Unit-5: Random Numbers

Introduction to Random Numbers:

Random numbers are characterized by the fact that their value cannot be predicted. Or, in other words, if one constructs a sequence of random numbers, the probability distribution of the following random numbers has to be completely independent of all the other generated numbers. Random numbers are samples drawn from a uniformly distributed random variable between some satisfied intervals they have equal probability of occurrence.

Random numbers are a necessary basic ingredient in the simulation of almost all discrete systems. Most computer languages have a subroutine, object, or function that will generate a random number. Similarly simulation languages generate random numbers that are used to generate event times and other random variables.

Random Number Tables

A table of numbers generated in an unpredictable, haphazard (hit-or-miss) that are uniformly distributed within certain interval are called random number table.

The random number in random number table exactly obeys (follow) two random number properties:

1. Uniformity
2. Independence

So, random number generated to form of table also called true random numbers.

Table of random numbers are used to create a Radom sample. A random number table is also called random sample table. There are many physical devices or process that can be used to generate a sequence of uniformly distributed random numbers i.e. true random numbers.

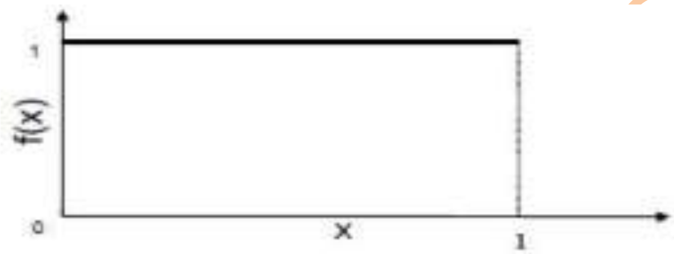
For example: An electrical pulse generator can be made to drive a counter cycling from 0 to 9. Using an electronic noise generator or radioactive source the pulse can be generated as random numbers.

<u>Properties of Random Numbers:</u>

1. A sequence of random numbers, $R1, R2, R3, \ldots\ldots Rn$ must have two important properties:
   - **Uniformity**, i.e. they are equally probable/distributed every where
   - **Independence**, i.e. the current value of a random variable has no relation with the previous values
2. Each random number R_t is an independent sample drawn from a continuous uniform distribution between 0 and 1

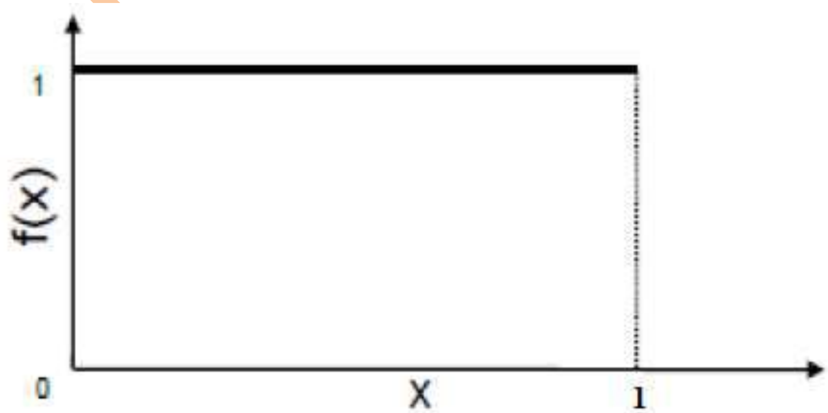
$$\text{pdf:} \quad f(x) = \begin{cases} 1, & 0 \le x \le 1 \\ 0, & \text{otherwise} \end{cases}$$

(Probability Distribution Function)



Expectation:

PDF:



$$E(R) = \int_0^1 x\,dx = [x^2/2]_0^1 = 1/2$$

Variance: Variance is Measure of the degree to which data are dispersed from the mean value

$$V(R) = \int_0^1 x^2 dx - [E(R)]^2$$
$$= [x^3/3]_0^1 - (1/2)^2 = 1/3 - 1/4$$
$$= 1/12$$

Types of Random Numbers:

There are three types of random numbers,

- True random numbers
- Pseudo random numbers
- Quasi random numbers

These different types of random numbers have different applications.

True Random Number:

True random numbers are gained from physical processes like radioactive decay (*decomposition*) or also rolling a dice. But rolling a dice is difficult perhaps someone could control the dice so well to determine the outcome.

The most often used example for "truly" random numbers is the decay of a radioactive material.

If a Geiger counter is put in front of such a radioactive source, the intervals between the decay events are truly random.

These random numbers are not generated by any algorithm but they are generated naturally from physical phenomenon and these are truly unpredictable, like events occurrence in nature.

Pseudo random numbers:

Pseudo means false, so false random numbers are being generated. These numbers are generated by a computer or that is to say, by an algorithm and because of this not truly random, these can be predictive.

Every new number is generated from the previous ones by an algorithm. This means that the new value is fully determined by the previous ones. But, depending on the algorithm, they often have properties making them very suitable for simulations.

Since the arithmetic operation is known and the sequence of random numbers can be repeated, the numbers cannot be called *truly random number*.

When generating pseudo-random numbers, certain problems or errors can occur. Some examples of errors includes the following

1. The generated numbers may not be uniformly distributed.
2. The generated numbers may be discrete-valued instead continuous valued
3. The mean of the generated numbers may be too high or too low.
4. The variance of the generated numbers may be too high or low
5. There may be dependence. The following are examples:
   - Autocorrelation between numbers.
   - Numbers successively higher or lower than adjacent numbers.
   - Several numbers above the mean followed by several numbers below the mean.

Quasi random numbers:

Quasi (Virtual) random numbers are not designed to appear random, rather to be uniformly distributed.

One aim of such numbers is to reduce and control errors in Monte Carlo simulations.

Properties of Good random Number Generators:

Usually, random numbers are generated by a digital computer as part of the simulation. Numerous methods can be used to generate the values. In selecting among these methods, or routines, there are a number of important considerations.

- **The routine should be fast**: The total cost can be managed by selecting a computationally efficient method of random-number generation.
- **The routine should be portable to different computers, and ideally to different programming languages.** This is desirable so that the simulation program produces the same results wherever it is executed.
- **The routine should have a sufficiently long cycle:** The cycle length, or period, represents the length of the random-number sequence before previous numbers begin to repeat themselves in an earlier order. Thus, if 10,000 events are to be generated, the period should be many times that long; a special case cycling is degenerating. A routine degenerates when the same random numbers appear repeatedly. Such an occurrence is certainly unacceptable. This can happen rapidly with some methods.
- **The random numbers should be replicable:** Given the starting point (or conditions), it should be possible to generate the same set of random numbers, completely independent of the system that is being simulated. This is helpful for debugging purpose and is a means of facilitating comparisons between systems.
- Most important, and as indicated previously, the generated random numbers should closely approximate the ideal statistical properties of uniformity and independences

Period in Random Number:

In random number generation, the term "period" refers to the length of the sequence before a random number generator (RNG) repeats itself and starts producing numbers that have been previously generated. In other words, the period is the number of unique values that an RNG can generate before it cycles back to its initial state and starts producing the same sequence of numbers again.

For example, consider a simple random number generator that produces values between 0 and 9. If this generator has a period of 10, it means that after generating 10 unique values, it will start producing the same sequence of values all over again.

Methods of Generating Random Numbers:

There are several methods of Generating Random Numbers some of them are:

- Linear Congruential Method
- Mid-Square Method

Linear Congruential Method:

The linear congruential method, initially proposed by Lehrer [1951], produces a sequence of integers, X_1, X_2,... between zero to m-1 according to the following recursive relationship:

$X_{i+1} = (a X_i + c) \bmod m$ ………………... Eq-1

Where, i=0,1,2,3………

The initial value $\boldsymbol{X_0}$ is called the seed, ***a*** is called the constant multiplier, ***c*** is the increment, and ***m*** is the modulus.

Case 1: If $c \neq 0$ in Eq-1, the form is called the **mixed congruential method.**

Case 2: When $c = 0$, the form is known as the **multiplicative congruential** method.

The selection of the values for ***a, c, m*** and $\boldsymbol{X_0}$ drastically affects the statistical properties and the cycle length.

Let's illustrate the different congruential Methods with the help of example.

Mixed Congruential Method: $c \neq 0$

Let a = 9, c = 3 , m = 31 & X_0 = 2

We know that,

$X_{i+1} = (a X_i + c) \bmod m$

Then,

$X_1 = (aX_0 + c) \bmod m$

$= (9*2 + 3) \bmod 31$

$= 21 \bmod 31$

$\mathbf{X_1 = 21}$

$X_2 = (aX_1 + c) \bmod m$

$= (9*21 + 3) \bmod 31$

$= 192 \bmod 31$

$\mathbf{X_2 = 6}$

And $\mathbf{X_3=26}$ **,** $\mathbf{X_4=20}$**,** $\mathbf{X_5=28}$

Hence, random numbers are **2,21,6,26,20,28,7,4,8**

Multiplicative Congruential Method: c = 0

Let a = 9, m = 31 & $X_0 = 2$

We know that,

$X_{i+1} = (a X_i + c) \bmod m$

If c = 0, then

$X_{i+1} = (a X_i) \bmod m$

$X_1 = (aX_0) \bmod m$

$= (9*2) \bmod 31$

$= 18 \bmod 31$

$\mathbf{X_1 = 18}$

$X_2 = (aX_1) \bmod m$

$= (9*18) \bmod 31$

= 162 mod 31

X_2 = 7

And **X_3=1, X_4=9, X_5=19**

Hence, random number are **2,18,7,1,9,19,16,20…..**

Additive Congrutial Method: a =1

Let c = 17, m = 29 & X0 = 7

We know that,

$X_{i+1} = (a X_i + c) \bmod m$

If a =1, then

$X_{i+1} = (X_i + c) \bmod m$

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

= (7 + 17) mod 29

= 24 mod 29

X_1= 24

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

= (24 + 17) mod 29

= 41 mod 29

X_2= 12

And **X_3=0, X_4=17, X_5=5**

Hence, random number are **7,24,12,0,17,5,22,10,27…….**

Arithmetic Congruential Method:

In this method random number are generated by the equation:

$X_{i+1} = (X_{i-1} + X_i) \bmod m$

Let, $X_1=9$, $X_2=13$, $m=17$

Then,

$X_3=(X_1 + X_2) \bmod m$

$= (9 + 13) \bmod 17$

$\mathbf{X_3= 5}$

$X_4=(X_2 + X_3) \bmod m$

$= (13 + 5) \bmod 17$

$\mathbf{X_4 = 1}$

………………..

So

The random numbers are **9,13,5,1,6,7,13,3………..**

Example 1:

Use the linear congruential method to generate a sequence of random numbers with $X_0 = 27$, $a= 17$, $c = 43$, and $m = 100$. Here, the integer values generated will all be between zero and 99 because of the value of the modulus. These random integers should appear to be uniformly distributed the integers from zero to 99. Random numbers between zero and 1 can be generated by:

$R_i =X_i / m$ ………………….Eq-2

Where, i = 1,2,3,……

Solution

The sequence of X_i and subsequent R_i values is computed as follows:

$X_0 = 27$

$X_1 = (aX_0 + c) \bmod m$

$X_1 = (17*27 + 43) \bmod 100$

$= 502 \bmod 100$

$\mathbf{X_1 = 2}$

$R_1 = X_1 / m$

$= 2/100$

$\mathbf{R_1 = 0.02}$

$X_2 = (aX_1 + c) \bmod m$

$X_2 = (17*2 + 43) \bmod 100$

$= 77 \bmod 100$

$\mathbf{X_2 = 77}$

$R_2 = X_2 / m$

$R_2 = 77/100$

$\mathbf{R_2 = 0.77}$

Likewise,

$X_3 = (17 \text{ x } 77 + 43) \bmod 100$

$= 1352 \bmod 100$

$\mathbf{X_3 = 52}$

$R_3 = 52/100$

$\mathbf{R_3 = 0.52}$

Example 2:

Let m = 100, a = 19, c = 0, and X_0 = 63. Generate a sequence of random integers using $X_{i+1} = (a\, X_i + c) \bmod m$.

Solution:

$X_0 = 63$

$X_1 = (19*63) \bmod 100$

$= 1197 \bmod 100$

$\mathbf{X_1 = 97}$

$X_2 = (19*97) \bmod 100$

$= 1843 \bmod 100$

$\mathbf{X_2 = 43}$

$X_3 = (19*43) \bmod 100$

$= 817 \bmod 100$

$\mathbf{X_3 = 17}$

………..

Example 3:

Using the multiplicative congruential method, find the period of the generator for a = 13, m = 64, and X_0 = 1, 2, 3, and 4. Prove that the solution is given , when the seed is 1 and 3, the sequence has period 16, a period of length eight is achieved when the seed is 2 and a period of length four occurs when the seed is 4.

Solution:

Following table shows the periods of random number with different seeds:

i	X_i	X_i	X_i	X_i
0	1	2	3	4
1	13	26	39	52
2	41	18	59	36
3	21	42	63	20
4	17	34	51	4
5	29	58	23	
6	57	50	43	
7	37	10	47	
8	33	2	35	
9	45		7	
10	9		27	
11	53		31	
12	49		19	
13	61		55	
14	25		11	
15	5		15	
16	1		3	

Here, we can see,

If seed is 1 and 3 period is 16 i.e. after 16th random number seed is repeated.

If seed is 2 period is 8 and if seed is 4 period is 4.

Assignment

1. Let m = 100, a = 19, c = 0, and X0 = 63, and generate a sequence random integers. Find first 7 random numbers generate using any suitable method?
2. Let $a = 7^5 = 16,807$, $m = 2^{31}-1 = 2,147,483,647$ (a prime number), and c= 0. These choices satisfy the conditions that insure a period of P = m-1. Further, specify a seed, $X_0 = 123,457$.
3. Let m = 47, a = 19, and X0 = 46, and generate a sequence c random integers. Find first 4 random numbers generate using any suitable method?
4. Let m = 100, a = 19, c = 6, and X0 = 63, and generate a sequence c random integers. Find first 5 random numbers generate using any suitable method?
5. Find random number, with first two random numbers is 10 and 15 respectively with modulus 50.

<u>Mid Square Method:</u>

This method was proposed by Van Neumann. In this method, we have a seed and then the seed is squared and its midterm is fetched as the random number.

Step 1: Take a **m** digits seed number $\mathbf{N_0}$

Step 2: Get square of seen number $\mathbf{N_0^2 = N_0 * N_0}$

Step 3: Square should be double digits of seed i.e. **2m**, if it is not then add zeros at the beginning to make double

Step 4: Take m digits from middle as a next random number. Then make this as a seed for next number.

Step 5: Go to Step 2

Step 6: Repeat these steps until same number chain occurs

Consider we have a seed having **N** digits we square that number to get a **2N** digits number if it doesn't become **2N** digits we add zeros before the number to make it **2N** digits.

This algorithm is basically the one which does not depend on the seed and the period should also be maximally long that it should almost touch every number in its range before it starts repeating itself as a rule of thumb remember that longer the period more random is the number.

Example:

Consider the seed to be 14 and we want a two digit random number.

Number --> Square --> Mid-term

14 --> 0**19**6 --> 19

19 --> 0**36**1 --> 36

36 --> 1**29**6 --> 29

29 --> 0**84**1 --> 84

84 --> 7**05**6 --> 05

05 --> 0**02**5 --> 02

02 --> 0**00**4 --> 00

00 --> 0**00**0 --> 00

In the above example, we can notice that we get some random numbers ***19,36,29,84,05,02,00*** which seem to be random picks, in this way we get multiple random numbers until we encounter a self-repeating chain.

We also get to know a disadvantage of this method that is if we encounter a **0** then we get a chain of **0s** from that point.

Also, consider that we get a random number **50** the square will be **2500** and the midterms are **50** again, and we get into this chain of **50**, and sometimes we may encounter such chains more often which acts as a disadvantage and because of these disadvantages this method is not practically used for generating random numbers.

Random Variate Generation (Non Uniform Random Number Generation)

A random variable is a measurable mapping having some distribution, and a random Variate is just a member of the co-domain of a random variable. A random Variate is a particular outcome of a random variable. Random Variates are the samples generated from a known distribution i.e. Random Variable and Random Variates have an inverse relationship.

Suppose ***X*** is a random variable which stands for the outcome of tossing a fair dice. So ***X*** can take value from **1** through **6** with equal probability of **1/6**. Now you actually toss a dice and get a number **4**. This number is a particular outcome of ***X***, and thus a random Variate. If you toss again, you may get another different value.

There are two popular methods for generate Random Variate:

- Inverse Transform Method
- Acceptance/Rejection Method

Inverse Transform Method(Non Uniform Transformation Method)

The inverse transform technique can be used to sample from the:

a) Exponential Distribution
b) Uniform Distribution
c) Triangular distribution

By inversing the Cumulative Distribution Function (CDF) of those probability distributions, the inverse transform technique can be utilized for any distribution when the CDF, F(x), is of a form that its inverse, F^{-1} can be computed easily.

a) Exponential Distribution

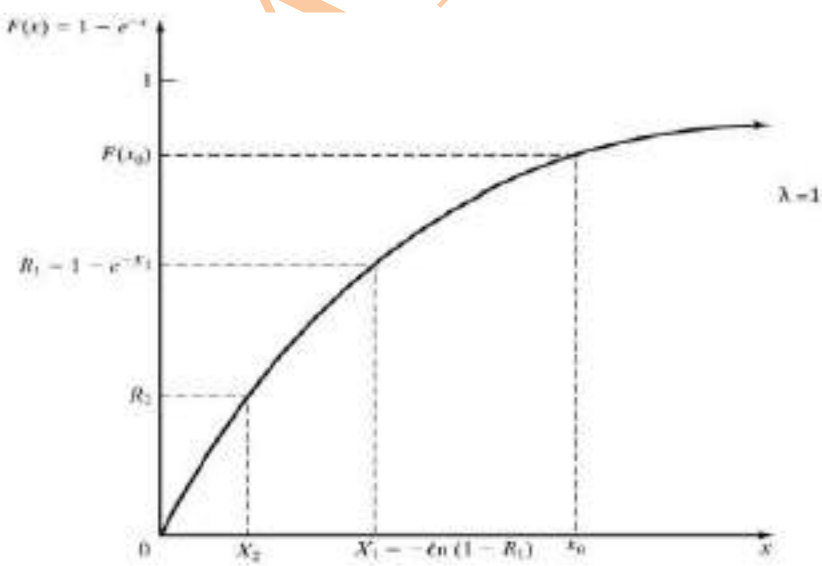
The exponential distribution has the Probability Distribution Function (PDF):

$$f(x) = \begin{cases} \lambda e^{-\lambda x}, & x \geq 0 \\ 0, & x < 0 \end{cases}$$

And the Cumulative Distribution Function (CDF)

$$F(x) = \int_{-\infty}^{x} f(t)dt = \begin{cases} 1 - e^{-\lambda x}, & x \geq 0 \\ 0, & x < 0 \end{cases}$$

λ is the number of occurrences per unit time.



The random Variate generation process is summarized in following steps:

Step 1: Compute the CDF of the random variable X for exponential distribution.

Step 2: Set F(X) = R on the range of X i.e. $1- e^{-\lambda X}=R$

Step 3: Solve the equation $1- e^{-\lambda X}=R$ in terms of X.

$1- e^{-\lambda X}=R$

$e^{-\lambda X}=1-R$

$-\lambda X= \ln(1- R)$

$$X = -\frac{1}{\lambda}\ln(1- R) \quad \text{…………………….. Eq-1}$$

Eq-1 is called random Variate generator for the exponential distribution. In general Eq-1 is written as

$X = F^{-1}(R)$

Example:

Generation of Exponential Variates X_i with mean 1 (i.e. λ=1), given random numbers R_i

i	1	2	3	4	5
R_i	0.1306	0.0422	0.6597	0.7965	0.7696

Solution:

$$R_1 = 1 - e^{-\lambda X}$$

$$X_1 = -\frac{1}{\lambda}\ln(1 - R_1)$$

$$X_1 = -\ln(1 - R_1) \qquad \text{(since } \lambda = 1)$$

X_1 = -ln(1-0.1306)

= -ln(0.8694)

X_1= 0.1400

X_2 = -ln(1-0.0422)

= -ln(0.9578)

X_2= 0.0431

X_3 = -ln(1-0.6597)

= -ln(0.3403)

X_3= 1.0779

X_4 = -ln(1-0.7965)

= -ln(0.2035)

X_4= 1.5920

X_5 = -ln(1-0.7696)

= -ln(0.2304)

X_5 = 1.4679

Then the veriats of given Random numbers are:

X_1	X_2	X_3	X_4	X_5
0.1400	**0.0431**	**1.0779**	**1.5920**	**1.4679**

Assignment:

1. Generation of Exponential Variates X_i with number of occurrence per unit time is 3, given random numbers R_i

i	1	2	3	4	5	6	7	8	9
Ri	0.32	0.2	0.01	0.121	0.55	0.11	0.4	0.7	0.22

b) Uniform Distribution:

The PDF for X in uniform distribution is given by

$$f(x) = \begin{cases} \dfrac{1}{b-a}, & a \le x \le b \\ 0, & otherwise \end{cases}$$

And the CDF is given by

$$F(x) = \begin{cases} 0, & x < a \\ \dfrac{x-a}{b-a}, & a \le x \le b \\ 1, & x > b \end{cases}$$

Now, set $F(X) = \frac{X-a}{b-a} = R$

$X = a + (b - a) R$ ………………… Eq-2

Eq-2 is the random Variate generator for the uniform distribution.

Assignment:

1. Generation of Uniform Distribution X_i with number of given random numbers R_i , where a=0.2, b=0.9

i	1	2	3	4	5	6	7	8	9
Ri	0.261	0.172	0.11	0.721	0.745	0.172	0.574	0.771	0.172

c) Triangular Distribution

Consider a random variable X that has PDF

$$f(x) = \begin{cases} x, & 0 \le x \le 1 \\ 2-x, & 1 \le x \le 2 \\ 0, & otherwise \end{cases}$$

This distribution is called triangular distribution with endpoints (0, 2) and mode at 1. Its CDF is given by

$$F(x) = \begin{cases} 0, & x \le 0 \\ \frac{x^2}{2}, & 0 \le x \le 1 \\ 1 - \frac{(2-x)^2}{2}, & 1 < x \le 2 \\ 1, & x > 2 \end{cases}$$

For $0 < X \le 1$,

$R = X^2/2$

$\therefore X = \sqrt{2R}$

For $1 \le X \le 2$,

$R = 1 - (2 - X)^2/2$

$\therefore X = 2 - \sqrt{2(1-R)}$

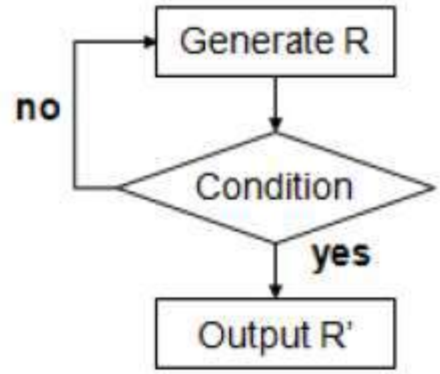
Assignment:

1. Generation of Triangular Distribution R_i with number of given random numbers X_i

i	1	2	3	4	5	6	7	8	9
xi	0.1	0.2	0.01	1.31	0.435	0.512	0.514	1.173	2.122

Acceptance/Rejection Method:

The Acceptance-Rejection method is a popular technique used in random variate generation to generate random numbers from a specified probability distribution that might be complex or hard to sample directly. This method involves using a simpler and easily samplable distribution to generate random samples that are then adjusted to match the desired distribution using acceptance or rejection criteria.



Use following steps to generate uniformly distributed random numbers between ***1/4 and 1.***

Step 1: Generate a random number *R~ U[0,1]*

Step 2:

a:

If $R \geq 1/4$, accept $X = R$, goto Step 3

b:

If $R < 1/4$, reject R, return to Step 1

Step 3: If another uniform random variate on [1/4, 1] is needed, repeat the procedure beginning at Step 1, Otherwise stop.

The efficiency:

Use this method in this particular example, the rejection probability is 1/4 on the average for each number generated. The number of rejections is a geometrically distributed random variable with probability of ``success'' being *p = 3/4*, mean number of rejections is *(1/p - 1) = 4/3 - 1 = 1/3* (i.e. 1/3 waste).

For this reason, the inverse transform (X = 1/4 + (3/4) R) is more efficient method.

Test for Randomness:

The desirable properties of random numbers are ***uniformity*** and ***independence*** to ensure that these desirable properties are achieved, a number of tests can be performed (fortunately, the appropriate tests have already been conducted for most commercial simulation software). The tests can be placed in two categories according to the properties of interest.

1. Testing for uniformity
2. Testing for independence.

Uniformity Test:

The testing for uniformity can be achieved through different frequency test. These tests use the Kolmogorov-Smirnov (K-S) or the Chi-Square test to compare the distribution of the set of numbers generated to a uniform distribution. Hence in this category we will discuss two types of test

- Kolmogorov-Smirnov (K-S) Test
- Chi-Square Test

In testing for uniformity, the hypotheses are as follows:

Null hypotheses (H_0): R_i ~U/[0, 1]

Alternative hypothesis (H_1): R_i !~U/[0, l]

Here U is the uniformity

The null hypothesis, H_0 reads that the numbers are distributed uniformly on the interval [0, 1] and alternative hypothesis H_1 reads that the numbers are distributed not uniformly on the interval [0, 1]. Failure to reject the null hypothesis means that no evidence of non-uniformity has been detected on the basis of this test. This does not imply that further testing of the generator for uniformity is unnecessary.

For each test, a level of significance (α) must be stated. The level α is the probability of rejecting the null hypothesis given that the null hypothesis is true, or

$\alpha = P(\text{reject } H_0 \mid H_0 \text{ true})$

The decision maker sets the value of α for any test. Frequently, α is set to 0.01 or 0.05.

The Kolmogorov-Smirnov (K-S) Test:

This test compares the continuous cumulative distribution function F(x), of the uniform distribution to the empirical cumulative distribution function $S_N(x)$, of the sample of N observations.

By definition,

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

If the sample from the random-number generator is R_1 R_2, • • •, R_N, then the empirical cumulative distribution function, $S_N(x)$, is defined by

$$S_N(x) = \frac{(\text{Number of } R_1\, R_2, \bullet\bullet\bullet, R_n \text{ which are} <= x)}{N}$$

Where N is the number of observation, N becomes larger $S_N(x)$ should become a better approximation to F(x), provided that the null hypothesis is true.

The **Kolmogorov-Smirnov (K-S) test** is based on the largest absolute deviation or difference between F(x) and $S_N(x)$ over the range of the random variable i.e. it is based on the statistic

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

For testing against a uniform cumulative distribution function, the test procedure follows these steps:

Algorithm for K-S Test:

Step 1: Define the hypothesis for testing uniformity

H_0 : R_i ~ Uniformity

H_1 : R_i !~ Uniformity

Step 2: Sort the data from smallest to largest. Let $R_{(i)}$ denoted the i th smallest observation, so that $R_{(1)} <= R_{(2)}$ ……..$<=R_{(N)}$

Step 2: Compute

$$D_{+} = \max\left\{\frac{i}{N} - R_i\right\} \quad D^{-} = \max\left\{R_i - \frac{(i-1)}{N}\right\}$$

Step 4: Compute

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

Step 5: Determine the critical value, D_{α} from K-S table for Critical Values for the specified significance level α and the given sample size N.

Step 6:

If $D <= D_{\alpha}$:

The null hypothesis (H_0) is accepted, i.e. the numbers are uniformly distributed.

Else

The null hypothesis (H_0) is rejecter, i.e. the numbers are not uniformly distributed.

K-S table for Critical Values

n \ α	0.01	0.05	0.1	0.15	0.2
1	0.995	0.975	0.950	0.925	0.900
2	0.929	0.842	0.776	0.726	0.684
3	0.828	0.708	0.642	0.597	0.565
4	0.733	0.624	0.564	0.525	0.494
5	0.669	0.565	0.510	0.474	0.446
6	0.618	0.521	0.470	0.436	0.410
7	0.577	0.486	0.438	0.405	0.381
8	0.543	0.457	0.411	0.381	0.358
9	0.514	0.432	0.388	0.360	0.339
10	0.490	0.410	0.368	0.342	0.322
11	0.468	0.391	0.352	0.326	0.307
12	0.450	0.375	0.338	0.313	0.295
13	0.433	0.361	0.325	0.302	0.284
14	0.418	0.349	0.314	0.292	0.274
15	0.404	0.338	0.304	0.283	0.266
16	0.392	0.328	0.295	0.274	0.258
17	0.381	0.318	0.286	0.266	0.250
18	0.371	0.309	0.278	0.259	0.244
19	0.363	0.301	0.272	0.252	0.237
20	0.356	0.294	0.264	0.246	0.231
25	0.320	0.270	0.240	0.220	0.210
30	0.290	0.240	0.220	0.200	0.190
35	0.270	0.230	0.210	0.190	0.180
40	0.250	0.210	0.190	0.180	0.170
45	0.240	0.200	0.180	0.170	0.160
50	0.230	0.190	0.170	0.160	0.150
OVER 50	$\frac{1.63}{\sqrt{n}}$	$\frac{1.36}{\sqrt{n}}$	$\frac{1.22}{\sqrt{n}}$	$\frac{1.14}{\sqrt{n}}$	$\frac{1.07}{\sqrt{n}}$

Example: Suppose that the five random numbers 0.44, 0.81, 0.14, 0.05, 0.93 were generated, and it is desire to perform a test for uniformity using the K-S test with a level of significance α of 0.05.

Solution:

Define hypothesis for testing uniformity

H_0 : R_i ~ uniformity

H_1 : R_i !~ uniformity

First, the numbers must be sorted from smallest to largest, i.e. the given random numbers are 0.05, 0.14, 0.44, 0.81, 0.93

N = 5

Now calculate the D^+ and D^- as:

I	1	2	3	4	5
R_i	0.05	0.14	0.44	0.81	0.93
i/N	1/5=0.2	2/5=0.4	3/5=0.6	4/5=0.8	5/5=1.0
(i-1)/N	(1-1)/5=0	(2-1)/5=0.2	(3-1)/5=0.4	(4-1)/5=0.6	(5-1)/5=0.8
$D^+ = (i/N)-R_i$	0.15	0.26	0.16	**-0.01**	0.07
$D^- = R_i - ((i-1)/N)$	0.05	**-0.06**	0.04	0.21	0.13

New, max $\{D^+\} = 0.26$ max $\{D^-\} = 0.21$

$D = \max\{D^+, D^-\} = \max\{0.26, 0.21\}$

D = 0.26

$D_{0.05}$ at N=5 = 0.565 (See in K-S Table for Critical Values)

$D_{0.05}$ = 0.565

Here $D < D_{0.05}$ hence null hypothesis is accepted

This shows that according to KS testing given numbers are uniformly distributed.

Assignment

1. Suppose that the five numbers 0.24, 0.80, 0.11, 0.05, 0.93 were generated, and it is desired to perform a test for uniformity using the Kolmogorov-Smirnov test with a level of significance α of 0.01.
2. Suppose that the four numbers 0.80, 0.14, 0.05, 0.5 were generated, and it is desired to perform a test for uniformity using the Kolmogorov-Smirnov test with a level of significance α of 0.10.
3. Suppose that the seven numbers 0.44, 0.81, 0.14, 0.05, 0.93, 0.01, 0.02 were generated, and it is desired to perform a test for uniformity using the Kolmogorov-Smirnov test with a level of significance α of 0.05.

Chi-Square Test:

Note: It is pronounce as Kaai Square Test

A chi-squared test (symbolically represented as X^2) is basically a data analysis on the basis of observations of a random set of variables. Usually, it is a comparison of two statistical data sets. This test was introduced by Karl Pearson in 1900 for categorical data analysis and distribution. So it was mentioned as Pearson's chi-squared test.

The chi-square test is used to estimate how likely the observations that are made would be, by considering the assumption of the null hypothesis as true.

A hypothesis is a consideration that a given condition or statement might be true, which we can test afterwards. Chi-squared tests are usually created from a sum of squared falsities or errors over the sample variance.

Chi-Square test uses the sample statistics as:

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

Where,

$\mathbf{O_i}$ is the observed number in the i-th class,

$\mathbf{E_i}$ is the expected number in the i-th class,

n is the number of classes.

For the uniform distribution, $\mathbf{E_i}$ the expected number in each class is given by:

$\mathbf{E_i = N/n}$

For equally spaced classes, **N** is the total number of observations. It can be shown that the sampling distribution of $\mathbf{X_0^2}$ is approximately the chi-square distribution with **n-1** degrees of freedom.

Algorithm for Chi-Square Test:

Step 1: Define the hypothesis for testing uniformity

$H_0 : R_i \sim$ Uniformity

$H_1 : R_i$!~ Uniformity

Step 2: Determine Order Statistics

$R_1 <= R_2 <= \ldots\ldots R_n$

Step 3: Divide Range $R_n - R_1$ in **n** equidistant intervals $[a_i, b_i]$, such that each interval has at least 5 observations.

Step 4: Calculate for i = 1,2,3.....,N

$O_i = N.\{S_N(b_i) - S_N(a_i)\}$,

$E_i = N.\{F(b_i) - F(a_i)\}$

Step 5: Calculate

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

Step 6: Determine the critical value of $X^2_{\alpha,n-1}$ from chi-square table

Step 7:

If $X^2 <= X^2_{\alpha,n-1}$

Null hypothesis is accepted, that means random numbers are uniformly distributed

Else

Null hypothesis is rejected, that means random numbers are not uniformly distributed

Percentage Points of the Chi-Square Distribution

Degrees of Freedom	Probability of a larger value of x^2								
	0.99	0.95	0.90	0.75	0.50	0.25	0.10	0.05	0.01
1	0.000	0.004	0.016	0.102	0.455	1.32	2.71	3.84	6.63
2	0.020	0.103	0.211	0.575	1.386	2.77	4.61	5.99	9.21
3	0.115	0.352	0.584	1.212	2.366	4.11	6.25	7.81	11.34
4	0.297	0.711	1.064	1.923	3.357	5.39	7.78	9.49	13.28
5	0.554	1.145	1.610	2.675	4.351	6.63	9.24	11.07	15.09
6	0.872	1.635	2.204	3.455	5.348	7.84	10.64	12.59	16.81
7	1.239	2.167	2.833	4.255	6.346	9.04	12.02	14.07	18.48
8	1.647	2.733	3.490	5.071	7.344	10.22	13.36	15.51	20.09
9	2.088	3.325	4.168	5.899	8.343	11.39	14.68	16.92	21.67
10	2.558	3.940	4.865	6.737	9.342	12.55	15.99	18.31	23.21
11	3.053	4.575	5.578	7.584	10.341	13.70	17.28	19.68	24.72
12	3.571	5.226	6.304	8.438	11.340	14.85	18.55	21.03	26.22
13	4.107	5.892	7.042	9.299	12.340	15.98	19.81	22.36	27.69
14	4.660	6.571	7.790	10.165	13.339	17.12	21.06	23.68	29.14
15	5.229	7.261	8.547	11.037	14.339	18.25	22.31	25.00	30.58
16	5.812	7.962	9.312	11.912	15.338	19.37	23.54	26.30	32.00
17	6.408	8.672	10.085	12.792	16.338	20.49	24.77	27.59	33.41
18	7.015	9.390	10.865	13.675	17.338	21.60	25.99	28.87	34.80
19	7.633	10.117	11.651	14.562	18.338	22.72	27.20	30.14	36.19
20	8.260	10.851	12.443	15.452	19.337	23.83	28.41	31.41	37.57
22	9.542	12.338	14.041	17.240	21.337	26.04	30.81	33.92	40.29
24	10.856	13.848	15.659	19.037	23.337	28.24	33.20	36.42	42.98
26	12.198	15.379	17.292	20.843	25.336	30.43	35.56	38.89	45.64
28	13.565	16.928	18.939	22.657	27.336	32.62	37.92	41.34	48.28
30	14.953	18.493	20.599	24.478	29.336	34.80	40.26	43.77	50.89
40	22.164	26.509	29.051	33.660	39.335	45.62	51.80	55.76	63.69
50	27.707	34.764	37.689	42.942	49.335	56.33	63.17	67.50	76.15
60	37.485	43.188	46.459	52.294	59.335	66.98	74.40	79.08	88.38

Example 1: Use the chi-square test with $\alpha = 0.05$ to test whether the data shown below are uniformly distributed.

0.34 0.83 0.96 0.47 0.79 0.99 0.37 0.72 0.06 0.18

0.90 0.76 0.99 0.30 0.71 0.17 0.51 0.43 0.39 0.26

0.25 0.79 0.77 0.17 0.23 0.99 0.54 0.56 0.84 0.97

0.89 0.64 0.67 0.82 0.19 0.46 0.01 0.97 0.24 0.88

0.87 0.70 0.56 0.56 0.82 0.05 0.81 0.30 0.40 0.64

0.44	0.81	0.41	0.05	0.93	0.66	0.28	0.94	0.64	0.47
0.12	0.94	0.52	0.45	0.65	0.10	0.69	0.96	0.40	0.60
0.21	0.74	0.73	0.31	0.37	0.42	0.34	0.58	0.19	0.11
0.46	0.22	0.99	0.78	0.39	0.18	0.75	0.73	0.79	0.29
0.67	0.74	0.02	0.05	0.42	0.49	0.49	0.05	0.62	0.78

Solution:

Here given,

$\alpha = 0.05$, N = 100

Now, we can make 10 classes with equal length of 10 to each class

i.e. n = 10

Define hypothesis for testing uniformity

H_0 : R_i ~ uniformity

H_1 : R_i !~ uniformity

The table for Chi-Square statistics is

Class Interval (i)	**O_i**	**E_i (N/n)**	**$(O_i - E_i)$**	**$(O_i - E_i)^2$**	**$X^2 = (O_i - E_i)^2/E_i$**
1 (0.0, 0.1)	8	10	-2	4	0.4
2 (0.1, 0.2)	8	10	-2	4	0.4
3 (0.2, 0.3)	10	10	0	0	0
4 (0.3, 0.4)	9	10	-1	1	0.1
5 (0.4, 0.5)	12	10	2	4	0.4
6 (0.5, 0.6)	8	10	-2	4	0.4
7 (0.6, 0.7)	10	10	0	0	0
8 (0.7, 0.8)	14	10	4	16	1.6
9 (0.8, 0.9)	10	10	0	0	0
10 (0.9, 1.0)	11	10	1	1	0.1
Total	**N=100**	**N=100**			**$\sum X^2 = 3.4$**

Note: *The value of O_i at interval 1 (0.0, 0.1) is total data between range greater than 0.0 and less than or equals to 0.1 i.e. (0.0 < data <= 0.1) there are 8 data presented in this range. Likewise for all the value of O_i for all intervals are calculated.*

The value of X^2 = **3.4**

Now, find the value of Chi-Square i.e. $X^2_{\alpha, n-1}$ from table.

Here degree of freedom is n-1=10-1=9 and α=0.05.

The tabulated value of $X^2_{0.05, 9}$ =**16.9**

Here, $X^2 < X^2_{0.05, 9}$ hence null hypothesis is accepted.

This shows that according to Chi-Square test the given numbers are uniformly distributed.

Example 2: Use the chi-square test with α = 0.05 to test whether the data shown below are uniformly distributed. In first ranges there are 15 random number , in second there are 5 random number, in 3rd there are 10 random number, in fourth there are 10 and in 5th there are 20 random number.

Solution:

Here given,

Total number of classes (n) = 5

Total random numbers (N) = 15+5+10+10+20 = 60

Define hypothesis for testing uniformity

H_0 : R_i ~ uniformity

H_1 : R_i !~ uniformity

The table for Chi-Square statistics is

Class Interval (i)	**O_i**	**E_i (N/n)**	**$(O_i - E_i)$**	**$(O_i - E_i)^2$**	**$X^2 = (O_i - E_i)^2/E_i$**
1	15	12	3	9	9/12=0.75
2	5	12	-7	49	49/12=4.08
3	10	12	-2	4	4/12=0.33
4	10	12	-2	4	4/12=0.33
5	20	12	8	64	64/12=5.33
Total	**N=60**	**N=60**			**$\sum X^2 = 10.82$**

Value of X^2 = **10.82**

Now, find the value of Chi-Square i.e. $X^2_{\alpha, n-1}$ from table.

Here degree of freedom is n-1=5-1= 4 and α=0.05.

The tabulated value of $X^2_{0.05, 4}$ = **9.49**

Here, $X^2 > X^2_{0.05, 4}$ hence null hypothesis is rejected.

This shows that according to Chi-Square test the given numbers are not uniformly distributed.

Conclusion:

Both the Kolmogorov-Smirnov and the chi-square test are acceptable for testing the uniformity of a sample of data, provided that the sample size is large. However, the Kolmogorov-Smirnov test is the more powerful of the two and is recommended. Furthermore, the Kolmogorov-Smirnov test can be applied to small sample sizes, whereas the chi-square is valid only for large samples, say N>=50.

Assignment:

1. Use the chi-square test with α = 0.99 to test whether the data shown below are uniformly distributed. In first ranges there are 10 random number, in second there are 10 random number, in 3rd there are 15 random number, in fourth there are 15 , in 5th there are 5 random number. And then 6th, 7th and 8th has 10, 5, 10 random numbers.
2. Use the chi-square test with a = 0.25 to test whether the data shown below are uniformly distributed. In first ranges there are 20 random number , in second there are 9 random number, in 3rd there are 15 random number, in fourth there are 15 and in 5th there are 13 random number.

K-S Test V/s Chi-Square Test:

K-S test	Chi-Square Test
Small samples	Large sample
Continuous Distribution	Discrete Distribution
Difference between Observed and Expected cumulative probabilities (CDF)	Differences between observed and hypothesized probabilities (PDFs or PMFs).
Uses each observation in the sample without any grouping => makes a better use of the data Cell size is not a problem	Group observation into a small number of cells =>Cell sizes effect the conclusion but no firm guidelines
Exact	Approximate

Independence Testing:

When a random number generator is devised, one needs to test its property. The two properties we are concerned most are ***uniformity*** and ***independence***.

An independence test of random numbers refers to a statistical test used to determine whether a sequence of random numbers generated by a random number generator (RNG) exhibits a significant level of independence between consecutive numbers. In other words, the test aims to assess whether the generated sequence of numbers is truly random or if there are patterns, correlations, or dependencies present that might indicate a lack of randomness.

There are three types of popular testing for independence as:

1. **Autocorrelation Test:** It tests the correlation between numbers and compares the sample correlation to the expected correlation of zero.
2. **Gap Test:** It counts the number of digits that appear between repetitions of particular digit and then uses the Kolmogorov-Smirnov test to compare with the expected size of gaps,
3. **Poker Test:** It treats numbers grouped together as a poker hand. Then the hands obtained are compared to what is expected using the chi-square test.

Autocorrelation Test:

Autocorrelation Test is a statistical test that determines whether a random number generator is producing independent random number in a sequence. The test for the auto correlation is concerned with the dependence between numbers in a sequence. The test computes the auto correlation between every **m** numbers (**m** is also known as lag) starting with i^{th} index.

The variables involved in this test are:

- **m** is the lag, the space between the number being tested.
- **i** is the index or number from we start the testing.
- **N** is the total number of random numbers generated.
- **M** is the largest integer such that $i+(M+1)m<=N$

Algorithm for Autocorrelation Test:

Step 1: Define the hypothesis for testing independence

$H_0 : R_i \sim$ independence

$H_1 : R_i \, !\sim$ independence

Step 2: Now the autocorrelation $\boldsymbol{\rho_{im}}$ between R_i, R_{i+m}, R_{i+2m},…….$R_{i+(M+1)m}$ is computed as:

$$\rho_{im} = \frac{1}{M+1}\left[\sum_{k=0}^{M} R_{i+km}R_{i+(k+1)m}\right] - 0.25$$

Step 3: Now, calculate the test statistics as:

$$Z_0 = \frac{\rho_{im}}{\sigma_{im}}$$

Where,

$$\sigma_{im} = \frac{\sqrt{13M+7}}{12(M+1)}$$

Step 4: find the critical value of $Z_{\alpha/2}$ from Z table

Step 5:

If $Z_0 <= Z_{\alpha/2}$,

Null hypothesis (H_0) is accepted, that means random number are independent

Else

Null hypothesis (H_0) is rejected, that means random number are not independence.

Example: Test whether the 3rd, 8th, 13th, and so on, numbers in the sequence at the beginning of this section are auto-correlated. (Use $\alpha = 0.05$.) Here, i = 3 (beginning with the third number), m = 5 (every five numbers), N = 30 (30 numbers in the sequence). Use $Z_{0.025} = 1.96$

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:

Define hypothesis for testing independence

$H_0 : R_i \sim$ independence

$H_1 : R_i\ !\sim$ independence

First we calculate the value of M using the condition

i + (M+1)m <= N

Since, i=3, m=5, and N=30 we have

3 + (M +1)5 <= 30

3 + 5M + 5 <=30

5M <= 22

M <= 22/5

M <= 4

Hence, M=4

Then,

$$\rho_{im} = \frac{1}{M+1}\left[\sum_{k=0}^{M} R_{i+km}R_{i+(k+1)m}\right] - 0.25$$

$\rho_{35} = 1/(4+1)[R_{3+0*5}\ R_{3+(0+1)5} + R_{3+1*5}\ R_{3+(1+1)5} + R_{3+2*5}\ R_{3+(2+1)5} + R_{3+3*5}\ R_{3+(3+1)5} + R_{3+4*5}\ R_{3+(4+1)5}] - 0.25$

$\rho_{35} = 1/(4+1)[R_3\ R_8 + R_8\ R_{13} + R_{13}\ R_{18} + R_{18}\ R_{23} + R_{23}\ R_{28}] - 0.25$

$\rho_{35} = 1/(4+1)[(0.23)(0.28) + (0.28)(0.33) + (0.33)(0.27) + (0.27)(0.05) + (0.05)(0.36)] - 0.25$

$\boldsymbol{\rho_{35} = -0.1945}$

Now,

$$\sigma_{im} = \frac{\sqrt{13M + 7}}{12(M+1)}$$

$\sigma_{35} = \sqrt{}\ (13(4) + 7) / 12(4 + 1)$

$\boldsymbol{\sigma_{35} = 0.1280}$

Now the test statistics is:

$$Z_0 = \frac{\rho_{im}}{\sigma_{im}}$$

$Z_0 = -0.1945/0.1280$

$\boldsymbol{Z_0 = -1.5195}$

Now, the critical value of $Z_{\alpha/2}$ is

$Z_{\alpha/2} = Z_{0.05/2} = Z_{0.025}$

$\boldsymbol{Z_{0.025} = 1.96}$

Here $Z_0 < Z_{0.025}$ hence null hypothesis is accepted

This shows that according to autocorrelation testing given data is independent.

Gap Test:

The gap test is used to determine the significance of the interval between the recurrences of the same digit. A gap of length **x** occurs between the recurrences of some specified digit.

The following example illustrates the length of gaps associated with the digit 3:

4, 1, **3**, 5, 1, 7, 2, 8, 2, 0, 7, 9, 1, **3**, 5, 2, 7, 9, 4, 1, 6, **3** , **3**, 9, 6, **3**, 4, 8, 2, **3**, 1, 9, 4, 4, 6, 8, 4, 1, **3**.

There are 7 three's are there. Thus only six gaps can occur. The first gap is of length 10 and second gap of length 7, Third gap of length 0, fourth gap of length 2, fifth gap of length 3 and sixth gap of length 8. Similarly, the gap associated with other digits can be calculated.

The probability of a particular gap length can be determined by:

P(gap of **n**) = P(**n** followed by exactly **x** non-n digits)

$$P(\text{gap of } n) = P(x \neq 3)P(x \neq 3)...P(x \neq 3)P(x = 3)$$

Where, n is the specified digit, x is the non-n digits. In above equation we consider the n = 3

If we are only concerned with digits between 0 and 9, then

$$P(\text{gap of } n) = 0.9^n 0.1$$

The **Theoretical Frequency Distribution F(x)** for randomly ordered digits is given by

$$P(gap \leq x) = F(x) = 0.1 \sum_{n=0}^{x} (0.9)^n = 1 - 0.9^{x+1}$$

Algorithm for Gap Test:

The procedure for the test follows the steps below. When applying the test to random numbers, class intervals such as [0, 0.1), [0.1, 0.2),. . . play the role of random digits.

Step 1: Define the hypothesis for testing independence

$H_0 : R_i \sim$ independence

$H_1 : R_i$!~ independence

Step 2: Specify the CDF for the Theoretical Frequency Distribution based on the selected class interval width.

$$P(gap \leq x) = F(x) = 0.1\sum_{n=0}^{x}(0.9)^n = 1 - 0.9^{x+1}$$

$\mathbf{F(x) = 1\text{-}0.9^{x+1}}$

Where **x** is the upper limit of gap length

Step3: Arrange the observed sample of gaps in a cumulative distribution with these same classes.

Step 4: Find D, the maximum deviation between F(x) and $S_N(x)$ as in K-S test.

Step 5: Determine the critical value, D_α, from Table for the specified value of α and the sample size N.

Step 6:

If $D < D_\alpha$

Null hypothesis is accepted, that means random numbers are independently distributed

Else

Null hypothesis is rejected, that means random numbers are not independently distributed

Example: Based on the frequency with which gaps occurs, analyze the 110 digits below to test whether they are independent. Use $\alpha = 0.05$.

4, 1, 3, 5, 1, 7, 2, 8, 2, 0, 7, 9, 1, 3, 5, 2, 7, 9, 4, 1, 6, 3, 3, 9, 6, 3, 4, 8, 2, 3, 1, 9, 4, 4, 6, 8, 4, 1, 3, 8, 9, 5, 5, 7, 3, 9, 5, 9, 8, 5, 3, 2, 2, 3, 7, 4, 7, 0, 3, 6, 3, 5, 9, 9, 5, 5 5, 0, 4, 6, 8, 0, 4, 7, 0, 3, 3, 0, 9, 5, 7, 9, 5, 1, 6, 6, 3, 8, 8, 8, 9, 2, 9, 1, 8, 5, 4, 4, 5, 0, 2, 3, 9, 7, 1, 2, 0, 3, 6, 3

Solution:

Define hypothesis for testing independence

H_0 : R_i ~ independence

H_1 : R_i !~ independence

The number of gaps is given by the number of data values minus the number of distinct digits 10 (0-9), i.e. 110-10 = 100 in the example. The numbers of gaps associated with the various digits are as follows:

Digits	Length of Gaps	Number of Gaps
0	47, 9, 3, 2, 2, 21, 6	7
1	2, 7, 6, 10, 6, 45, 9, 10	8
2	1, 6, 12, 22, 0, 38, 8, 4	8
3	10, 7, 0, 2, 3, 8, 5, 5, 2, 4, 1, 14, 0, 9, 14, 5,1	17
4	17,7, 5, 0, 2, 18, 12, 3, 23, 0	10
5	10, 26, 0, 3, 2, 11, 2, 0, 0, 12, 2, 12, 2	13
6	3, 9, 24, 9, 14, 0, 22	7
7	4, 5, 26, 10, 1, 16, 6, 22	8
8	19, 7, 3, 8, 21, 16, 0, 0, 4	9
9	5, 5, 7, 8, 4, 1, 14, 0, 14, 2, 8, 1, 9,	13

The Calculation for Gap Test is shown in the following table:

We can take any gap length, here we take gap length = **0-3**

Frequency is the number of occurrence of digits with in gap length, i.e. if gap length is 0-3 then frequency is the total occurrence of digits 0, 1, 2, and 3.

Gap Length	Frequency	Relative Frequency	Cum. Frequency S(X)	Theoretical Frequency F(X)	$\lvert F(x) - S_N(x)\rvert$
0-3	35	0.35	0.35	0.3439	0.0061
4-7	22	0.22	0.57	0.5695	0.0005
8-11	17	0.17	0.74	0.7176	**0.0224**
12-15	9	0.09	0.83	0.8147	0.0153
16-19	5	0.05	0.88	0.8784	0.0016
20-23	6	0.06	0.94	0.9202	0.0198
24-27	3	0.03	0.97	0.9497	0.0223
28-31	0	0.00	0.97	0.9657	0.0043
32-35	0	0.00	0.97	0.9775	0.0075
36-39	2	0.02	0.99	0.9852	0.0043
40-43	0	0.00	0.99	0.9903	0.0003
44-47	1	0.01	1.00	0.9936	0.0064

$F(x) = 1-0.9^{x+1}$

Where x is upper boundary of gap length i.e. 3, 7, 11, 15 ……. 43, 47

Now, Calculate the value of D

$D = \max \{F(x)-S_N(x)\}$

D = 0.0224

Now, Find the value of D_α at N=100 from K-S table of critical value

$$D_{0.05} = \frac{1.36}{\sqrt{N}}$$

$$D_{0.05} = \frac{1.36}{\sqrt{100}}$$

$D_{0.05} = 0.136$

Here, $D < D_{0.05}$ hence null hypothesis is accepted

This shows that according to gap test given random numbers are independent.

Assignment:

1. Based on the frequency with which gaps occur, analyze the digits below to test whether they are independent. Use α = 0.05.

Gap Length	Frequency
0-4	25
5-9	15
10-14	10
15-19	3
20-24	2
25-29	0
30-34	5
35-39	10
40-44	30

2. Based on the frequency with which gaps occur, analyze the digits below to test whether they are independent. Use α = 0.2.

Gap Length	Frequency
0-5	15
6-10	20
11-15	5
16-20	3
21-25	12
26-30	15

Poker Test:

The poker test for independence is based on the frequency with which certain digits are repeated in a series of numbers. This test not only tests for the randomness of the sequence of numbers, but also the digits comprising of each of the numbers.

The expected value of each of the combination of digits in a number is compared with the observed value by means of the chi-square test for independence. The acceptance is done if the observed value of ***chi- square*** sums for all the possible combinations of digits is less than the acceptable value for the given degree of freedom at the specified confidence interval.

- This test gets its name from a game of cards called poker
- This test not only tests the randomness of the sequence of numbers, but also the digits comprising of each number
- Every random number of five digits or every sequence of five digits is treated as poker hand.

The poker test uses the Chi-Square statistics to accept or reject the null hypothesis.

In 3-digit numbers there are only three possibilities, as follows:

1. The individual numbers can all be different.
2. The individual numbers can all be the same.
3. There can be one pair of like digits.

The probability associated with each of these possibilities is given by the following

P (three different digits) = P (second different from the first) x P (third different from the first and second)

= (0.9) (0.8)

P (three different digits) = 0.72

P (three like digits) = P (second digit same as the first) x P (third digit same as the first)

= (0.1) (0.1)

P (three like digits) = 0.01

P (exactly one pair) = 1 – [P(three different digits)+ P(three like digits)]

= 1 – [0.72 + 0.01]

P (exactly one pair) = 0.27

Algorithm for Poker Test:

Step 1: Define hypothesis for testing independence as:

H_0 : R_i ~ Independent

H_1 : R_i !~ Independent

Step 2: Generate frequency distribution for 3 combinations and apply Chi-Square test

Step 3: Compute test statistics as

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

Step 4: Determine for significant level α, $X^2_{\alpha,n-1}$

Step 5:

If $X^2 < X^2_{\alpha,n-1}$

Null hypothesis (H_0) is accepted that means random numbers are independence

Else

Null hypothesis (H_0) is rejected, that means random numbers are not independence

Example: A sequence of 1000 three-digit numbers has been generated and an analysis indicates that 680 have three different digits, 289 contain exactly one pair of like digits, and 31 contain three like digits. Based on the poker test, are these numbers independent? Let α = 0.05. Test these numbers using poker test for three digits.

Solution:

Define hypothesis for testing independence as:

$H_0 : R_i \sim$ Independent

$H_1 : R_i\ !\sim$ Independent

Now generate the frequency distribution table and apply chi-square test

Combination, i	Observed Frequency, (Oi)	Expected Frequency (Ei)	(Oi-Ei)	$(Oi-Ei)^2 / Ei$
Three Different digit	680	0.72X1000=720	-40	2.22
Three Like digit	31	0.01X1000=10	21	44.10
Exactly one pair	289	0.27X1000=270	19	1.33
	1000	1000		47.65

There are three conditions so, **n=3**

The test statistics

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

$\mathbf{X^2 = 47.65}$

Then, the critical value of $X^2_{\alpha, n-1}$ is

$\mathbf{X^2_{0.05, 3-1} = X^2_{0.05, 2} = 5.99}$

Here, $X^2 > X^2_{0.05,\,2}$ hence null hypothesis is rejected

This shows that according to Poker Test given random numbers are not independent.

Assignment:

1. The sequence of three-digit numbers has been generated and an analysis indicates that 300 have three different digits, 500 contain exactly one pair of like digits, and 200 contain three like digits. Based on the poker test, are these numbers independent? Let $\alpha = 0.05$. Test these numbers using poker test for three digits.

Poker test for 4-digit numbers:

In four digit number, there are five different possibilities

P (four different digits) = 4c4 x 10/10 x 9/10 x 8/10 x 7/10

P (four different digits) = 0.504

P (one pair) = 4c2 x10/10x1/10x9/10x8/10

P (one pair) = 0.432

P (two pair) = 4c2/2 x 10/10 x 1/10 x 9/10 x 1/10

P (two pair) = 0.027

P (three digits of a kind) = 4c3 x 10/10 x 1/10 x 1/10 x 9/10

P (three digits of a kind) = 0.036

P (four digits of a kind) = 4c4 x 10/10 x 1/10 x 1/10 x 1/10

P (four digits of a kind) = 0.001

Example: A sequence of 1000 four digit numbers has been generated and an analysis indicates the following combinations and frequencies.

Combination (I)	Observed frequency (Oi)
Four different digits	560
One pair	394
Two pair	32
Three digits of a kind	13
Four digit of a kind	1
	1000

Based on poker test, test whether these numbers are independent. Use α=0.05 and N=4 is 9.49.

Solution:

Define hypothesis for testing independence as:

$H_0 : R_i \sim$ Independent

$H_1 : R_i$!~ Independent

Now generate the frequency distribution table and apply chi-square test

Combination (i)	Observed frequency (Oi)	Expected frequency(Ei)	(oi-Ei)	$(oi-Ei)^2/Ei$
Four different digits	560	0.504x1000=504	56	6.22
One pair	394	0.432x1000=432	-38	3.343
Two pair	32	0.027x1000=27	5	0.926
Three digits of a kind	13	0.036	-23	14.694
Four digit of a kind	1	0.0001x1000=1	0	0.000
	1000	1000		25.185

There are five conditions so, **n=5**

The test statistics

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

$\mathbf{X^2 = 25.185}$

Then, the critical value of $X^2_{\alpha, n-1}$ is

$\mathbf{X^2_{0.05,\,5-1} = X^2_{0.05,\,4} = 9.49}$

Here, $X^2 > X^2_{0.05,\,4}$ hence null hypothesis is rejected

This shows that according to Poker Test given random numbers are not independent.

Assignment:

Define frequency test for random numbers. Develop the Poker test for tour digit numbers, and use it to test whether a sequence of following 1000-four digit numbers are independent. (Use $\alpha = 0.05\ and\ N = 4\ is\ 9.49$)

Combination i	Observed Frequency 0_i
Four different digit	565
One pair	392
Two pair	17
Three like digits	24
Four like digits	2
	1000

Assignment:

Short Questions

1. What do you mean by Pseudo random numbers?
2. Explain non-uniform random number generation.
3. Use the linear congruential method to generate a sequence of three two-digit random integers. Let X0=29, a=9, c=49 and m=100.
4. Use the mixed congruential method to generate a sequence of three two digit random numbers with Xo=37, a=7, c=29 and m=100.
5. Explain the congruence method of generating random numbers.

Long Questions

1. What is the main objective of gap test? Explain gap test algorithm with example.
2. Explain the process of Poker test for four digit numbers.
3. What do you mean by uniformity test? Explain the poker test with example.

End of unit-5