{(n)!} \times n = \sum_{n=0}^{\infty} \frac{1}{(n)!} = e \approx 2.718$$

3. Simulation Methodology

In this problem, I continuously generate variable U_i in (0,1) until $U_1 \leq U_2 \leq U_3 \cdots U_{n-1} > U_n$ where the value N is recorded. Repeat simulation 100000 times, and record each N and calculate mean value in every simulation.

4. Experiments and Results

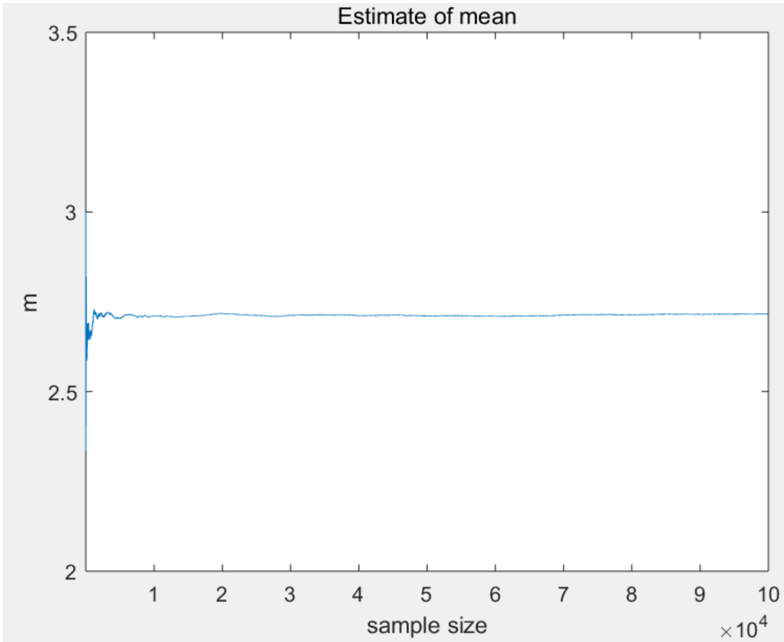


Figure 3.4 Estimate of mean in different sample size

From the figure3.4, when the experiment time start from 1 to 100000, we can see that the larger the sample size is, the closer to e the value of mean is. Also, we can see the value of estimate is floating up and down around e, and the range of the floating is becoming smaller and smaller. Therefore, we can predict that when the sample size is infinite, the estimate will be the true value, e.

When sample size is 100000, the frequency distribution of N is presented below,

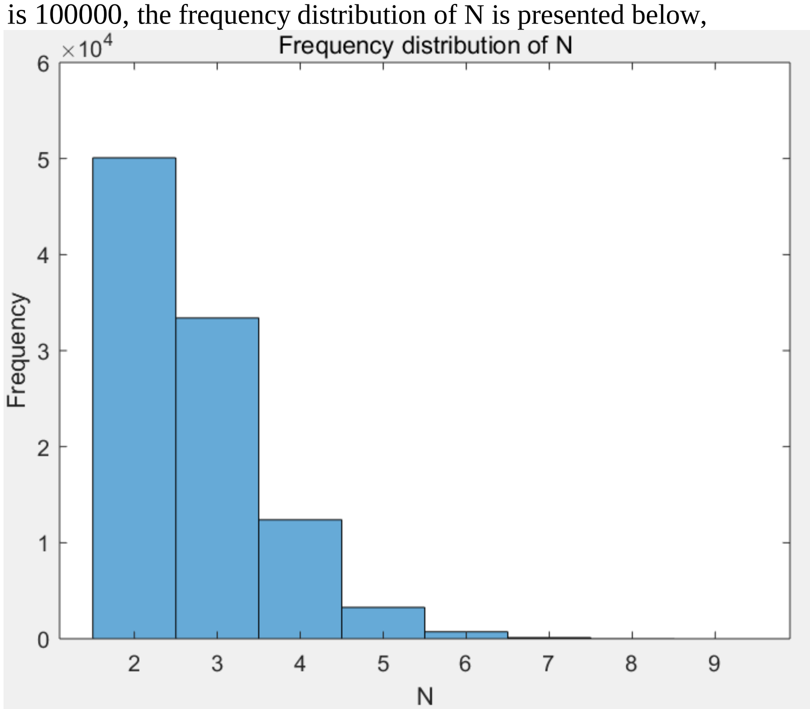


Figure 3.5 Frequency distribution of N

```
ans =

    2.7183
```

Figure 3.6 Mean value by simulation(100000 times)

In figure 3.6, the simulation of the mean value is 2.7183, which is close to the true value of $E[N]=e\approx 2.718$.

5. Source Code

The code for Problem B:

```
N=zeros(1,100000);
m=zeros(1,100000);
for i=1:100000
    j=1;
    s1=rand();
    s2=rand();
    while s2 >= s1
        s1=s2;
        s2=rand();
        j=j+1;
    end
    N(i)=j+1;
    m(i)=mean(N(1:i));
end
m(100000)
figure(1);
plot(m(1,1:100000))
axis([1 100000 2 3.5]);
title('Estimate of mean');
ylabel('m');
xlabel('sample size');
figure(2);
histogram(N)
title('Frequency distribution of N');
ylabel('Frequency');
xlabel('N');
```

Problem C

1. Problem Statement

Maxima of Uniform RV's

Consider the 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 $P(X_2 = k)$? (Hint: condition on U_1 and then uncondition.) What does this say about $E[X_2]$?

2. Theoretical Analysis

Assume we generate k random numbers { $U_1, U_2 \cdots U_k$}, which are uniform random variables in (0,1). there are $k!$ results. If we want to meet the condition that U_k is the biggest number and U_1 is the second biggest number, there are overall $(k-2)!$ results. It means that:

$$\mathrm{P}[X_2 = (k-1)] = \frac{(k-2)!}{(k)!}$$

$$\mathrm{P}[X_2 = k] = \frac{(k-1)!}{(k+1)!}$$

When $k \to \infty$:

$$\mathrm{E}[X_2] = \sum_{n=1}^{\infty} \frac{(k-1)!}{(k+1)!} \times k = \sum_{k=1}^{\infty} \frac{1}{k+1} = \sum_{k=0}^{\infty} \frac{1}{k+2}$$

The function $\sum_{k=0}^{\infty} \frac{1}{k+2}$ isn't a convergent function, so $\mathrm{E}[X_2]$ isn't a constant.

3. Simulation Methodology

In this problem, I generate a variable U_1which is the first maximum. Then I continuously generate variables U_i 1000 times and record the N when the second and third maximum appear. X_2 is the distance from the first record to the second record and X_3 is the distance from the second record to the third record. Repeat experiment 8 times and record X_2 and X_3.

4. Experiments and Results

I did the experiment with sample size 1000 for 8 times. In the last experiment, the sample means of X2 = 7.6850, X3 = 70.3330. The figure 3.7 and the figure 3.8 are the frequency distribution of X2 and X3.

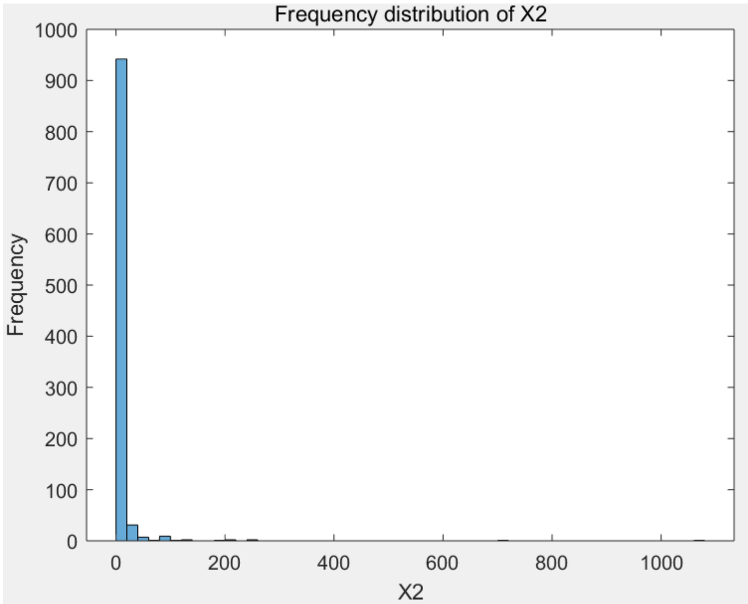


Figure 1.7 The Frequency distribution of X2

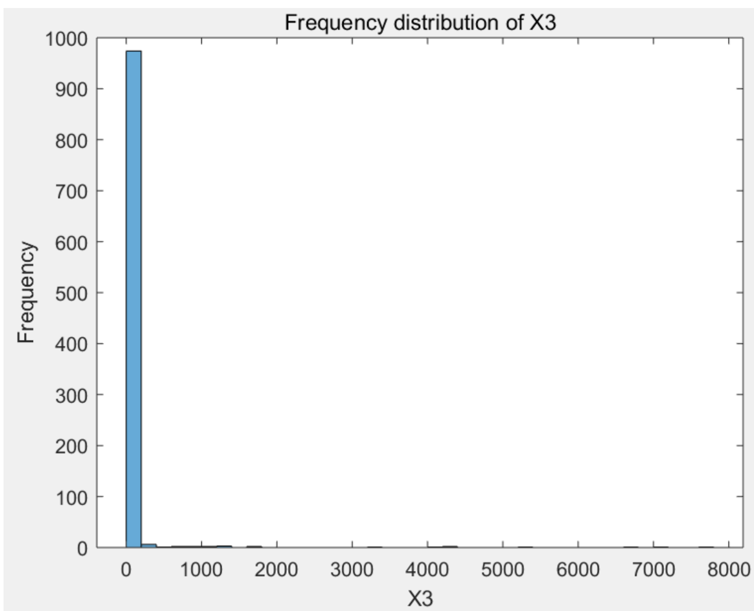


Figure 2.8 The Frequency distribution of X3

After the 8-time experiments, I get the result below:

x_2

1x8 double

1	2	3	4	5	6	7	8
6.5360	9.2820	8.6820	9.5630	8.0560	9.0150	10.2840	7.6850

Figure 3.9 X2 in 8-time experiments

x_3

1x8 double

1	2	3	4	5	6	7	8
42.8610	55.2550	46.7580	35.8900	527.9780	85.7480	34.0270	70.3330

Figure 4.10 X3 in 8-time experiments

From the figure 3.9 and the figure 3.10, we can find that X2 and X3 are not stable at a constant. They change a lot in different experiments.

5. Source Code

```
time=8;
x_2=zeros(1,8);
x_3=zeros(1,8);
for k=1:8
    N=zeros(1,1000);
    M=zeros(1,1000);
    for i=1:1000
        j=1;
        s1=rand();
        s2=rand();
        max=s1;
        while s2 <= max
            s2=rand();
            j=j+1;
        end
        N(i)=j;
        j=1;
        max=s2;
```

```
        s3=rand();
        while s3 <= max
            s3=rand();
            j=j+1;
        end
        M(i)=j;
end
x_2(k)=mean(N);
x_3(k)=mean(M);
end
figure(1);
histogram(N);
title('Frequency distribution of X2');
ylabel('Frequency');
xlabel('X2');
figure(2);
histogram(M);
title('Frequency distribution of X3');
ylabel('Frequency');
xlabel('X3');
```