Project 3 Some Interesting DRV's

EE 511 – Section Thursday 9 am

Name: Junquan Yu

Student ID #: 3372029142

1. Problem Statement

1) Sum of Uniform RV's

Define:

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

where {Ui} 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) Minima of Uniform RV's

Define:

$$N = \mathrm{Min}\{n : U_1 \le U_2 \le U_3 \le \ldots \le U_{n-1} > U_n\}$$

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

3) Maxima of Uniform RV's

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

Example: the records are underlined.

$\{Ui\} = \{0.2314, 0.4719, 0.1133, 0.5676, 0.4388, 0.9453...\}$

(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 $i\text{-}1^{th}$ 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 Exploration or Analysis

For the problem 1:

I generate a series of random numbers which obeys the uniform distribution U(0,1) until the sum of these random numbers is greater than 1. Then I record the index value of the last one, which is the minimum value of n for this trial. After that, I repeat this experiment for 10000 times and record the minimum value of n for each trial. I would end up getting 10000 values of N, which is the minimum value of n for each trial and calculate the sample mean from these values of N, which is the estimator for the mean of N.

For the problem 2:

I generate a series of random numbers which obeys the uniform distribution U(0,1) until the certain term of this sequence is less than its predecessor. Then I record the index value of this certain term, which is the minimum value of n for this trial. After that, I repeat this experiment for 10000 times and record the minimum value of n for each trial. After that, I repeat this experiment for 10000 times and record the minimum value of n for each trial. I would end up getting 10000 values of N, which is the minimum value of n for each trial and calculate the sample mean from these values of N, which is the estimator for the mean of N.

For the problem 3:

Consider the sequence of iid Uniform RV's {Ui}. If $U_j > \max_{i=1:j-1}\{U_i\}$ we say U_j is a record. According to this, I generate a series of iid random numbers which obeys the uniform distribution U(0,1) until the third record of this sequence emerges. Then I record the index value of the second and third record (the index value of the first record is always 1). After that, I calculate the difference of the index value of the first and second record, which is the value of X_2 and the difference of the index value of the second and third record, which is the X_3 (the value of X_1 is always 1). Afterwards, I repeat this experiment for 10000 times and record the value of X_2 and X_3 for each trial. I would end up getting 10000 values of X_2 and X_3 and calculate the sample means from these values of X_2 and X_3.

3. Simulation Methodology

For the problem 1:

I used the function input( ) to enable user to set the number of repetitions, which is

M, for this experiment. Then I created the “for j=1:M” loop structure in order to repeat the following experiment for M times. Every time through the “for” expression, the value of the accumulator p which is the sum of a series of random numbers generated would be initialized to ensure that the value of p is 0 at start of each loop and the index value i would also be initialized to 0. Then I used the “while (p<=1)” loop structure to judge whether the accumulator p is greater than 1. If not, the index value i would plus 1 and a new random number would be generated and assigned to the element of the array, which is a(1,i). Then the value of this random number is added to the accumulator p. This process was repeated through the “while” loop until the value of p is greater than 1 and thus jump out of the “while” loop. Then I saved the current index value, which is i, to the element of the other array, which is b(1,j), at the end of each loop. When the whole “for” loop finished, I ended up getting M index values i, which are M values of N. I can use the function mean( ) to calculate the sample mean of N, which is $\hat{m}$.

For the problem 2:

I used the function input( ) to enable user to set the number of repetitions, which is M, for this experiment. Then I created the “for j=1:M” loop structure in order to repeat the following experiment for M times. Every time through the “for” expression, the index value i was initialized to 1, and the first two random numbers of the sequence were generated and assigned to the first two elements of an array, which are a(1,1) and a(1,2). Then I used the “while (a(1,i)<=a(1,i+1))” loop structure to judge successively whether each term in the sequence is less than its predecessor. If not, the index value i plus 1 and a new random number would be generated and assigned to the element of the array, which is a(1,i+1). This process was repeated through the "while" loop until the certain term is first less than its predecessor and thus jump out of the "while" loop. Then I saved the current index value of this term, which is i+1, to the element of the other array, which is b(1,j) at the end of each loop. When the whole "for" loop finished, I ended up getting M index values i+1, which are M values of N. I can use the function mean( ) to calculate the sample mean of N, which is $\hat{m}$.

For the problem 3:

I used the function input( ) to enable user to set the number of repetitions, which is M, for this experiment. Then I created the "for j=1:M" loop structure in order to repeat the following experiment for M times. I created an array, which is b( ), to accommodate the index value of the first three records of the sequence of random numbers. Because the index value of the first record is always 1, I assigned 1 to the element of this array, which is b(1,1). Every time through the "for" expression, the index value i was initialized to 1, and the first two random numbers of the sequence were generated and assigned to the first two elements of an array, which are a(1,1) and a(1,2). Then I used the "while (a(1,i+1)<=max(a(1,1:i)))" loop structure to judge successively whether each term in the sequence is greater than the maximum value among all of its predecessors. If not, the index value i plus 1 and a new random number would be generated and assigned to the element of the array, which is a(1,i+1). This process was repeated through the "while" loop until the certain term emerged to be greater than the maximum value among all of its predecessors and thus jump out of the "while" loop. Then I saved the current index value of this term, which is index value of the second record, to the element of the array, which is b(1,2). Then I used the same way to find out the index value of the third record with the other "while" loop structure and saved it to the b(1,3). After that, I calculated the difference from the first to the second record, which is X_2, and from the second to the third record, which is X_3, and assigned them to the array element c(1,j) and d(1,j) respectively at the end of the loop. When the whole "for" loop finished, I ended up getting the array c( ) and d( ) with M elements in each of them, which are values of X_2 and X_3. I can used the function mean( ) to calculate the sample mean of X_2 and X_3 and used the function histogram( ) to plot the probability graph for them.

4. Experiments and Results

For the problem 1:

I did this experiment for 8 times in total. I set M=10000 for all of times and the results are as follows:

Experiment	1st	2nd	3rd	4th	5th	6th	7th	8th
E[N]	2.7311	2.7193	2.7163	2.7063	2.7172	2.7078	2.7163	2.7159

I guess the true value for E[N] is the natural base e.

Assuming that $U_1, U_2, \ldots, U_n$ is a series of iid random variables which obey the uniform distribution U(0,1). The cdf of the sum of these random variables is:

$$F_n(x) = P[\sum_{i=1}^{n} U_i \le x] = \frac{x^n}{n!}$$

We can prove this by the mathematical induction.

We also know that:

$$P(N > n) = F_n(x) = \frac{x^n}{n!}$$

Then we can get:

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

The true value for E[N] is:

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

For the problem 2:

I did this experiment for 8 times in total. I set M=10000 for all of times and the results are as follows:

Experiment	1st	2nd	3rd	4th	5th	6th	7th	8th
E[N]	2.7243	2.7075	2.7016	2.7170	2.7291	2.7201	2.7157	2.7092

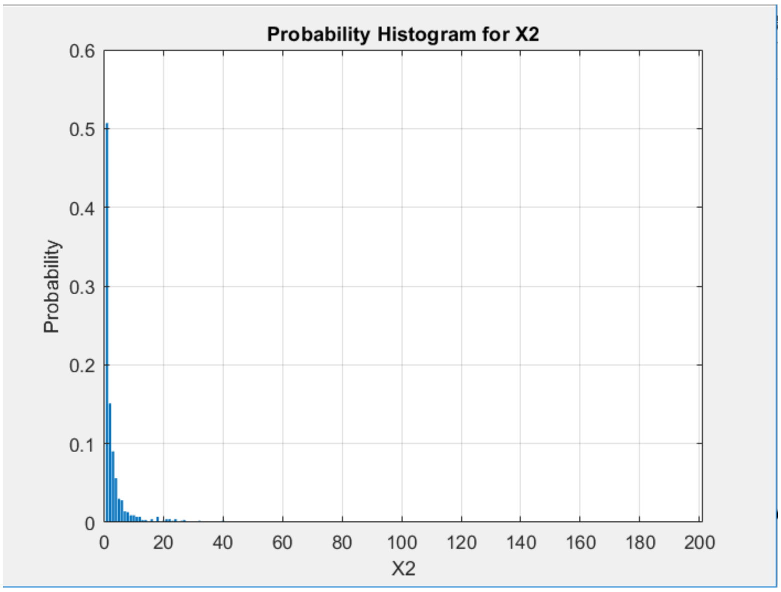
So I guess the true value for E[N] is the natural base e.

For the problem 3:

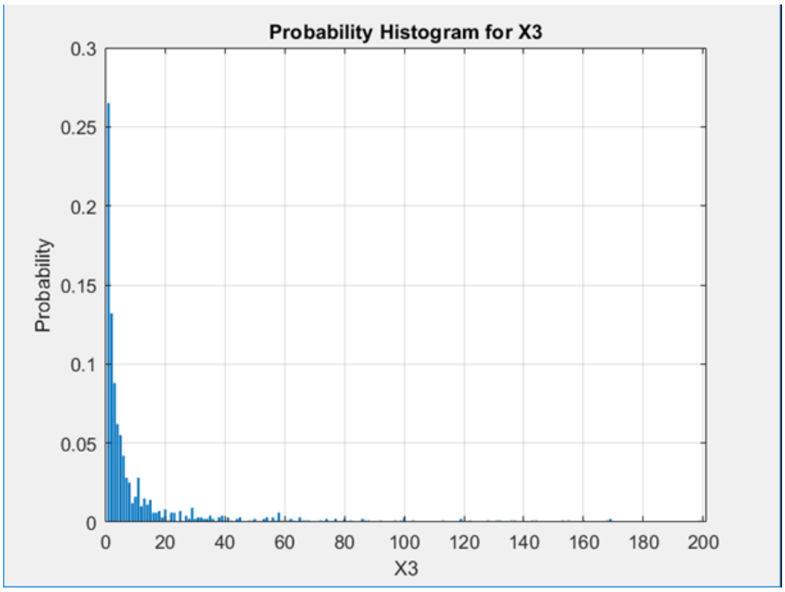
I did this experiment just for 3 times. I set M=1000 for all of times and the results are as follows:

In the first time, the sample mean of X_2 is 9.718 while the sample mean of X_3 is 50.517.

The probability histogram for X_2 is:

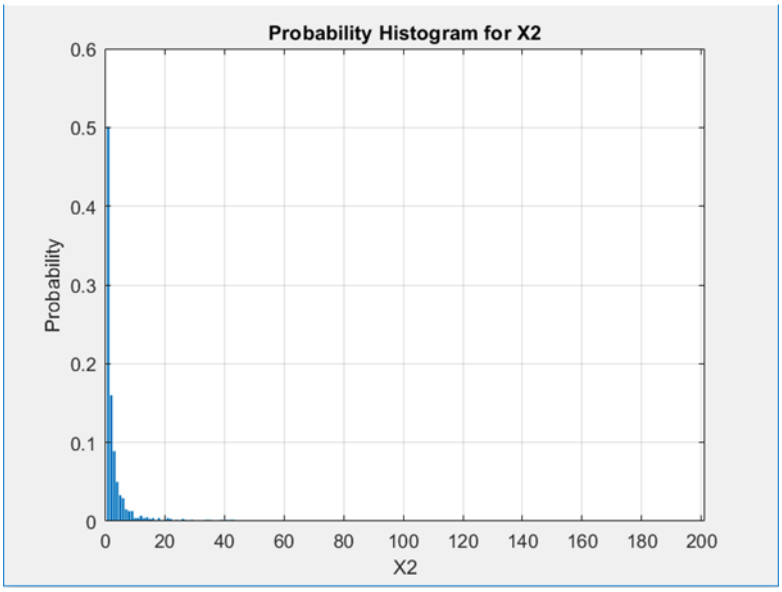


The probability histogram for X_3 is:

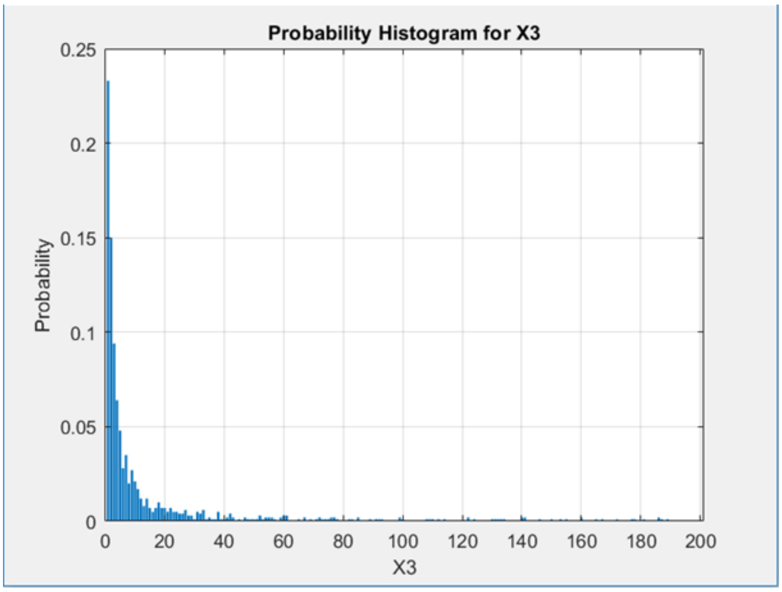


In the second time, the sample mean of X_2 is 8.164 while the sample mean of X_3 is 48.396.

The probability for histogram for X_2 is:



他



5. References

1. *Alberto Leon-Garcia. (2008). Probability, Statistics, and Random Processes for Electrical Engineering. Upper Saddle River, NJ 07458. Pearson Education, Inc.*

2. *Zhou Sheng, Shiqian Xie, Chengyi Pan. (2008). Probability Theory and Mathematical Statistics. No.4, Dewai Street, Xicheng District, Beijing. Higher Education Pres*

6. Source Code

The program for the problem 1:

```
M=input('please enter the number of repetitions: ');
for j=1:M          %repeat this experiment for M times
    i=0;
    p=0;
    while (p<=1)    %find N for this trial
        i=i+1;
        a(1,i)=rand;
        p=p+a(1,i);
    end
    b(1,j)=i;
end
u=mean(b(1,1:M));    %calculate the sample mean of N
```

the program for the problem 2:

```
M=input('please enter the number of repetitions: ');
for j=1:M                    %repeat this experiment for M times
    i=1;
    a(1,i)=rand;         %generate the first two random numbers of the sequence
    a(1,i+1)=rand;
    while (a(1,i)<=a(1,i+1))    %find N for this trial, which is the index value of term in the
sequence that is first less than its predecessor
        i=i+1;
        a(1,i+1)=rand;
    end
    b(1,j)=i+1;
end
u=mean(b(1,1:M));    %calculate the sample mean of N
```

the program for the problem 3:

```
M=input('please enter the number of repetitions: ');
for j=1:M           %repeat this experiment for M times
    b(1,1)=1;
    i=1;
    a(1,i)=rand;         %generate the first two random numbers of the sequence
    a(1,i+1)=rand;
        while (a(1,i+1)<=max(a(1,1:i)))    %find the second record of sequence for this trial
          i=i+1;
```

```
        a(1,i+1)=rand;
      end
      b(1,2)=i+1;
      i=i+1;
      a(1,i+1)=rand;
      while (a(1,i+1)<=max(a(1,1:i)))    %find the third record of sequence for this trial
        i=i+1;
        a(1,i+1)=rand;
      end
      b(1,3)=i+1;
      c(1,j)=b(1,2)-b(1,1);       %calculate the X2 for this trial
      d(1,j)=b(1,3)-b(1,2);       %calculate the X3 for this trial
end
figure              %plot the probability histogram for X2 and X3
tbl1=tabulate(c(:));
X1=tbl1(1:100,1);
P1=tbl1(1:100,3)/100;
bar(X1,P1);
grid on
ylabel('Probability')
xlabel('X2')
title('Probability Histogram for X2')
figure
tbl2=tabulate(d(:));
X2=tbl2(1:100,1);
P2=tbl2(1:100,3)/100;
bar(X2,P2);
grid on
ylabel('Probability')
xlabel('X3')
title('Probability Histogram for X3')
u1=mean(c);       %calculate the sample mean of X2 and X3
u2=mean(d);
```