Find answers to the following questions by simulation using SAS (at least 10000 simulations). You must provide a final answer once you have observed the output from SAS and give reasons for your answers. That is, if there is convergence, why and what is the final answer. However, if there is divergence, why.

1. Suppose X_i for i=1, 2, 3… has uniform (0, 1) distribution.
   A. Let M = min (n: $X_1 + X_2 + \ldots + X_n > 1$). Find expected value of M; E(M) = Mean of M.
   B. Let N = min (n+1: $X_n > X_{n+1}$). Find expected value of N; E(N) = Mean of N.

2. A one unit stick is broken randomly into two pieces.
   A. Find E (the short piece divided by the long piece).
   B. Find E (the long piece divided by the short piece).

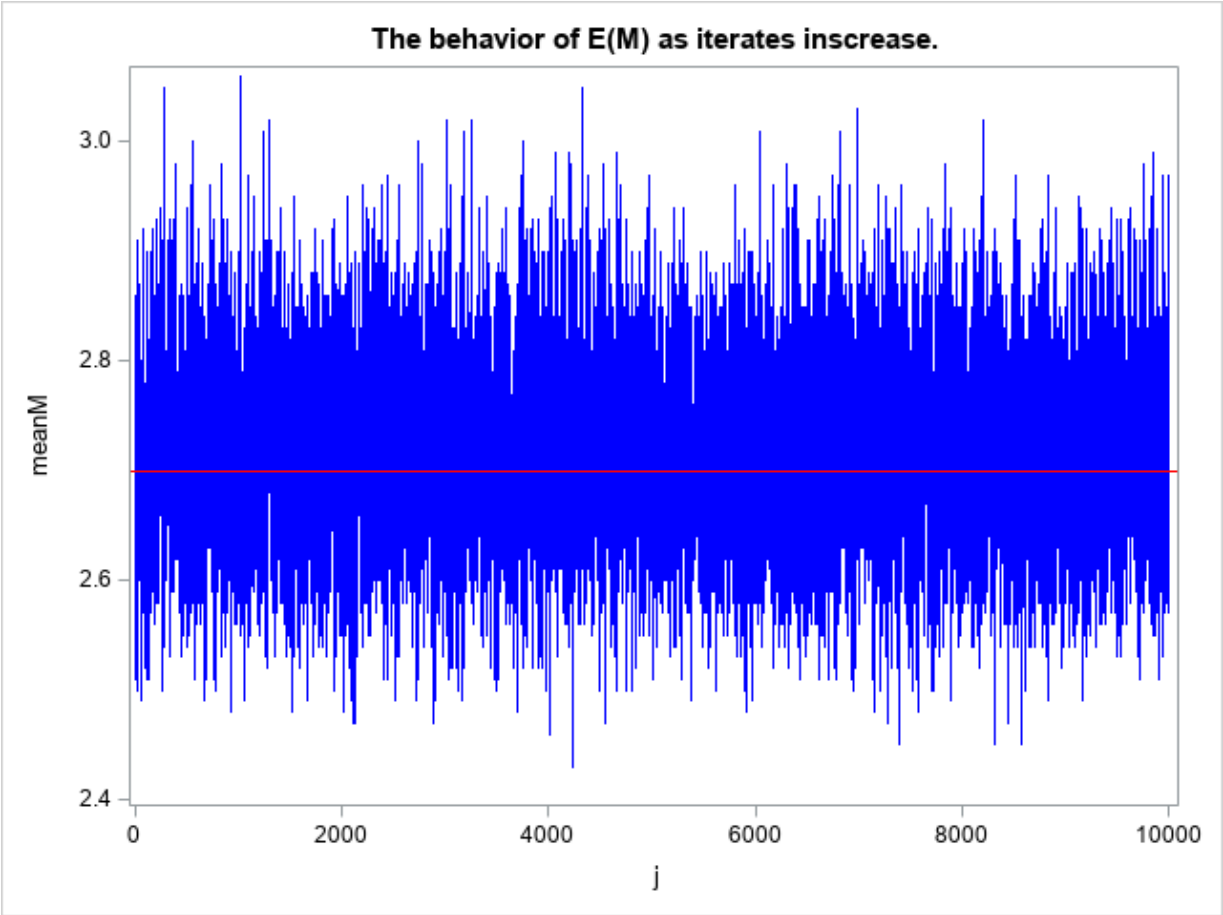
```
/***** 1A ** Nhi Vu *****/
* Let M = min {n: X1 + X2 + ... + Xn> 1}. Find E(M)=expected value of M ;
data Q1A ;
do j=1 to 10000;  *10000 simulations;
*Find expected value of M for a samples with 100 random value M;
meanM=0;
sumM=0;
do i=1 to 100;
*Set all initial value at n=0;
      sum=0;
      M=0;
   do until (sum>1) ;
      x = rand('uniform');
      sum+x;
      M+1 ;
   end ;
   sumM=sumM+M;
end;
meanM=sumM/100;
output;
end;
run;

proc means data=Q1A;
title "The expected value of M";
var meanM;
run;

ods graphics on;
proc sgplot data=Q1A;
series  x = j y = meanM / lineattrs=(color=blue);
refline 2.7 / axis= y lineattrs=(color=red);
title "The behavior of E(M) as iterates inscrease.";
run;
ods graphics off;
quit;
```

The expected value of M 09:14 Tuesday, April 6, 2021 1

The MEANS Procedure

Analysis Variable: meanM

N	Mean	Std Dev	Minimum	Maximum
10000	2.7193720	0.0881743	2.4300000	3.0600000



Even though the graph of E(M) does not show a tendency, it bounces around a value of mean of E(M), approximately 2.7193720. Besides, the summary statistic table shows a distance between E(M) of samples and their mean is close, the value of Std Dev is very small so that we can conclude that the values in a statistical data set are closely cluster to the mean of the data set, the distance between min value(2.43), max values (3.06) and mean are not far. Therefore, the data of ecpetecd value of M is convergence and my final answer that the expected value of M is approximately 2.7193720.

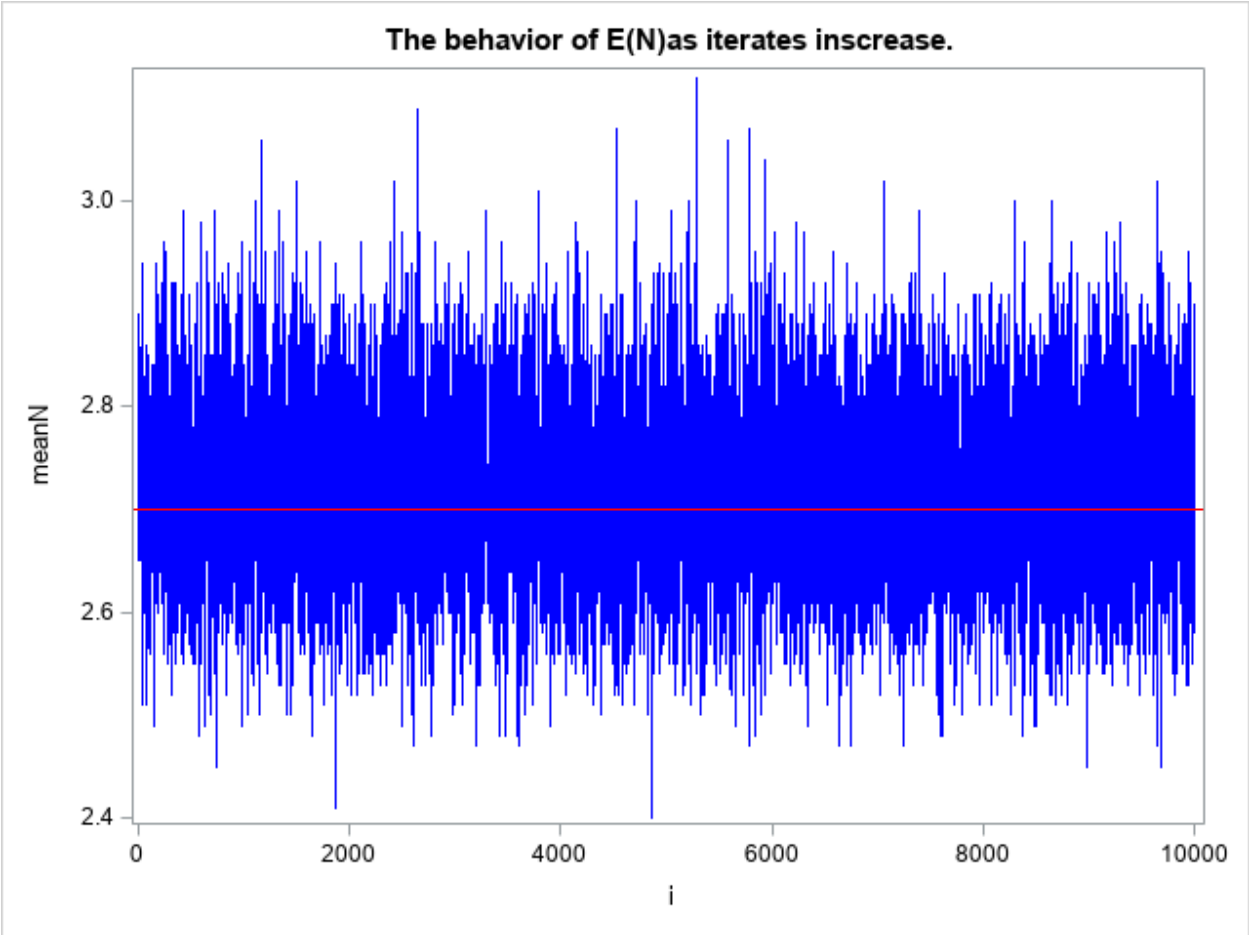
```
/***** 1B ** Nhi Vu *****/
* Let N = min (n+1: Xn > Xn+1). Find expected value of  N, E(N) = Mean of N;
data Q1B;
do i=1 to 10000;  * 10000 simulations;
*Find expected value of M for a samples with 100 random value N;
sumN=0;
do j=1 to 100;
*Set all initial value at n=0;
   N=1;
   x = rand('uniform');
   prevX = x;
   do until (prevX > x) ;
      prevX = x;            * remember last x ;
      x = rand('uniform'); * generate new x ;
      N+1 ;
   end;
   sumN=sumN+N;
end;
meanN=sumN/100;
output;
end;
run;

proc means data=Q1B;
title "The expected value of N";
var meanN;
run;

ods graphics on;
proc sgplot data=Q1B;
series  x = i y = meanN / lineattrs=(color=blue);
refline 2.7 / axis= y lineattrs=(color=red);
title "The behavior of E(N)as iterates inscrease. ";
run;
ods graphics off;
quit;
```

The expected value of N 09:14 Tuesday, April 6, 2021 2

The MEANS Procedure

Analysis Variable : meanN

N	Mean	Std Dev	Minimum	Maximum
10000	2.7197910	0.0875835	2.4000000	3.1200000

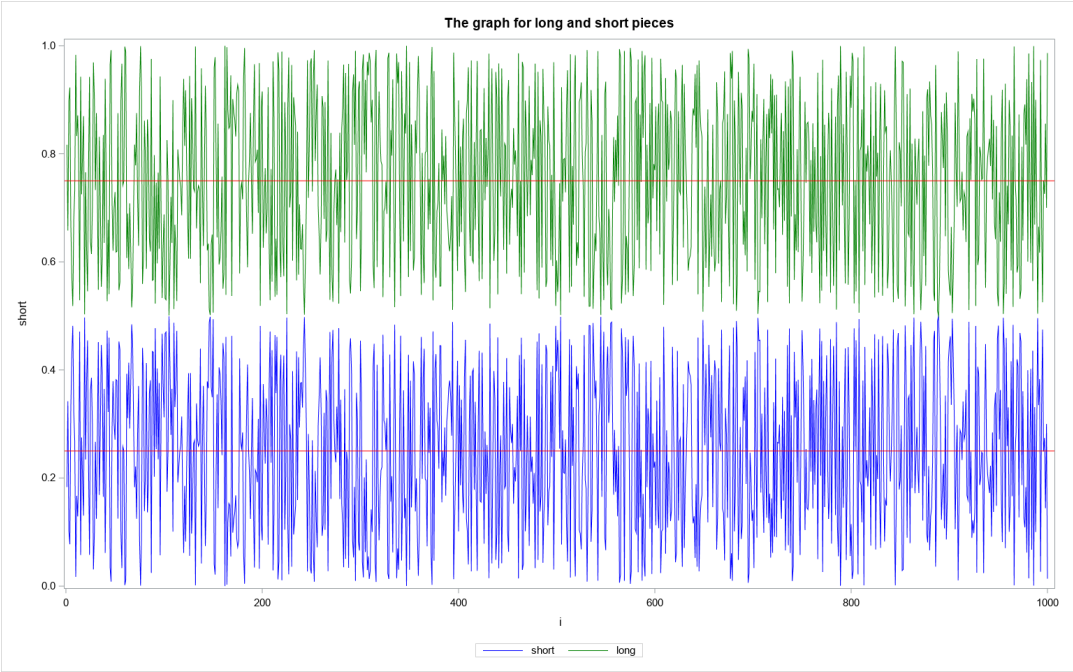


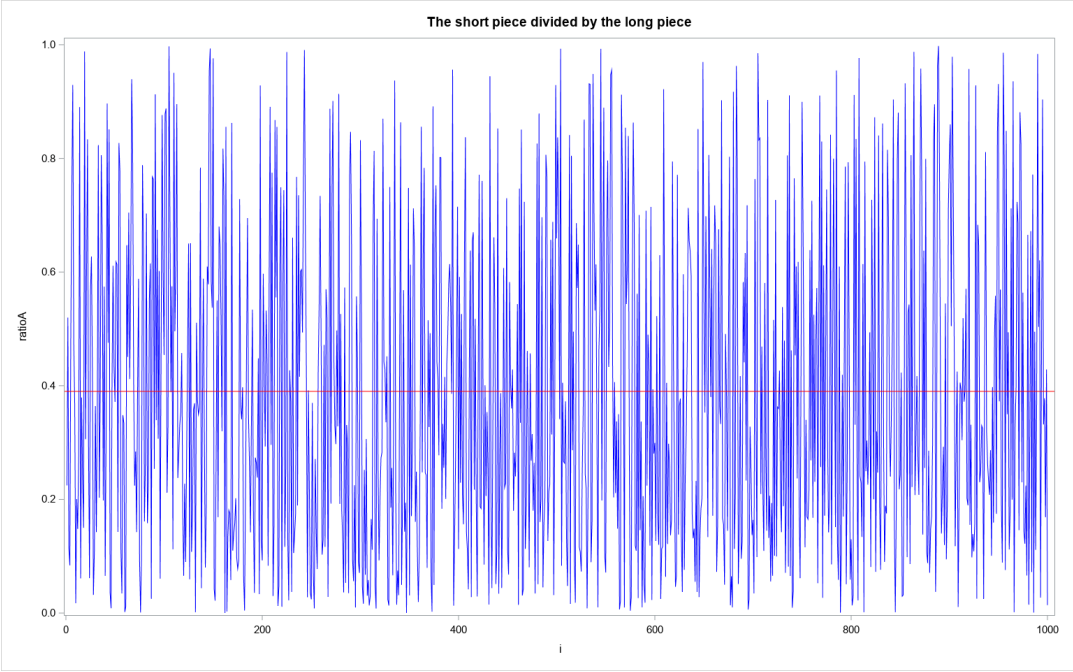
Similarly, to question 1A, even though the graph of E(N) does not show a tendency as iterates i increase, it bounces around a value of mean of E(N), approximately 2.7197910. Besides, the summary statistic table shows a distance between E(N) of samples and their mean is close, the value of Std Dev is very small so that we can conclude that the values in a statistical data set are closely cluster to the mean of the data set, the distance between min value(2.4), max values (3.12) and mean are not far. Therefore, the data of ecpetecd value of N is convergence and my final answer that the expected value of N is approximately 2.7197910.

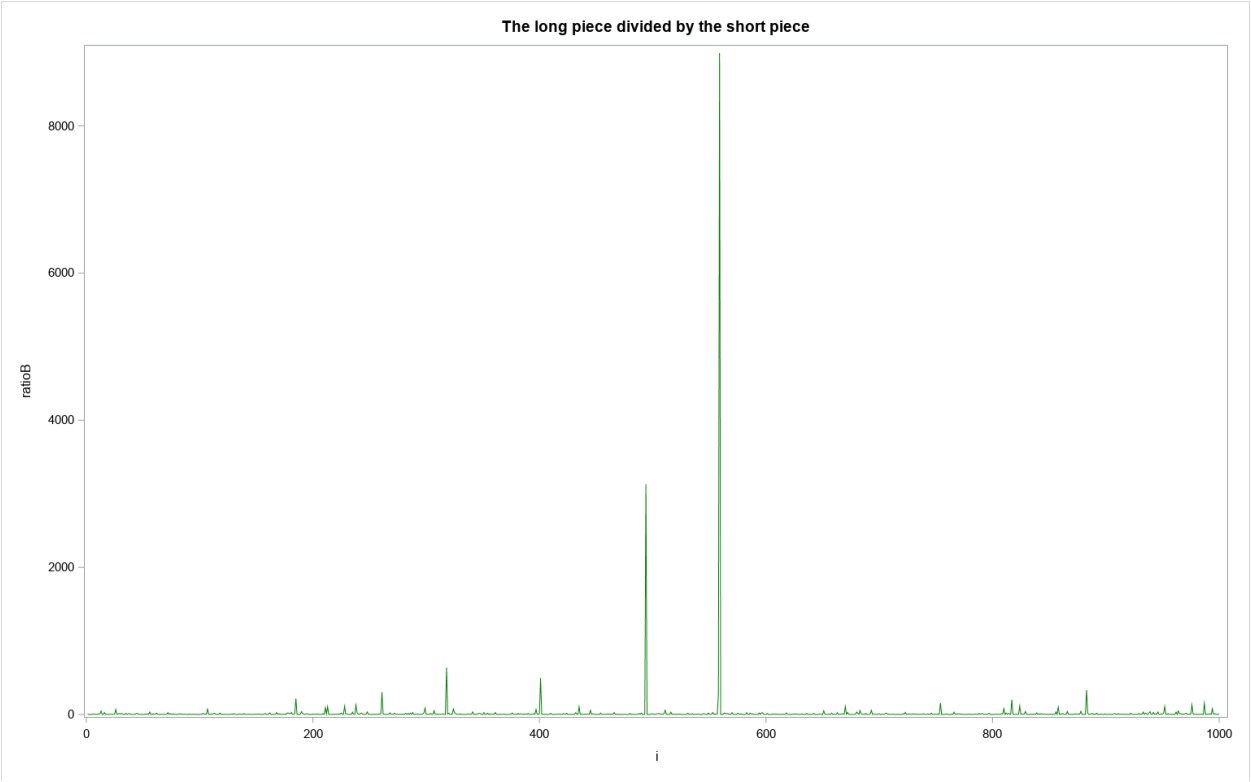
```
/***** Q2 ** Nhi Vu *****/
data Q2;
   do i= 1 to 1000;
      x= rand('uniform');    * break randomly at any point on one unit stick;
      if x>0.5 then long=x; * find a long piece;
      else long=1-x;
      short=1-long;            * find a short piece;
      ratioA=short/long;       * for question 2a;
      ratioB=1/ratioA;         * for question 2b;
      output;
   end;
run;

proc means data=Q2;
title "The expected values for question 2";
var short long ratioA ratioB;
run;

ods graphics on;
proc sgplot data=Q2;
series  x = i y = short / lineattrs=(color=blue);
refline 0.25 / axis= y lineattrs=(color=red);
series  x=i y=long / lineattrs=(color=green);
refline 0.75 / axis= y lineattrs=(color=red);
series  x=i y=ratioA / lineattrs=(color=green);
run;

proc sgplot data=Q2;
series  x = i y = ratioA / lineattrs=(color=blue);
refline 0.39 / axis= y lineattrs=(color=red);
title "The short piece divided by the long piece ";
run;

proc sgplot data=Q2;
series  x=i y=ratioB / lineattrs=(color=green);
title "The long piece divided by the short piece ";
run;
quit;
```


The MEANS Procedure

Variable	N	Mean	Std Dev	Minimum	Maximum
short	1000	0.2496936	0.1431633	0.000324537	0.4994578
long	1000	0.7503064	0.1431633	0.5005422	0.9996755
ratioA	1000	0.3847069	0.2767202	0.000324642	0.9978334
ratioB	1000	13.6175732	102.0854298	1.0021713	3080.31


The graph for long and short pieces
short
1.0
0.8
0.6
0.4
0.2
0.0
0
200
400
600
800
1000
i
short
long


The short piece divided by the long piece
ratioA
1.0
0.8
0.6
0.4
0.2
0.0
0
200
400
600
800
1000
i



By 10000 simulations, I found each data sets of short piece, of long piece, of the short divide by the long piece and of the long divide by short piece. Each data sets, I calculate the mean and plot them which x-axis is iteration.

For data set of short pieces with 10000 observations, I found their mean is approximately 0.25, also other summary numbers are statistically significant enough to conclude that data converge to their mean. The graph of short piece bounces around their mean, so I conclude that the short piecce is approximately 0.25.

Similarly, the distribution in long piece data show convergence to their mean, around 0.75. Thus, I conclude that the long piece is around 0.75.

However, 2 ratio data sets do not share the same pattern. I found the convergence in the data of “the short divide by the long piece” by looking at its summary statistics and its graph which goes around its mean, while the tendency in data of “the long piece divided by the short piece” shows divergence, datas spread out of their mean, its Std Dev is very high, which indicates that the data points are far from the mean, so we cannot use their mean to represent all data. The distance between min, max and mean value are very far, the graph also does not show tendency of data.

Thus, my final answer is that the data of the short divide by the long piece converge to their mean, which is around 0.3847069 but the data of the long pecie divied by short one is divergence even though the data of long peice and short piece converge to their own means.