Weining Bao weiningb@usc.edu 8287723653

Assignment 4

1. Sum of Uniform RVs:

Define: $N = \min\{n : \sum_{i=1}^{n} U_i > 1\}$ where $U_i s$ are IID uniform (0,1) RVs.

Find by simulation: $\hat{m} = E[N]$ as an estimator for the mean. Derive the true value of $E[N]$.

2. Minima of Uniform RVs:

Define: $N = \min\{n : U_1 \le U_2 \le ... \le U_{n-1} > U_n\}$ where $U_i s$ are IID uniform (0,1) RVs.

Find by simulation: $\hat{m} = E[N]$ as an estimator for the mean. Derive the true value of $E[N]$.

3. Maxima of Uniform RVs:

Consider the sequence of IID Uniform RVs $U_i s$. If $U_j > \max_{i=1:j-1}\{U_i\}$ then it is a record.
Let X_i be a RV for the distance from the i-1st record to the ith record. Clearly $X_1 = 1$ always.

Find by simulation: $\hat{m}_j$ as an estimator for $E[X_j]$ for j = 1…6.

Matlab Simulation

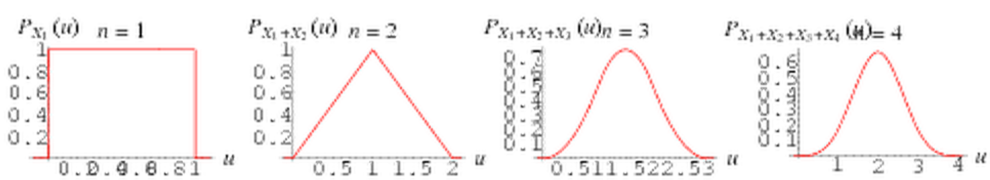
with code at last

1. Sum of Uniform RVs:

Define: $N = \min\{n : \sum_{i=1}^{n} U_i > 1\}$ where $U_i s$ are IID uniform (0,1) RVs.

Find by simulation: $\hat{m} = E[N]$ as an estimator for the mean. Derive the true value of $E[N]$.

Referenced on WolframlAlpha: Uniform Sum Distribution[1]



The distribution for the sum $X_1 + X_2 + \ldots + X_n$ of n uniform variates on the interval $[0, 1]$ can be found directly as

$$P_{X_1+\ldots+X_n}(u) = \underbrace{\int\int\cdots\int}_{n} \delta(x_1 + x_2 + \ldots + x_n - u)\, dx_1\, dx_2 \cdots dx_n,$$

where $\delta(x)$ is a delta function.

[1] http://mathworld.wolfram.com/UniformSumDistribution.html

A more elegant approach uses the characteristic function to obtain

$$P_{X_1+\ldots+X_n}(u) = \mathcal{F}_t^{-1}\left[\left(\frac{i\left(1-e^{i\,t}\right)}{t}\right)^n\right](u) \tag{2}$$
$$= \frac{1}{2\,(n-1)!}\sum_{k=0}^{n}(-1)^k\binom{n}{k}(u-k)^{n-1}\,\mathrm{sgn}\,(u-k),$$

where the Fourier parameters are taken as $(1,\,1)$. The first few values of $P_n\,(u)$ are then given by

$$P_{X_1}\,(u) = \tfrac{1}{2}\,[\mathrm{sgn}\,(1-u)+\mathrm{sgn}\,u] \tag{3}$$
$$P_{X_1+X_2}\,(u) = \tfrac{1}{2}\,[(-2+u)\,\mathrm{sgn}\,(-2+u)-2\,(-1+u)\,\mathrm{sgn}\,(-1+u)+u\,\mathrm{sgn}\,u] \tag{4}$$
$$P_{X_1+X_2+X_3}\,(u) = \tfrac{1}{4}\left[-(-3+u)^2\,\mathrm{sgn}\,(-3+u)+3\,(-2+u)^2\,\mathrm{sgn}\,(-2+u)-3\,(-1+u)^2\,\mathrm{sgn}\,(-1+u)+u^2\,\mathrm{sgn}\,u\right] \tag{5}$$
$$P_{X_1+X_2+X_3+X_4}\,(u) = \tfrac{1}{12}\left[(-4+u)^3\,\mathrm{sgn}\,(-4+u)-4\,(-3+u)^3\,\mathrm{sgn}\,(-3+u)+6\,(-2+u)^3\,\mathrm{sgn}\,(-2+u)-4\,(-1+u)^3\,\mathrm{sgn}\,(-1+u)+u^3\,\mathrm{sgn}\,u\right], \tag{6}$$

illustrated above.

Interestingly, the expected number of picks n of a number x_k from a uniform distribution on $[0,\,1]$ so that the sum $\sum_{k=1}^{n} x_k$ exceeds 1 is e (Derbyshire 2004, pp. 366-367). This can be demonstrated by noting that the probability of the sum of n variates being greater than 1 while the sum of $n-1$ variates being less than 1 is

$$P_n^{(1)} = \int_1^n P_{X_1+\ldots+X_n}\,(u)\,d\,u - \int_1^{n-1} P_{X_1+\ldots+X_{n-1}}\,(u)\,d\,u \tag{7}$$
$$= \left(1-\frac{1}{n!}\right)-\left[1-\frac{1}{(n-1)!}\right] \tag{8}$$
$$= \frac{1}{n\,(n-2)!}. \tag{9}$$

The values for $n = 1,\,2,\,\ldots$ are 0, 1/2, 1/3, 1/8, 1/30, 1/144, 1/840, 1/5760, 1/45360, ... (OEIS A001048). The expected number of picks needed to first exceed 1 is then simply

$$\langle n_1\rangle = \sum_{n=1}^{\infty} n\,P_n^{(1)} = \sum_{n=1}^{\infty}\frac{1}{(n-2)!} = \sum_{n=0}^{\infty}\frac{1}{n!} = e. \tag{10}$$

So the probability mass function(PMF) of N is $P(n) = \dfrac{1}{n(n-2)!}\quad n = 2,3\ldots$

and the true value of $E[N]$ is $e \approx 2.7182818$.

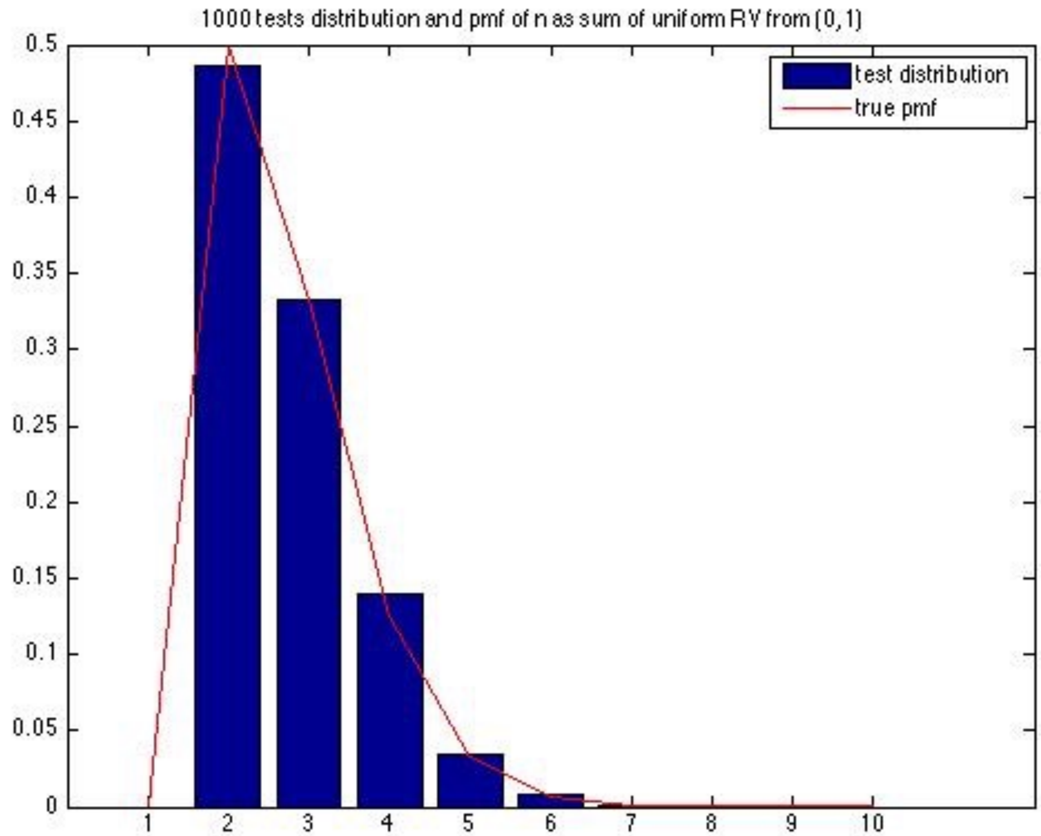
Here are the results and the PMF distributions from 1000, 10000 and 100000 times tests:

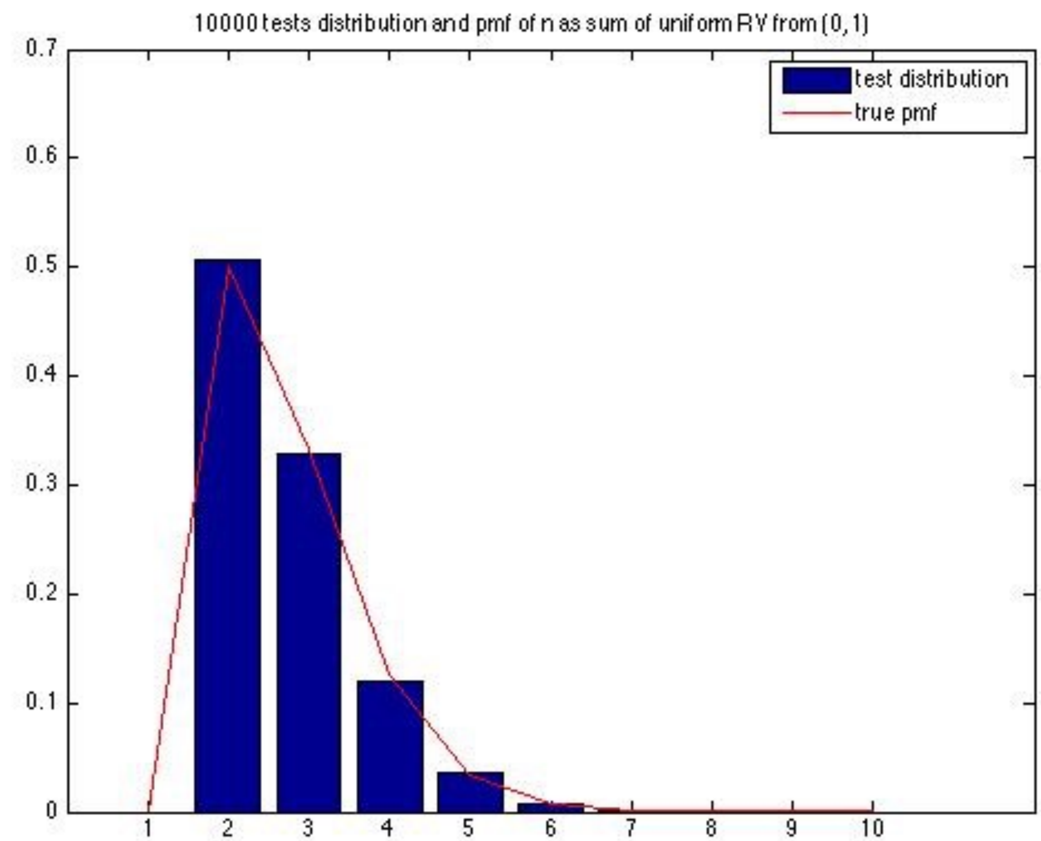
Prob 1:

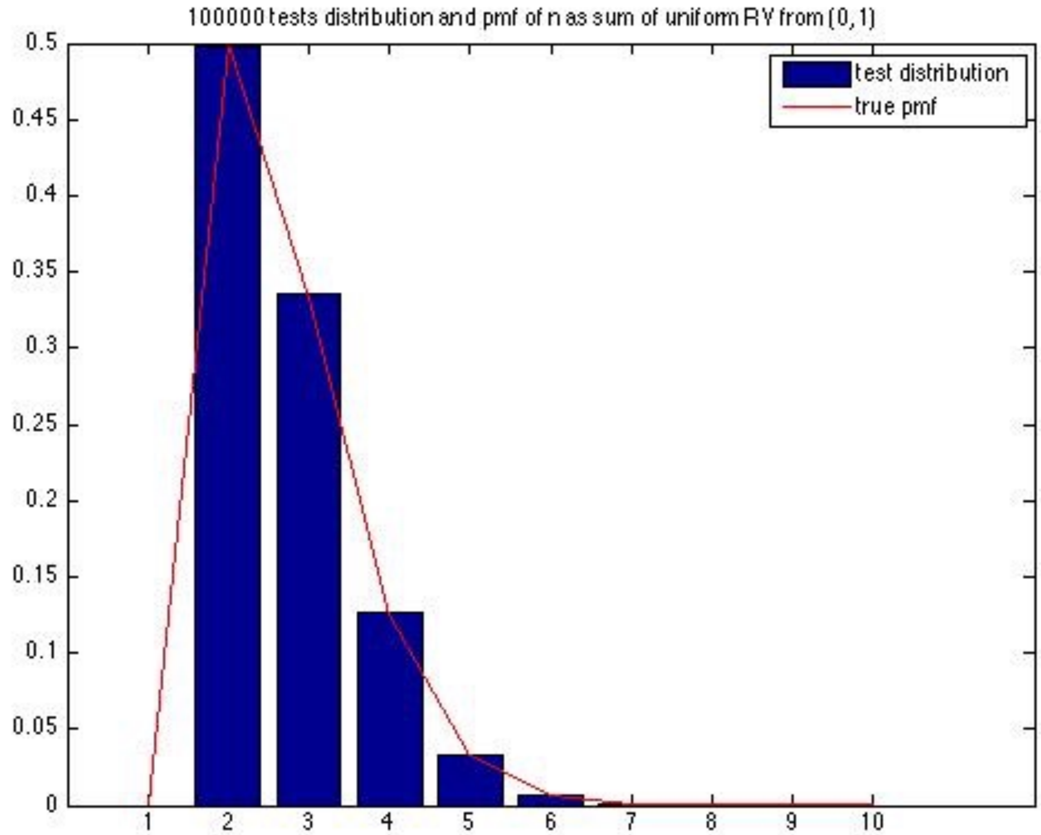
Estimating 1000 times for n:
Mean: 2.749000. Variance: 0.786786.

Estimating 10000 times for n:
Mean: 2.711800. Variance: 0.773218.

Estimating 100000 times for n:
Mean: 2.720090. Variance: 0.762348.


1000 tests distribution and pmf of n as sum of uniform RV from (0,1)
test distribution
true pmf
0.5
0.45
0.4
0.35
0.3
0.25
0.2
0.15
0.1
0.05
0
1
2
3
4
5
6
7
8
9
10


10000 tests distribution and pmf of n as sum of uniform RV from (0,1)
test distribution
true pmf
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
1
2
3
4
5
6
7
8
9
10



The results get close to the true value e with the increase of tests number. Test distributions fit the theoretical probability mass function.

2. Minima of Uniform RVs:

Define: $N = \min\{n : U_1 \le U_2 \le ... \le U_{n-1} > U_n\}$ where $U_i s$ are IID uniform (0,1) RVs.

Find by simulation: $\hat{m} = E[N]$ as an estimator for the mean. Derive the true value of $E[N]$.

The PMF of n can be calculated in following way: it is impossible for n = 1;

if n = 2, then $U_2 < U_1$. The possibility is $\int_{u_1=0}^{1}\int_{u_2=0}^{u_1} du_2\, du_1 = \frac{1}{2}$;

if n = 3, the possibility is $\int_{u_1=0}^{1}\int_{u_2=u_1}^{1}\int_{u_3=0}^{u_2} du_3\, du_2\, du_1 = \frac{1}{3}$;

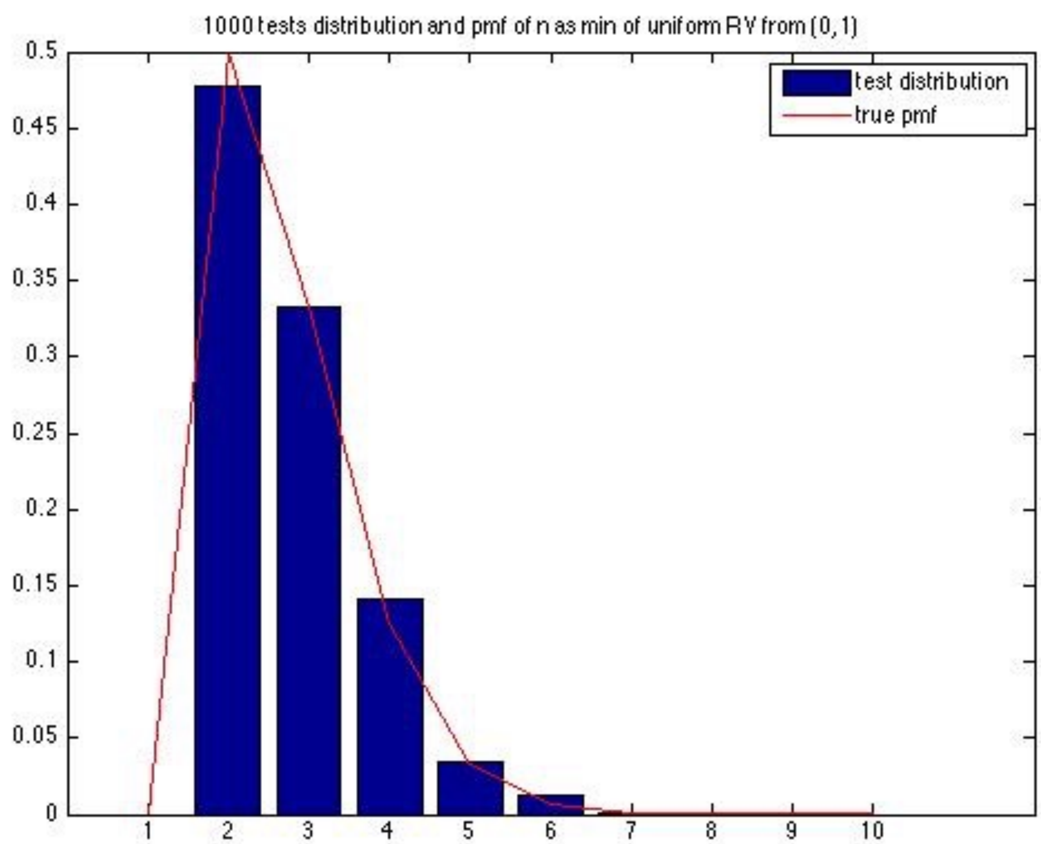
n = 4, the possibility is $\int_{u_1=0}^{1}\int_{u_2=u_1}^{1}\int_{u_3=u_2}^{1}\int_{u_4=0}^{u_3} du_4\, du_3\, du_2\, du_1 = \frac{1}{8}$ … It can be found that the probability mass function(PMF) of N is also $P(n) = \dfrac{1}{n(n-2)!}$ $n = 2,3...$ and the true value of $E[N]$ is $e \approx 2.7182818$. It is the same to the result of problem 1. Here are the results and PMF graphs:

Prob 2:

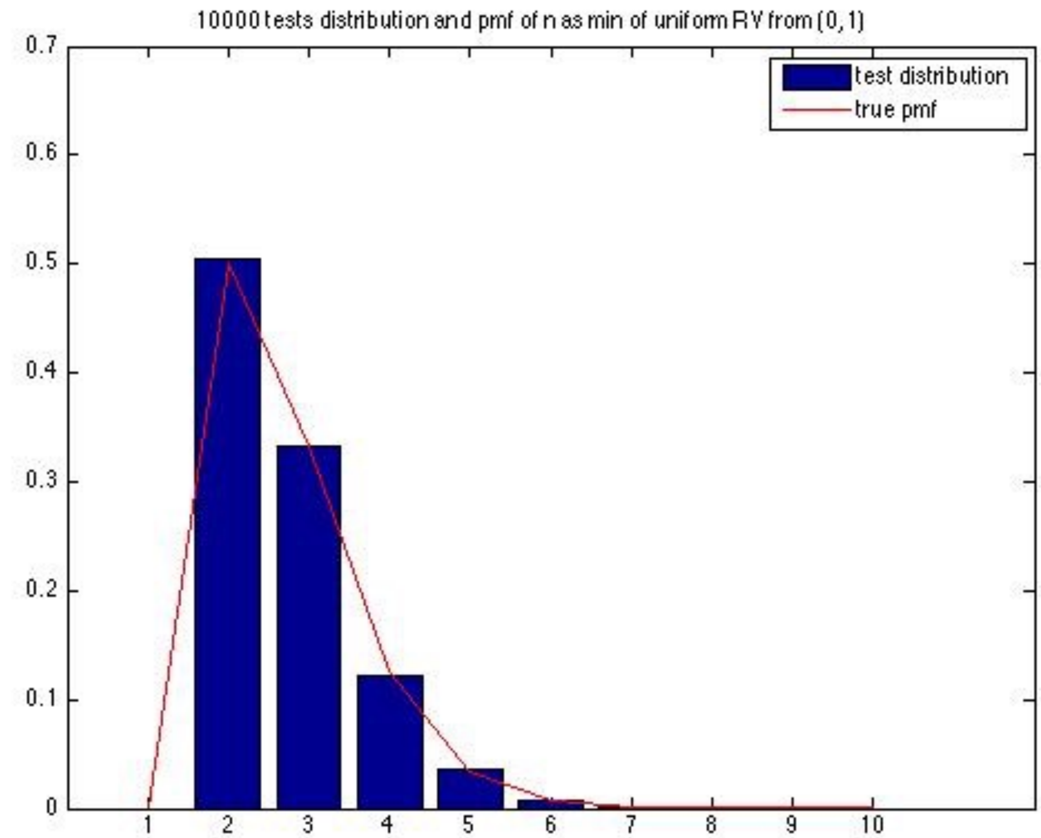
Estimating 1000 times for n:
Mean: 2.780000. Variance: 0.864464.

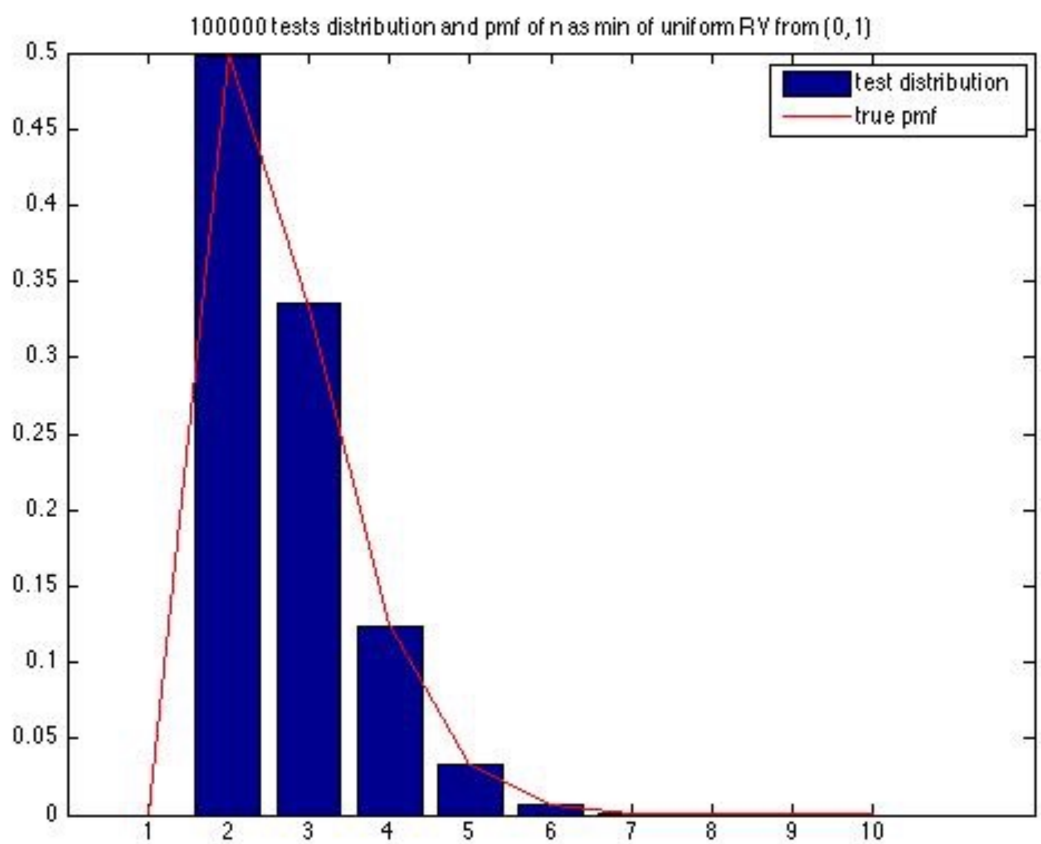
Estimating 10000 times for n:
Mean: 2.712400. Variance: 0.761562.

Estimating 100000 times for n:
Mean: 2.719730. Variance: 0.766126.



Similar to the results from problem 1, these results get close to the true value e with the increase of tests number. Test distributions fit the theoretical probability mass function.


10000 tests distribution and pmf of n as min of uniform RV from (0,1)
test distribution
true pmf
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
1
2
3
4
5
6
7
8
9
10


100000 tests distribution and pmf of n as min of uniform RV from (0,1)
test distribution
true pmf
0.5
0.45
0.4
0.35
0.3
0.25
0.2
0.15
0.1
0.05
0
1
2
3
4
5
6
7
8
9
10

3. Maxima of Uniform RVs:
Consider the sequence of IID Uniform RVs $U_i s$. If $U_j > \max_{i=1:j-1}\{U_i\}$ then it is a record.
Let X_i be a RV for the distance from the i-1st record to the ith record. Clearly $X_1 = 1$ always.
Find by simulation: $\hat{m}_j$ as an estimator for $E[X_j]$ for j = 1…6.

No theoretical analysis is done. Only simulation is done with 1000 and 10000 times of sequence generation. Results are following and have big difference and large variance in the distribution:

Prob 3:

Estimating 1000 times for x(1-6):
Mean of x1: 1.000000. Variance of x1: 0.000000.
Mean of x2: 10.176000. Variance of x2: 15587.056080.
Mean of x3: 81.263000. Variance of x3: 864160.770602.
Mean of x4: 483.804000. Variance of x4: 24308295.603187.
Mean of x5: 3013.998000. Variance of x5: 2187337126.112109.
Mean of x6: 14477.350000. Variance of x6: 76179836483.244781.

Estimating 10000 times for x(1-6):
Mean of x1: 1.000000. Variance of x1: 0.000000.
Mean of x2: 20.847600. Variance of x2: 1254945.518326.
Mean of x3: 83.193000. Variance of x3: 2159762.816232.
Mean of x4: 862.635900. Variance of x4: 1710259213.469481.
Mean of x5: 33029.831400. Variance of x5: 7499714255287.833984.
Mean of x6: 44357.669600. Variance of x6: 10828068417298.447266.

With hope to approach the true value of mean with 100000 times sequence generation, the simulation can not be finished even after 10 hours. So only these two times simulations are used for estimate. It can be seen that two estimates have big difference which implies that they may not be a good estimate for $\hat{m}_j = E[X_j]$ for j = 1…6.

Code:

```
% prob 1
fprintf('\nProb 1: \n\n');
p = zeros(10, 1); % generate the pmf of n = 1...10
for i = 2:10
   p(i) = 1 / (i*factorial(i-2));
end
for count = 3:5 % test 1000, 10000, 100000 times
   num = 10^count; % number of tests
   ns = zeros(num, 1);
   fprintf('Estimating %d times for n:\n', num);
   for i = 1:num
      total = 0;
      n = 0;
      while total <= 1 % generate random number from (0,1) till sum > 1
```

```
        total = total + rand();
        n = n + 1;
      end
      ns(i) = n;
   end
   fprintf('Mean: %f. Variance: %f.\n\n', mean(ns), var(ns));
   d = zeros(10, 1); % generate the test distribution of n = 1...10
   for i = 2:10
      d(i) = sum(ns == i) / num;
   end
   figure;
   bar(1:10, d); hold on;
   plot(1:10, p, 'r'); hold off;
   title([num2str(num), ...
   ' tests distribution and pmf of n as sum of uniform RV from (0,1)']);
   legend('test distribution', 'true pmf');
end

% prob 2
fprintf('\nProb 2: \n\n');
for count = 3:5 % test 1000, 10000, 100000 times
   num = 10^count; % number of tests
   ns = zeros(num, 1);
   fprintf('Estimating %d times for n:\n', num);
   for i = 1:num
      last = -1; % last random number from (0,1)
      curr = rand(); % generate a new random number
      n = 1;
      while curr >= last % generate random number from (0,1) till curr < last
         last = curr;
         n = n + 1;
         curr = rand();
      end
      ns(i) = n;
   end
   fprintf('Mean: %f. Variance: %f.\n\n', mean(ns), var(ns));
   d = zeros(10, 1); % generate the test distribution of n = 1...10
   for i = 2:10
      d(i) = sum(ns == i) / num;
   end
   figure;
   bar(1:10, d); hold on;
   plot(1:10, p, 'r'); hold off;
   title([num2str(num), ...
   ' tests distribution and pmf of n as min of uniform RV from (0,1)']);
   legend('test distribution', 'true pmf');
end
```

```
% prob 3
fprintf('\nProb 3: \n\n');
for count = 3:4 % test 1000, 10000 times
    num = 10^count; % number of tests
    x = zeros(num, 6);
    fprintf('Estimating %d times for x(1-6):\n', num);
    for j = 1:num
        max = -1; % curr max random number from (0,1)
        mi = 0; % index of curr max random number
        r = 1; % curr record number
        i = 0; % curr index
        while r < 7 % generate 6 records
            i = i + 1;
            curr = rand();
            if curr > max
                x(j, r) = i - mi;
                mi = i;
                max = curr;
                r = r + 1;
            end
        end
    end
    fprintf('Mean of x1: %f. Variance of x1: %f.\n', mean(x(:,1)), var(x(:,1)));
    fprintf('Mean of x2: %f. Variance of x2: %f.\n', mean(x(:,2)), var(x(:,2)));
    fprintf('Mean of x3: %f. Variance of x3: %f.\n', mean(x(:,3)), var(x(:,3)));
    fprintf('Mean of x4: %f. Variance of x4: %f.\n', mean(x(:,4)), var(x(:,4)));
    fprintf('Mean of x5: %f. Variance of x5: %f.\n', mean(x(:,5)), var(x(:,5)));
    fprintf('Mean of x6: %f. Variance of x6: %f.\n\n', mean(x(:,6)), var(x(:,6)));
end
```