UNIT-3

Random Number Generation

Random numbers are essential part of simulation of almost all discrete systems. Most of the programming languages have builtin functions/objects to generate a random number. Basically, random numbers are used to generate various random variables like event time, service time etc. Here, we will discuss various methods of generating random numbers, testing for randomness, using random numbers to generate a random variable with the help of probability distributions etc.

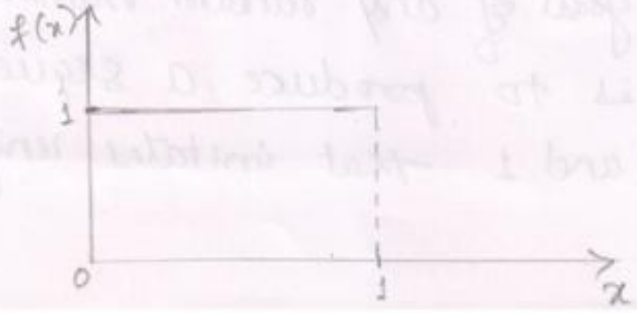
3.1 Properties of Random Numbers

A sequence of random numbers $R_1, R_2, \ldots$ must have two statistical properties: uniformity and independence.

Each random number R_i must be an independent sample drawn from continuous uniform distribution between 0 and 1. That is, pdf is given by—

$$f(x) = \begin{cases} 1, & 0 \le x \le 1 \\ 0, & \text{otherwise} \end{cases}$$

The density function is shown in the following figure—



The expectation of random number R_i is —

$$E(R) = \int_0^1 x.dx = \frac{x^2}{2}\Big|_0^1$$

$$= \frac{1^2}{2} - \frac{0^2}{2}$$

$$= \frac{1}{2}$$

$$V(R) = E(R^2) - [E(R)]^2$$

$$= \int_0^1 x^2 dx - \left(\frac{1}{2}\right)^2$$

$$= \frac{x^3}{3}\Big|_0^1 - \frac{1}{4} = \frac{1}{3} - 0 - \frac{1}{4}$$

$$= \frac{1}{12}$$

3.2 Generation of Pseudo-Random numbers

There are several methods to generate random numbers. As methods are known to us well in advance, the term 'random' may not exactly mean what it indicates. Hence, the term 'Pseudo-random' is used.

The goal of any random number generation scheme is to produce a sequence of numbers between 0 and 1 that imitates uniform distribution.

In the generation of random numbers, certain problems may occur as below -

* Generated numbers may not be uniformly distributed.
* Generated numbers may be discrete, instead of continuous.
* The mean of generated numbers may be too low or too high.
* The variance of the numbers may be too low or too high.
* There may be dependency due to auto correlation

There are several methods to generate random numbers using computer with the following conditions:

(i) The method should be fast.
(ii) The method should be portable to different computers, and to different progearmming languages.
(iii) The method should have sufficient long cycle. That is - the random numbers generated may start repeating. From one sequence to another sequence, there must be enough gap.
(iv) The random numbers should be repeatable independent of the system which generated it.
(v) Generated random numbers should closely approximate the statistical properties of uniformity and independence.

3.3 Techniques for Generating Random Numbers

Various techniques for generating random numbers are discussed here.

3.3.1 Linear Congruential Method

The linear congruential method produces a sequence of integers $X_1, X_2, \ldots$ between 0 and $m-1$ using the formula –

$$X_{i+1} = (a.X_i + c) \bmod m, \quad i = 0, 1, 2, \ldots \quad \text{—— Eq ①}$$

Here, the initial value X_0 is called as seed

a is .. .. multiplier

c is .. .. increment

m is .. .. modulus.

If $c \neq 0$ in the above equation, then it is called as mixed congruential method. If $c = 0$, then it is known as multiplicative congruential method. The selection of the values for a, c, m and X_0 drastically affects the statistical properties and the cycle-length.

Note that, using the above equation, random integers are generated. But, the requirement is random numbers in the range of 0 to 1. Hence, they are calculated as –

$$R_i = \frac{X_i}{m}, \quad i = 1, 2, 3, \ldots \quad \text{—— Eq ②}$$

Example:

Use the linear congruential method to generate a sequence of random numbers with $X_0 = 27$, $a = 17$, $c = 43$ and $m = 100$.

Solution: The sequence of seeds X_i and the subsequent random numbers R_i are computed as shown below –

$X_0 = 27$

$X_1 = (a \cdot X_0 + c) \bmod m$

$= (17 \cdot 27 + 43) \bmod 100$

$= 2$

Now, $R_1 = \frac{2}{100} = 0.02$

$X_2 = (a \cdot X_1 + c) \bmod m$

$= (17 \cdot 2 + 43) \bmod 100$

$= 77$

$R_2 = \frac{77}{100} = 0.77$

$X_3 = (17 \cdot 77 + 43) \bmod 100$

$= 52$

$R_3 = 0.52$

Continuing in the above manner, we can generate as many random numbers as required.

*****

① Generate Random no's using Linear Congruential method $X_0 = 23$, $a = 8$ $C = 47$ and $m = 100$. $[X_{i+1} = (aX_i + C) \bmod m]$

Ans :- $X_1 = (8(23) + 47) \bmod (100)$

$= (231) \bmod 100 = 31$

$\boxed{X_1 = 31}$

$R_1 = \frac{X_1}{m} = \frac{31}{100} = 0.31$

$\boxed{R_1 = 0.31}$

$X_2 = (8(31) + 47) \bmod (100)$

$= (295) \bmod 100 = 95$

$\boxed{X_2 = 95}$

$R_2 = \frac{X_2}{m} = \frac{95}{100} = 0.95$

$\boxed{R_2 = 0.95}$

$X_3 = (8(95) + 47) \bmod (100)$

$= (807) \bmod (100) = 7$

$\boxed{X_3 = 7}$

$R_3 = \frac{X_3}{m} = \frac{7}{100} = 0.07$

$\boxed{R_3 = 0.07}$

$X_4 = [8(7) + 47) \bmod (100)$

$= (103) \bmod (100) = 3$

$\boxed{X_4 = 3}$

$R_4 = \frac{X_4}{m} = \frac{3}{100} = 0.03$

$\boxed{R_4 = 0.03}$

$X_5 = (8(3) + 47) \bmod (100)$

$= (71) \bmod 100 = 71$

$\boxed{X_5 = 71}$

$R_5 = \frac{X_5}{m} = \frac{71}{100} = 0.71$

$\boxed{R_5 = 0.71}$

NOTE: Apart from uniformity and independence properties of random numbers, there are two more: maximum density and maximum period.

It can be observed from equation ② that the random numbers will be in the set

$$I = \left\{0, \frac{1}{m}, \frac{2}{m}, \cdots \frac{m-1}{m}\right\}$$

because each X_i is an integer in the set $\{0, 1, \cdots m-1\}$ Thus, each R_i is discrete in I instead of continuous on $[0, 1]$. But, if m is very large, then this problem can be reduced. The meaning of maximum density is there is no large gap among R_i in $[0, 1]$

To achieve maximum period (of cycle of repetition of R_i), the proper choice of a, c, m and X_0 are made:

(i) If $m = 2^b$ and $c \neq 0$, the longest possible period is $P = m = 2^b$, when c is relatively prime to m. And, $a = 1 + 4k$, where k is integer.

(ii) If $m = 2^b$ and $c = 0$, then longest possible period $P = \frac{m}{4} = 2^{b-2}$. This is achieved when X_0 is odd and $a = 3 + 8k$ or $a = 5 + 8k$.

(iii) If m is prime number and $c = 0$, then $P = m - 1$. This is achieved when a has a property that the smallest integer k such that $a^k - 1$ is divisible by m is $k = m - 1$.

Tests for Random Numbers:

To check whether the random numbers satisfy the properties uniformity and independence, few tests are performed :-

1. Frequency Test : Uses Kolmogorov-Smirnov Test or Chi-square test to compare distribution of the set of numbers generated to verify whether they follow Uniform distribution.
2. Auto correlation Test: Tests correlation between numbers. The expected correlation is zero. Hence the sample correlation is compared with zero.

The hypotheses for testing uniformity are as follows:

$H_0 : R_i \sim \text{Uniform}[0, 1]$

$H_1 : R_i \not\sim \text{Uniform}[0, 1]$

(NOTE: Here H_0 is called as Null hypothesis, that is, our assumption. H_1 is known as Alternative hypothesis. In sampling theory, H_0 is tested for its truthness. If it is found to be false, the H_1 is accepted.)

If numbers are uniformly distributed, then H_0 is accepted.

Testing for independence has the hypotheses as -

H_0 : $R_i \sim$ independently

H_1 : $R_i \not\sim$ independently

If H_0 is accepted, it means that evidence of dependency is not found in this test.

For each test, a level of significance α must be stated. The α indicates the probability of rejecting the null hypothesis when it is true.

i.e. $\alpha = P(\text{reject } H_0 \mid H_0 \text{ is true})$

In general, α is set as 0.01 or 0.05. When $\alpha = 0.01$, we say that the test result is 99% confident. And if $\alpha = 0.05$, the confidence level is 95%. Also, if we reject H_0, when it is true, it is known as Type - I Error.

Frequency Tests

The test of uniformity is very basic to validate the random number generator. We will discuss two methods: K-S Test and Chi-square Test. The null hypothesis in both of these tests is - there is no significant difference between sample distribution and theoretical distribution.

The Kolmogorov - Smirnov Test (K-S Test):

This test compares cdf $F(x)$ of uniform distribution with the empirical cdf $S_N(x)$ of the sample of N observations. By definition

$$F(x) = x, \quad 0 \le x \le 1$$

If the sample from random numbers generator is $R_1, R_2, \ldots, R_N$, then $S_N(x)$ is —

$$S_N(x) = \frac{\text{number of } R_1, R_2 \ldots R_N \text{ which are} \le x}{N}$$

As N increases, $S_N(x)$ will be nearer to $F(x)$.

The K-S test is based on the statistic

$$D = \max\left(|F(x) - S_N(x)|\right)$$

The sampling distribution of D is known. It is tabulated as a function of N and will be given in a pre-defined Table.

The procedure for K-S test is given below —

step(1) Arrange the given data in ascending order. i.e.

$$R_1 \le R_2 \le R_3 \le \ldots \le R_N$$

step(2) Compute

$$D^+ = \max_{1 \le i \le N} \left\{ \frac{i}{N} - R_i \right\}$$

$$D^- = \max_{1 \le i \le N} \left\{ R_i - \frac{i-1}{N} \right\}$$

step(3) Compute $D = \max(D^+, D^-)$.

Step(4) Locate the critical value D_α in a pre-defined table for a specified α and for given N.

Step(5) If the sample statistic $D > D_\alpha$ then the null hypothesis (that the sample is of uniform distribution) is rejected. If $D \leq D_\alpha$, then we can conclude that there is no significant difference between the distribution of $\{R_1, \ldots R_N\}$ and the uniform distribution.

Example:

Five random numbers are generated: 0.44, 0.81, 0.14, 0.05 and 0.93. Test for uniformity using K-S Test, with the level of significance $\alpha = 0.05$.

Solution: Arrange the random numbers in the ascending order and mark them as $R_1, R_2, \ldots$. The calculations are done as shown in the table given below—

i	R_i	$\frac{i}{N}$	$\frac{i}{N} - R_i$	$R_i - \frac{i-1}{N}$
1	0.05	0.20	0.15	0.05
2	0.14	0.40	0.26	—
3	0.44	0.60	0.16	0.04
4	0.81	0.80	—	0.21
5	0.93	1.00	0.07	0.13

Now, $D^+ = \max\left\{\frac{i}{N} - R_i\right\}$

$= 0.26$

and, $D^- = \max\left\{R_i - \frac{i-1}{N}\right\}$

$= 0.21$

$\therefore D = \max\{D^+, D^-\}$

$= 0.26$

Now, refer the table "Kolmogorov-Smirov critical values" for $N = 5$ and $\alpha = 0.05$. It is found from table that $D_\alpha = 0.565$.

Thus, computed value $D \leq D_\alpha$. Hence the null hypothesis can be accepted. That is, we can accept that the sequence of given random numbers follows uniform distribution.

The Chi-Square Test:

The chi-square test uses the sample statistic -

$$\chi_0^2 = \sum_{i=1}^{n} \frac{(O_i - E_i)^2}{E_i}$$

Here, O_i is the observed number in the i^{th} class,
E_i is the expected number in the i^{th} class
n is number of classes.

For uniform distribution, $E_i = \frac{N}{n}$, where N is total number of observations. It can be shown that χ_0^2 follows chi-square distribution with $n-1$ degrees of freedom.

If computed $\chi_0^2 \leq$ table value of χ^2 for a given α, then null hypothesis is accepted.

Example:

Use χ^2 test with $\alpha = 0.05$ for the following data to check the uniformity:

0.34	0.90	0.25	0.89	0.87	0.44	0.12	0.21	0.46	0.67
0.83	0.76	0.79	0.64	0.70	0.81	0.94	0.74	0.22	0.74
0.96	0.99	0.77	0.67	0.56	0.41	0.52	0.73	0.99	0.02
0.47	0.30	0.17	0.82	0.56	0.05	0.45	0.31	0.78	0.05
0.79	0.71	0.23	0.19	0.82	0.93	0.65	0.37	0.39	0.42
0.99	0.17	0.99	0.46	0.05	0.66	0.10	0.42	0.18	0.49
0.37	0.51	0.54	0.01	0.81	0.28	0.69	0.34	0.75	0.49
0.72	0.43	0.56	0.97	0.30	0.94	0.96	0.58	0.73	0.05
0.06	0.39	0.84	0.24	0.40	0.64	0.40	0.19	0.79	0.62
0.18	0.26	0.97	0.88	0.64	0.47	0.60	0.11	0.29	0.78

Solution: There are 100 numbers in a given list in the interval $[0, 1]$. That is, $N = 100$. Let us take $n = 10$ intervals of equal length like—

$[0, 0.1)$, $[0.1, 0.2)$, $[0.2, 0.3)$ ..... $[0.9, 1.0)$.

We expect all 100 values should be equally distributed over these 10 intervals. That is, each interval should have 10 values. But, the actual data may not fit exactly with this expectation. Now, we will calculate the test statistic χ^2 using the formula

$$\chi^2 = \sum_{i=1}^{10} \frac{(O_i - E_i)^2}{E_i}$$

Here, each $E_i = 10$.

And, O_i indicates total number of elements actually observed in the interval i. That is,

O_1 indicates no of elements in the range $[0, 0.1)$

O_2 " " " " " " $[0.1, 0.2)$

and so on.

Interval	O_i	E_i	$O_i - E_i$	$(O_i - E_i)^2$	$\frac{(O_i - E_i)^2}{E_i}$
1	7	10	-3	9	0.9
2	9	10	-1	1	0.1
3	8	10	-2	4	0.4
4	9	10	-1	1	0.1
5	14	10	4	16	1.6
6	7	10	-3	9	0.9
7	10	10	0	0	0.0
8	15	10	5	25	2.5
9	9	10	-1	1	0.1
10	12	10	2	4	0.4
Total	100	100			7.0

Thus, $\chi_0^2 = 7.0$

Now, using χ^2 table for $\alpha = 0.05$ and the degrees of freedom $= n - 1 = 9$, we will get-

$\chi^2_{0.05,\,9} = 16.9$

Since, $\chi_0^2 \leq \chi^2_{0.05,\,9}$ we can accept the null hypothesis. That is, the random numbers are uniformly distributed.

Tests for autocorrelation:

There is a chance that the generated random numbers have some kind of dependency like -

(i) numbers are repeated at regular intervals.

(ii) numbers at certain positions (like 5th, 10th, 15th etc) are very low or very high compared to other numbers.

Here, we will discuss whether such relationship exist among numbers or not.

The test for autocorrelation computes the auto-correlation between every l (ie lag) numbers starting with i^{th} number. Thus the autocorrelation ρ_{il} between following numbers has to be computed -

$$R_i, R_{i+l}, R_{i+2l}, \ldots R_{i+(M+1)l}.$$

Here, M is largest integer such that

$$i + (M+1)l \leq N.$$

And, N is total number of random numbers in a given sequence.

If the autocorrelation value is zero, then we can say that numbers are independent.

ie

$$H_0 : \rho_{il} = 0$$

$$H_1 : \rho_{il} \neq 0$$

The test statistic for autocorrelation test is given by-

$$Z = \frac{\hat{\rho}_{iL}}{\sigma_{\hat{\rho}_{iL}}}$$

Here,

$$\hat{\rho}_{iL} = \frac{1}{M+1}\left[\sum_{k=0}^{M} R_{i+kL} \cdot R_{i+(k+1)L}\right] - 0.25$$

and

$$\sigma_{\hat{\rho}_{iL}} = \frac{\sqrt{13M+7}}{12(M+1)}$$

The test statistic Z_0 follows standard normal distribution with mean 0 and variance 1. Hence, Z_0 is tested against $\alpha = 0.05$ or 0.01. The null hypothesis is accepted if

$$-Z_{\alpha/2} \leq Z_0 \leq Z_{\alpha/2}$$

The value of $Z_{\alpha/2}$ can be obtained from pre-defined cumulative normal distribution table.

Example:

Test for whether 3rd, 8th, 13th.... numbers in the following sequence are autocorrelated, for $\alpha = 0.05$.

0.12	0.01	0.23	0.28	0.89	0.31	0.64	0.28	0.83
0.93	0.99	0.15	0.33	0.35	0.91	0.41	0.60	0.27
0.75	0.88	0.68	0.49	0.05	0.43	0.95	0.58	0.19
0.36	0.69	0.87						

Solution: We have to start with 3rd number.

Hence $i = 3$.

Every 5th number has to be checked as per the question (3rd, 8th, 13th etc, difference is 5).

Hence, $L = 5$.

Total no of values, $N = 30$.

So, M is calculated as -

$$i + (M+1)L \leq N$$

$$\Rightarrow 3 + (M+1).5 \leq 30$$

$$\Rightarrow M \leq \frac{27}{5} - 1$$

$$\Rightarrow M \leq 4.4$$

M is largest integer ≤ 4.4

$$\Rightarrow M = 4.$$

Now, we need to compute,

$$\hat{\rho}_{iL} = \frac{1}{M+1}\left[\sum_{k=0}^{M} R_{i+kL} \cdot R_{i+(k+1)L}\right] - 0.25$$

$$\Rightarrow \hat{\rho}_{35} = \frac{1}{4+1}\left[R_3 \cdot R_8 + R_8 \cdot R_{13} + R_{13} \cdot R_{18} + R_{18} \cdot R_{23} + R_{23} \cdot R_{28}\right] - 0.25$$

$$= \frac{1}{5}\left[(0.23)(0.28) + (0.28)(0.33) + (0.33)(0.27) + (0.27)(0.05) + (0.05)(0.36)\right] - 0.25$$

$$= \frac{1}{5} \cdot 0.2774 - 0.25$$

$$= -0.1945$$

Now,

$$\sigma_{\hat{\rho}_{35}} = \frac{\sqrt{13M+7}}{12(M+1)} = \frac{\sqrt{13 \cdot 4 + 7}}{12(4+1)}$$

$$= \frac{\sqrt{59}}{60}$$

$$= 0.1280.$$

$$\therefore Z_0 = \frac{-0.1945}{0.1280} = -1.5195$$

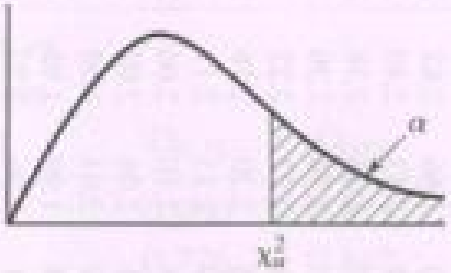
Now, from the table, $Z_{0.025} = 1.96$

As $-Z_{\alpha/2} \leq Z_0 \leq Z_{\alpha/2}$ is satisfied —

$$-1.96 \leq -1.5195 \leq 1.96$$

Therefore, H_0 is accepted. That is, random numbers are independent.

Table A.6 Percentage Points of The Chi-Square Distribution with ν Degrees of Freedom



ν	$\chi^2_{0.005}$	$\chi^2_{0.01}$	$\chi^2_{0.025}$	$\chi^2_{0.05}$	$\chi^2_{0.10}$
1	7.88	6.63	5.02	3.84	2.71
2	10.60	9.21	7.38	5.99	4.61
3	12.84	11.34	9.35	7.81	6.25
4	14.96	13.28	11.14	9.49	7.78
5	16.7	15.1	12.8	11.1	9.2
6	18.5	16.8	14.4	12.6	10.6
7	20.3	18.5	16.0	14.1	12.0
8	22.0	20.1	17.5	15.5	13.4
9	23.6	21.7	19.0	16.9	14.7
10	25.2	23.2	20.5	18.3	16.0
11	26.8	24.7	21.9	19.7	17.3
12	28.3	26.2	23.3	21.0	18.5
13	29.8	27.7	24.7	22.4	19.8
14	31.3	29.1	26.1	23.7	21.1
15	32.8	30.6	27.5	25.0	22.3
16	34.3	32.0	28.8	26.3	23.5
17	35.7	33.4	30.2	27.6	24.8
18	37.2	34.8	31.5	28.9	26.0
19	38.6	36.2	32.9	30.1	27.2
20	40.0	37.6	34.2	31.4	28.4
21	41.4	38.9	35.5	32.7	29.6
22	42.8	40.3	36.8	33.9	30.8
23	44.2	41.6	38.1	35.2	32.0
24	45.6	43.0	39.4	36.4	33.2
25	49.6	44.3	40.6	37.7	34.4
26	48.3	45.6	41.9	38.9	35.6
27	49.6	47.0	43.2	40.1	36.7
28	51.0	48.3	44.5	41.3	37.9
29	52.3	49.6	45.7	42.6	39.1
30	53.7	50.9	47.0	43.8	40.3
40	66.8	63.7	59.3	55.8	51.8
50	79.5	76.2	71.4	67.5	63.2
60	92.0	88.4	83.3	79.1	74.4
70	104.2	100.4	95.0	90.5	85.5
80	116.3	112.3	106.6	101.9	96.6
90	128.3	124.1	118.1	113.1	107.6
100	140.2	135.8	129.6	124.3	118.5

Table A.8 Kolmogorov–Smirnov Critical Values

Degrees of Freedom (*N*)	$D_{0.10}$	$D_{0.05}$	$D_{0.01}$
1	0.950	0.975	0.995
2	0.776	0.842	0.929
3	0.642	0.708	0.828
4	0.564	0.624	0.733
5	0.510	0.565	0.669
6	0.470	0.521	0.618
7	0.438	0.486	0.577
8	0.411	0.457	0.543
9	0.388	0.432	0.514
10	0.368	0.410	0.490
11	0.352	0.391	0.468
12	0.338	0.375	0.450
13	0.325	0.361	0.433
14	0.314	0.349	0.418
15	0.304	0.338	0.404
16	0.295	0.328	0.392
17	0.286	0.318	0.381
18	0.278	0.309	0.371
19	0.272	0.301	0.363
20	0.264	0.294	0.356
25	0.24	0.27	0.32
30	0.22	0.24	0.29
35	0.21	0.23	0.27
Over 35	$\frac{1.22}{\sqrt{N}}$	$\frac{1.36}{\sqrt{N}}$	$\frac{1.63}{\sqrt{N}}$

Source: F. J. Massey, "The Kolmogorov–Smirnov Test for Goodness of Fit," *The Journal of the American Statistical Association*, Vol. 46. © 1951, p. 70. Adapted with permission of the American Statistical Association.