Answer of Assignment 2 Monte Carlo Method

Name: Jian Wang FSUID: JW09R

1. Compute $D_N = \text{Max}(D^+, D^-)$ for the data set $x_1 = 0.2, x_2 = 0.6, x_3 = 07$,Take F to be the c.d.f. of U(0; 1); the uniform distribution on (0; 1):(Do these computations by hand - no computer code.) What do you think D^+, D^-, D_N measure, intuitively?

Answer: The Formula of D^+, D^- are as follows:

$D^+ = \text{Max}_{k=1,\ldots,N}(\frac{K}{N} - F(x_k))$, $D^- = \text{Max}_{k=1,\ldots,N}(F(x_k) - \frac{k-1}{N})$,

From the assumptions of the problem $x_1 = 0.2, x_2 = 0.6, x_3 = 07$, Then we can calculate the D^+, D^-, Which were listed as follows:

$$D^+ = \underset{k=1,\ldots,N}{\text{Max}}(\frac{K}{N} - F(x_k)) = \text{Max}\left(\frac{1}{3} - 0.2, \frac{2}{3} - 0.6, 1 - 0.7\right) = 0.3$$

$$D^- = \underset{k=1,\ldots,N}{\text{Max}}(F(x_k) - \frac{k-1}{N}) = \text{Max}\left(0.2 - 0, 0.6 - \frac{1}{3}, 0.7 - \frac{2}{3}\right) = 0.267$$

So the $D_N = \text{Max}(D^+, D^-)$=Max(0.3,0.267)=0.3

2. Prove

$$\mathrm{Max}[F(X_1), \underset{k=1,\dots,N-1}{\mathrm{Max}}\left(F(x_{k+1}) - \frac{k}{N}, \frac{k}{N} - F(x_k)\right), 1 - F(X_N)]$$

$$= \mathrm{Max}[\mathrm{Max}_{k=1,\dots,N}(\tfrac{k}{N} - F(X_k)), \mathrm{Max}_{k=1,\dots,N}(F(X_k) - \tfrac{k-1}{N}))$$

Which was used in the derivation of D_N, See your lecture note.

Answer: The proof is as follows:

Firstly, We can see that $F(X_1) = F(X_1) - 0 = F(X_1) - \frac{0}{N} = F(X_1) - \frac{1-1}{N}$ which is the first term of $F(X_k) - \frac{k-1}{N}$. Similarly, $1 - F(X_N) = \frac{N}{N} - F(X_N)$, which is the last part of $\frac{k}{N} - F(X_k)$.

Secondly, $F(x_{k+1}) - \frac{k}{N}$ $(k = 1,2,\dots,N-1)$ is equal to $F(x_k) - \frac{k-1}{N}$ $(k = 2,3,\dots,N)$

Finally, from the above two results. we can divide $F(X_1)$, $\left(F(x_{k+1}) - \frac{k}{N}, \frac{k}{N} - F(x_k)\right)$ $(k = 1,2,\dots,N)$, $1 - F(X_N)$ into two parts as follows

$F(x_{k+1}) - \frac{k}{N}$ $(k = 1,\dots,N)$ and $F(X_k) - \frac{k-1}{N}$ $(k = 1,\dots,N)$,

Hence,

$$\mathrm{Max}[F(X_1), \underset{k=1,\dots,N-1}{\mathrm{Max}}\left(F(x_{k+1}) - \frac{k}{N}, \frac{k}{N} - F(x_k)\right), 1 - F(X_N)]$$

$$= \mathrm{Max}[F(x_{k+1}) - \tfrac{k}{N}\ (k = 1,\dots,N) \text{ and } F(X_k) - \tfrac{k-1}{N}\ (k = 1,\dots,N)]$$

$$= \mathrm{Max}[\mathrm{Max}_{k=1,\dots,N}(\tfrac{k}{N} - F(X_k)), \mathrm{Max}_{k=1,\dots,N}(F(X_k) - \tfrac{k-1}{N})]$$

The Proof is completed.

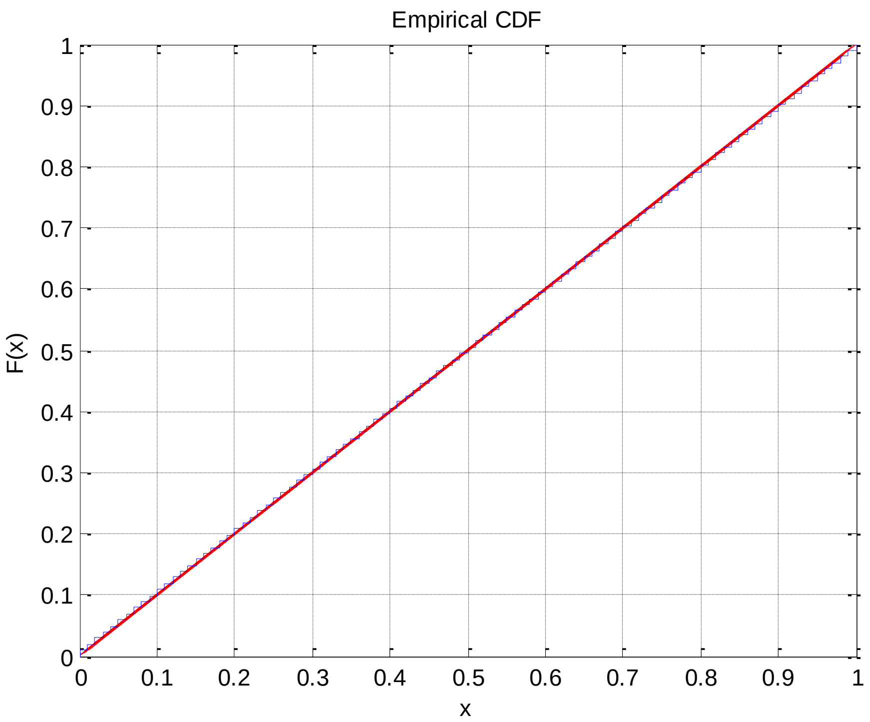
3. Consider the MCG with parameters: a = 23; M = 10^8+1; and let the seed be 47594118. (This is the original MCG proposed by Lehmer in 1948.) Apply the Kolmogorov-Smirnov test to the First 1000 random numbers (including the seed) from this generator. Compute the KS-statistic and Find its p-value. What is your conclusion for the generator?

Answer:

The formula of the MCG is that $X_{n+1} \equiv a * X_n \text{ (Mod M)}$, the seed is 47594118.
With the help of the software Matlab, using the following language,
[H,P,KSSTAT,CV] = kstest(x, [x unifcdf(x,0,1)],0.01)
Where H is whether accepting the Hypothesis or not and the x is the number I generate from the MCG, the unifcdf represent the Uniformly distribution of of [0,1]

We get the p value is the **0.1275 and the KS-Statistics is 0.037.**

It is far bigger than 0.05, so we can conclude that the number generated from this MCG is coming from the uniformly distribution [0,1] at 5% significant level. Besides, from the following chart, we can also see that the difference of the KS test is really small, which demonstrate that the numbers are random.



4. Implement an obviously "bad" random number generator of your choice - you should explain why it is bad. Then apply the χ^2 test together with the KS-statistic as explained in Remark 2, part 3. Take M = 10; N = 1000; and k = 10: What are your conclusions?

Answer:

An immediate example of a "bad" random numbers is the Randu, which generated the numbers laid on the 15 plane.

Firstly, to generate the 10 blocks, I choose the seeds from 1 to 10. for each group(with the different seed), I generate 1000numbrs. As in the formal assignment, The numbers should lied on the 15 plane in R^3, the figure of one group for example is as follows:

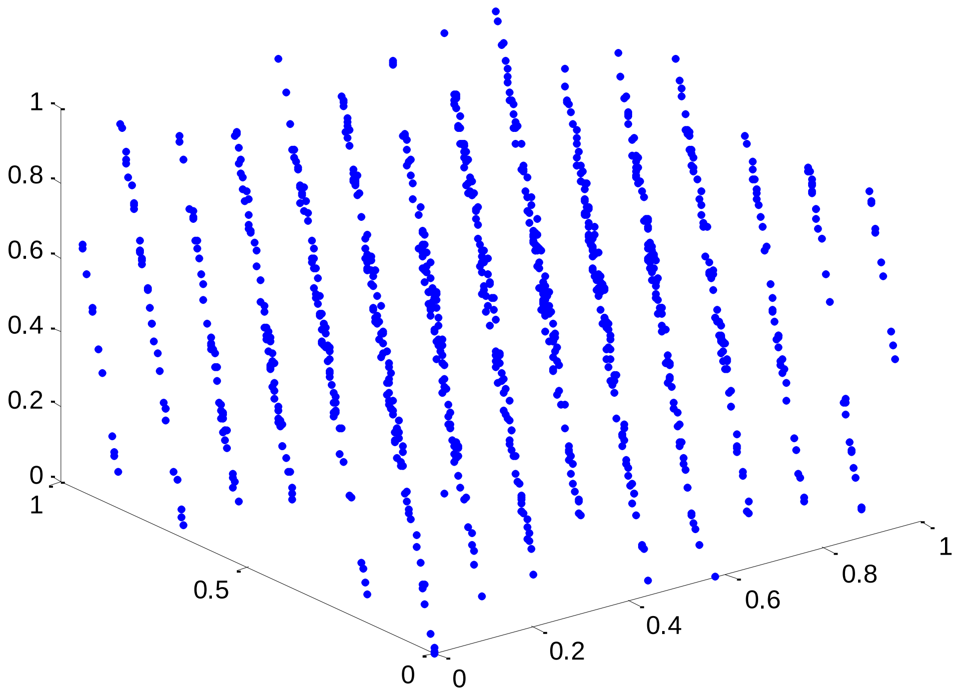


Figure 2 The distribution of Randu random numbers

Secondly, generated the chi square value.

Since the k=10, so the P_i is equal to 0.1 for each $i = 1,..,10$, we just need calculate the Y_i for each group. The results are as follows:

Group	1	2	3	4	5	6	7	8	9	10
Chi square Value	14.20	6.34	12.30	5.06	12.42	6.64	22.04	16.88	8.18	8.16

Since we know that the chi square value correspondent to the $\chi^2(k-1)$ distribution. So we can use the KS Test to verify if the Randu random value correspond the uniform distribution. With the help of the software Matlab, I write the program and got the results as follows:

```
[H,P,KSSTAT,CV] = kstest(x, [x chi2cdf(x,9)],0.05)
```

The p value is the 0.2643 and the KS-Statistics is 0.3031. Although the p value is greater than the significant I chose 0.05, the value is not very good, which means that the chi square I generated are only of nearly 26.43% chance to occur.

My conclusion is that the Randu random numbers are not very good random numbers.

5. Write a computer program for the gap test and apply it to the Fibonacci generator. See the previous lecture notes and examples for the parameters of the generator. I want you to skip the First 100 terms of the generator, and start with the 101st term. Use J = (0.3,0.8) as the subinterval. All the other parameters of the test should be the same as the parameters discussed in Gap Test Examples. What is your conclusion?

Answer:

The formula of the Fibonacci generator is as follows:

$$x_n \equiv x_{n-1} + x_{n-2} (\text{Mod } 2^{31})$$

The $x_1 = 1$ and $x_2 = 1$

Since the J=(0.3,0.8), hence the probability is 0.5, so we have the 4 outcomes with probabilities.

$$P = \{0.5, 0.5^2, 0.5^3, 0.5^3\}$$

Out=$\{0.5,0.25,0.125,0.125\}$

So the 100P should be

$$\{50,25,12.5,12.5\}$$

And the Yi are as follows:

$$\{45,43,1,11\}$$

With the formula of the χ^2 distribution

$$Q(k-1) = \sum_{i=1}^{k} \frac{(Yi - NPi)^2}{Npi}$$

The Chi square value is 24.22, which is a big value and the p value with the freedom degree 3 is 0.0000224726.

6. Do you think we can apply the χ^2 test directly to the run-up counts, assuming that we know the probability of run-up of length i? (Hint: think about the basic assumptions required to apply the χ^2 test, and investigate whether these assumptions are satisfied by "run-up events"

Answer:
My answer is that we can not apply the χ^2 test directly to the run-up counts, despite assuming that we know the probability of run-up of length i.

The main reason is that the statistics of the run-up and the run-down test are not independent. The basic assumption is of χ^2 test is that the experiment should be repeated independently, which means that the Yi should be independently.

From the formula of the Run up Test:

$$u = \frac{1}{n}\sum_{i=1}^{6}\sum_{j=1}^{6}\big(u(i) - nb(i)\big)\big(u(j) - nb(j)\big)a_{ij}$$

where the a_{ij} is the coefficient Matrix as follows:

$$\begin{bmatrix} 4529.4 & 9044.9 & 13568 & 18091 & 22615 & 27892 \\ & 18097 & 27139 & 27139 & 36187 & 45234 \\ & & 40721 & 54281 & 67852 & 83685 \\ & & & 72414 & 90470 & 111580 \\ & & & & 113262 & 139476 \\ & & & & & 172860 \end{bmatrix}$$

We can see that the coefficient matrix is not zero at i<>j, which means that the samples of the experiments are not independent.

However the u has χ^2 distribution with 6 degrees of freedom when n is large (commonly >=4000)

7. Implement the collision test and apply it to the classical MCG by Lehmer that was discussed in the previous lecture examples. Use M = 10, 000 and N = 2000: Let the seed of the MCG be 47594118. You should apply the test to the first 2000 numbers, including the number that corresponds to the seed (i.e., 0.475941). What is your conclusion?

Answer:

$X_{n+1} \equiv a * X_n \text{ (Mod M)}$, where a=23, M=10,000. The seed (X_0) is 47594118.

Besides, the formula of the collision test is that

$$P_{j,n} = \frac{j}{M} P_{j,n-1} + \frac{M-j+1}{M} P_{j-1,n-1}; j = 1,2,\dots,n; n = 1,2,\dots,N$$

With the software VBA, I wrote a program to calculate the $P_{j,n}$, the program is as attachment(Attachment 1).

Since the M is 10000, so the Urns will be 10000, and we should calculate the counts of the random values which have the same first four digits.

The result is that there are 212 collisions, and I listed the CDF of the probability of the collision as follows:

Number of collisions	Probability
150	0.01
160	0.012
180	0.289
190	0.611
200	0.866
210	0.973
215	0.99

We can see that the probability that the collisions are less than 210 are 0.973. **My result is 212, so we can reject the hypothesis that the numbers are random at the 2.5% significant level.**

8. Design a statistical test for random number generators, based on the following result. Then apply the test to any generator you want and explain the results.
FACT: A coin is flipped consecutively until the number of heads obtained equals the number of tails. The output of a flip is heads with probability p. Define the random variable X as: X = the first time the total number of heads is equal to the total number of tails. Observe that X takes values 2, 4, 6, ... For example, if the outcomes of one experiment are: H, H, T, H, T, T then the value of X for this outcome is 6. Here is the probability density function of X:

$$P\{x = 2n\} = \frac{1}{2n-1}\binom{2n}{n} p^n (1-p)^n$$

(For the interested student, a proof of this statement can be found in "Introduction to Probability Models", Sheldon Ross, 8th edition, page 128.)

Answer:
Firstly, I calculated the probability of the distribution and list the table as follows:

Number of n	Probability
1	0.5
2	0.125
3	0.0625
4	0.0390625
5	0.02734375
>=6	0.24609375

Secondly, the process of the Statistics Test is as follows:

1) Generate the random value with the MCG(Lehmer).
2) Select the interval (0,0.5). If the random value is in the interval then the outcome is "H", and vice versa. Since we chose the interval (0,0.5), so the p is 0.5.
3) Calculate the X_i, which is the first time that the total number of the heads equal to the total numbers of tails.
4) Use the χ^2 distribution test to test if we accept or reject that the random value we generated from the MCG are good enough. Choose n = 100 to meet the rule of thumb and let k equals to 4 which is as follows:

Number of n	Probability
1	0.5
2	0.125
3,4,5	0.12890625
>=6	0.24609375

Finally, the result is as follows:

Xi	Expected value	(Yi-Npi)^2/Npi
57	50	0.98
12	12.5	0.02
11	12.890625	0.277291667
20	24.609375	0.863343254

The Q_{k-1} is 2.140634921 and the p value of 3 degree of freedom is 0.54373589, which shows that the MCG random generator is fairly good.
The program of the VBA to calculate the Xi is as attachment 2.

Attachment 1

```
Sub Collision_Test()
Dim i, j As Integer
Dim a(2000, 2000) As Double
m = 10000
n = 2000
a(1, 1) = 1
For i = 2 To n
a(i, i) = a(i - 1, i - 1) * (m - i + 1) / m
Next i
For j = 1 To 2000
For k = j + 1 To 2000
a(j, k) = j / m * a(j, k - 1) + (m - j + 1) / m * a(j - 1, k - 1)
Next k
Next j
For i = 1 To n
Worksheets("sheet1").Range("f3").Select
ActiveCell.Offset(i - 1, 0).Value = a(i, 2000)
Next i

End Sub
```

Attachment 2

```
Sub Number_test()
Dim i, j, k As Integer
k = 0
n = 15000
Number = 0
number_ind = 0
Worksheets("sheet2").Select
Range("c2").Select
For i = 1 To n
If ActiveCell.Value = "H" Then
k = k + 1
Else
k = k - 1
End If
Number = Number + 1
If k = 0 Then
number_ind = number_ind + 1
Cells(number_ind + 1, "d").Value = Number
Number = 0
k = 0
End If
ActiveCell.Offset(1, 0).Select
Next i
End Sub
```