Chapter 5

Random Number Generations

5.1 Introduction

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.

3.2 Random Number Tables

A table of numbers generated in an unpredictable, haphazard that are uniformly distributed within certain interval are called random number table. The random number in random number table exactly obey two random number properties: uniformity and independence so random number generated form 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.

3.3 Pseudo Random Numbers

Pseudo means false, so false random numbers are being generated. The goal of any generation scheme is to produce a sequence of numbers between zero and 1 which simulates, or imitates, the ideal properties of uniform distribution and independence as closely as possible. 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.

3.4 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

3.5 Method to Generate Random Numbers

3.5.1 Linear Congruential Method

The linear congruential method, initially proposed by Lehrer [1951], produces a sequence of integers, X1, X2,... between zero and m — 1 according to the following recursive relationship:

$X_{i+1} = (a X_i + c) \bmod m$, i = 0, 1, 2 ...   Equation (3.1)

The initial value X0 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 Equation (3.1), 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 Xo drastically affects the statistical properties and the cycle length. An example will illustrate how this technique operates.

EXAMPLE 3.1

Use the linear congruential method to generate a sequence of random numbers with X0 = 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 zero to 99. Random numbers between zero and 1 can be generated by $R_i = X_i/m$, i= 1,2,…… equation (3.2)

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

X0 = 27

X1 = (17•27 + 43) mod 100 = 502 mod 100 = 2

R1=2∕100=0. 02

X2 = (17 • 2 + 43) mod 100 = 77 mod 100 = 77

R2=77∕100=0. 77

X3 = (17•77+ 43) mod 100 = 1352 mod 100 = 52

R3=52∕100=0. 52

First, notice that the numbers generated from Equation (3.2) can only assume values from the set I = {0,1 /m, 2/m,..., (m — l)/m), since each X_i is an integer in the set {0,1,2,..., m -1}. Thus, each Ri is discrete on I, instead of continuous on the interval [0, 1], This approximation appears to be of little consequence, provided that the modulus m is a very large integer. (Values such as m = 231 -1 and m = 248 are in common use in generators appearing in many simulation languages). By maximum density is meant that the values assumed by Ri = 1,2,..., leave no large gaps on [0,1].

EXAMPLE 3.2

Let m = 100, a = 19, c = 0, and X0 = 63, and generate a sequence c random integers using Xi+1

= (a Xi + c) mod m.

X0 = 63

X1 = (19)(63) mod 100 = 1197 mod 100 = 97

X2 = (19) (97) mod 100 = 1843 mod 100 = 43

X3 = (19) (43) mod 100 = 817 mod 100 = 17

EXAMPLE 3.3

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, X0 = 123,457.
The first few numbers generated are as follows:

X1= 7^5(123,457) mod (2^31 - 1) = 2,074,941,799 mod (2^31

- 1) X1 = 2,074,941,799

R1= X1∕2^31

X2 = 7^5(2,074,941,799) mod (2^31 - 1)

= 559,872,160 R2 = X2 /2^31= 0.2607

X3 = 7^5(559,872,160) mod (2^31 - 1) =

1,645,535,613 R3 = X3 /2^31= 0.7662

3.5.2 Midsquare Method

- First arithmetic generator: Midsquare method
  - von Neumann and Metropolis in 1940s
- The Midsquare method:
  - Start with a four-digit positive integer Z_0
  - Compute: $Z_0^2 = Z_0 \times Z_0$ to obtain an integer with up to eight digits
  - Take the middle four digits for the next four-digit number

i	Z_i	U_i	$Z_i \times Z_i$
0	7182	-	51581124
1	5811	0.5811	33767721
2	7677	0.7677	58936329
3	9363	0.9363	87665769
...			

- The main problem of this method is that 0 can be generated in the sequence.

3.6 Non Uniform Random Number Generation/Random Variate 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.

1. Non Uniform Transformation Method /Inverse Transform Method

The inverse transform technique can be used to sample from the exponential, uniform, triangular distribution etc. by inversing the 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.
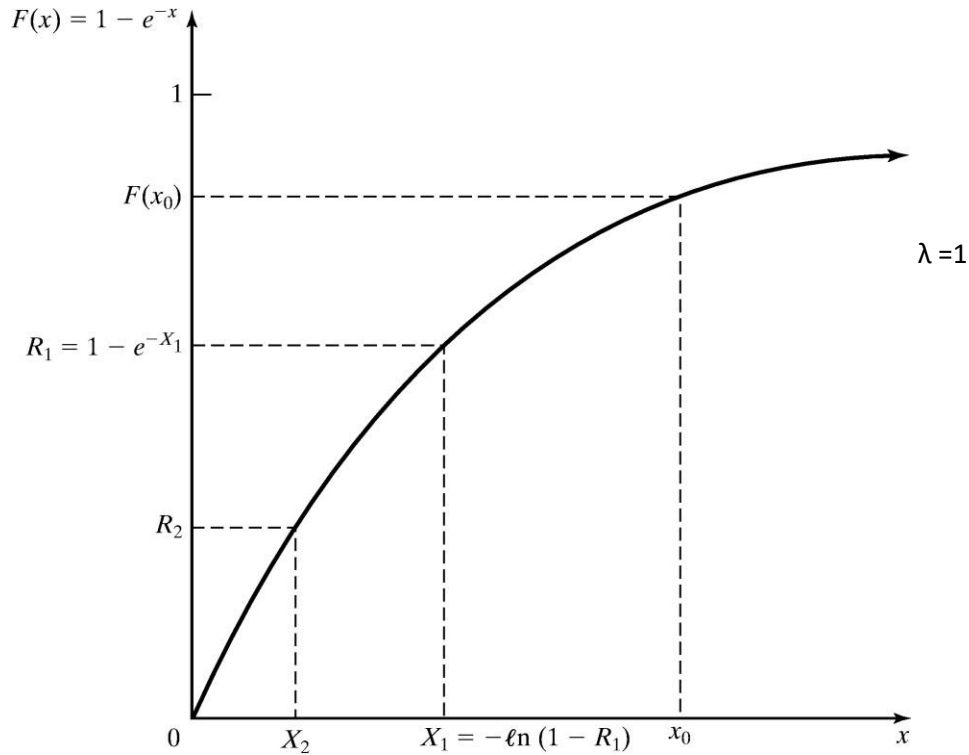
a) Exponential Distribution

The exponential distribution has the probability 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.

$F(x) = 1 - e^{-x}$

$F(x_0)$

$\lambda = 1$

$R_1 = 1 - e^{-X_1}$

R_2

0 X_2 $X_1 = -\ln(1 - R_1)$ x_0 x

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.

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

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

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

$X = - \frac{1}{\lambda}\ln(1 - R)$ (1)

Equation (1) is called random Variate generator for the exponential distribution. In general equation (1) is written as $X = F^{-1}(R)$

Example: Generation of Exponential Variates X_i with mean 1 (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:

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

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

$X_1 = -\ln(1 - R_1)$ (since $\lambda = 1$)

$X_1 = -\ln(0.1306) = 0.1400$ and so on.

Then, required random Variates are:

X_i	0.1400	0.0431	1.078	1.592	1.468

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 $F(X) = \dfrac{X-a}{b-a} = R$

$X = a + (b - a)\,R$ is the random Variate generator for the uniform distribution.

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

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

For 0 < X ≤ 1,

$R = X^2/2$

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

For 1 ≤ X ≤ 2,

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

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

2. Acceptance /Rejection method

- Useful particularly when inverse cdf does not exist in closed form
- Illustration: To generate random Variates, $X \sim U(1/4, 1)$

Procedures:

Step 1. Generate R ~ U[0,1]

Step 2a. If R >= ¼, accept X=R.

Step 2b. If R < ¼, reject R, return to Step 1
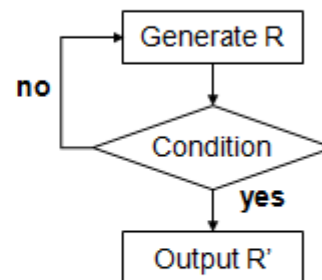
- R does not have the desired distribution, but R conditioned (R') on the event $\{R \geq \frac{1}{4}\}$ does.
- Efficiency: Depends heavily on the ability to minimize the number of rejections.

3.7 Testing for Randomness

The desirable properties of random numbers — 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.

a) Testing for uniformity

b) Testing for independence.

Testing for uniformity

The testing for uniformity can be achieved through different frequency test. These tests use the Kolmogorov-Smirnov 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

1) Kolmogorov-Smirnov test

2) Chi- square test

The detail description of each of these tests is given below. In testing for uniformity, the hypotheses are as follows:

H_0: R_i ~U/[0, 1]

H_1: R_i ~U/[0, l]

The null hypothesis, H_0 reads that the numbers are distributed 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

α = P (reject H_0 | H_0 true)

The decision maker sets the value of & for any test. Frequently, α is set to 0.01 or O.05.

1. The Kolmogorov-Smirnov test.

This test compares the continuous cdf, F(X), of the uniform distribution to the empirical cdf, SN(x), of the sample of N observations.

By definition,

F(x) = x, 0 <= x <= 1

If the sample from the random-number generator is R_1 R_2, • • •, RN, then the empirical cdf, SN(X), is defined by

SN(X) = (Number of R_1 R_2, • • •, R_n which are <= x)/N

As N becomes larger, SN(X) should become a better approximation to F(X), provided that the null hypothesis is true.

The **Kolmogorov-Smirnov test** is based on the largest absolute deviation or difference between F(x) and SN(X) over the range of the random variable. I.e. it is based on the statistic

D = max | F(x) - SN(x)|

For testing against a uniform cdf, the test procedure follows these steps:

Algorithm for K-S test

Step 1. Rank the data from smallest to largest. Let $R_{(i)}$ denote the i th smallest observation, so that $R_{(1)} <= R_{(2)} <= \bullet \bullet \bullet <= R_{(N)}$

Step 2. Compute

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

Step 3: Compute $D = \max\{D_+, D^-\}$

Step 4. Determine the critical value, $D\alpha$, from Table A.8(in your Text book) for the specified significance level α and the given sample size N.

Step 5.

If the sample statistic D is greater than the critical value, $D\alpha$, the null hypothesis that the data are a sample from a uniform distribution is rejected.

If D <= Da, conclude that no difference has been detected between the true distribution of { R1 R2, ,$\bullet \bullet \bullet$, Rn } and the uniform distribution. Hence the null hypothesis is accepted.

Example

Suppose that the five numbers 0.44, 0.81, 0.14, 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 a of 0.05.

Solution

First, the numbers must be ranked from smallest to largest. I.e the given numbers 0.05, 0.14 0.44 , 0.81, 0.93. The calculations can be facilitated by use of Table below.

R_i	0.05	0.14	0.44	0.81	0.93
i/N	0.2	0.4	0.6	0.8	1.0
i/N-R_i	0.15	0.26	**0.16**	-	0.07
R_i –(i-1)/N	0.15	-	**0.04**	0.21	0.13

For example

At $R_{(3)}$ the value of D+ is given by 3/5 - $R_{(3)}$ = 0.60 - 0.44 =0.16 and of D- is given by $R_{(3)}$ = 2/5 = 0.44 - 0.40 = 0.04. and other value also can be computed similarly.

Now The statistics are computed as D+ = 0.26 (Maximum of the row i/N-Ri) and D- = 0.21 (maximum of the row R_i –(i-1)/N) . Therefore, D = max{0.26, 0.21} = 0.26.

The critical value of **D,** obtained from Table A.8 (in Text book) for a = 0.05 and N = 5, is 0.565.

Since the computed value, 0.26, is less than the tabulated critical value, 0.565, the hypothesis of no difference between the distribution of the generated numbers and the uniform distribution is not rejected.

0.07 0.29 0.47   0.55
0.88

2. Chi- Square Test

The chi-square test uses the sample statistic

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

Where O_i; is the observed number in the ith class, E_i is the expected number in the ith class, and n is the number of classes. For the uniform distribution, E_i the expected number in each class is given by:

E_i = N/n

N is the total number of observation. It can be shown that the sampling distribution of X_0^2 is approximately the chi-square distribution with n - 1 degree of freedom.

Example

Use the chi-square test with a = 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

<u>Solution</u>

The table for chi square statistics is

Class interval(i)	Oi	Ei	(Oi – Ei)	$(Oi-Ei)^2$	$(Oi-Ei)^2/Ei$
1	8	10	-2	4	0.4
2	8	10	-2	4	0.4
3	10	10	0	0	0.0
4	9	10	-1	1	0.1
5	12	10	2	4	0.4
6	8	10	-2	4	0.4
7	10	10	0	0	0.0
8	14	10	14	16	0.16
9	10	10	0	0	0.0
10	11	10	1	1	0.1
Total	N=100	N=100			$\sum = 3.4$

Above Table contains the essential computations for chi square test. The test uses n = 10 intervals of equal length, namely [0.0, 0.1), [0.1, 0.2), . . . , [0.9, 1.0). The value of X^2 is 3.4.

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.Since X_0^2 is much smaller than the tabulated value of chi square, the null hypothesis of a uniform distribution is not rejected.

Test for independence includes the three types of tests as given below:

1) Autocorrelation Test tests the correlation between numbers and compares the sample correlation to the expected correlation of zero.

2) Gap test 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: Treats numbers grouped together as a poker hand. Then the hands obtained are compared to what is expected using the chi-square test.

1. Tests for Autocorrelation

The tests for autocorrelation are concerned with the dependence between numbers in a sequence. As an example, consider the following sequence of numbers:

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

From a visual inspection, these numbers appear random, and they would probably pass all the tests presented to this point. However, an examination of the 5th, 10th, 15th (every five numbers beginning with the fifth), and so on indicates a very large number in that position.

Now, 30 numbers is a rather small sample size to reject a random-number generator, but the notion is that numbers in the sequence might be related. In this particular section, a method for determining whether such a relationship exists is described. The relationship would not have to be all high numbers. It is possible to have all low numbers in the locations being examined, or the numbers may alternately shift from very high to very low.

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 form we start.
- **N** is the number of random numbers generated.
- **M** is the largest integer such that i+(M+1)m<=N

Now the autocorrelation between R_i, R_{i+m}, R_{i+2m},.......$R_{i+(M+1)m}$ is computed as

Now the autocorrelation ρ_{im} between R_i, R_{i+m}, R_{i+2m}, ...$R_{i+(M+1)m}$ is computed as

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

Now the test statistics is

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

Where

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

After computing Z0, do not reject the null hypothesis of independence if - $z_{a/2} <= Z0 <= z_{a/2}$, where a is the level of significance.

Example 3.1

Test whether the 3rd, 8th, 13th, and so on, numbers in the sequence at the beginning of this section are auto-correlated. (Use a = 0.05.) Here, i = 3 (beginning with the third number), m = 5 (every five numbers), N = 30 (30 numbers in the sequence).

Solution:

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.

i.e. 3+5M+5<=30 I.e.5M<=22 i.e. M<=22/5 4

Hence M=4

Then,

ρ35 = 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

= -0.1945

And

σ35= √ (13(4) + 7) / 12(4 + 1) = 0.1280

Then, the test statistic assumes the value

$Z_0 = -0.1945/0.1280 = -1.516$

Now, the critical value is $Z0.025 = 1.96$ ($Z\alpha/2$ is taken in this test)

Therefore, the hypothesis of independence cannot be rejected on the basis of this test.

4. 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, 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 and third gap of length zero. And so on. Similarly the gap associated with other digits can be calculated. The theoretical probability of first gap (of length 10 for digit 3) can be calculated as

The probability of a particular gap length can be determined by a Bernoulli trail.

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

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 for randomly ordered digits is given by

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

Gap Test Algorithm

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: Specify the cdf for the theoretical frequency distribution given by Equation above based on the selected class interval width.

Step2: Arrange the observed sample of gaps in a cumulative distribution with these same classes.

Step 3: Find D, the maximum deviation between F(x) and SN(X) as in K-S test.

Step 4: Determine the critical value, D_α, from Table for the specified value of α and the sample size N.

Step 5: If the calculated value of D is greater than the tabulated value of D_α, the null hypothesis of independence is rejected.

EXAMPLE

Based on the frequency with which gaps occur, analyze the 110 digits below to test whether they are independent. Use α = 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

The number of gaps is given by the number of data values minus the number of distinct digits, or 110 −10 = 100 in the example. The numbers of gaps associated with the various digits are as follows:

Digit	0	1	2	3	4	5	6	7	8	9
No of Gap	7	8	8	17	10	13	7	8	9	13

The Calculation for gap test is shown in the following table:

Gap Length	Frequency	Relative Frequency	Cumulative frequency S(X)	Theoretical Frequency F(X)	D=|F(X)-S(X)|
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	0.97	0.9657	0.0043
32-35	0	0	0.97	0.9775	0.0075
36-39	2	0.02	0.99	0.9852	0.0043
40-43	0	0	0.99	0.9903	0.0003
44-47	1	0.01	1.00	0.9936	0.0064

The critical value of D is given by $D_{0.05} = 1.36 / \sqrt{100} = 0.136$

Since D = max |F(x) - S$_N$(x)| = 0.0224 is less than D$_{0.05}$, we do not reject the hypothesis of independence on the basis of this test.

5. Poker Test

The poker test for independence is based on the frequency with which certain digits are repeated in a series of numbers. The following example shows an unusual amount of repetition:

0.255, 0.577, 0.331, 0.414, 0.828, 0.909, 0.303, 0.001...

The poker test uses the chi square statistics to accept or reject the null hypothesis.

In each case, a pair of like digits appears in the number that was generated. In three-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) = 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) = 0.01

P (exactly one pair) = 1 - 0.72 - 0.01 = 0.27

Example 1: 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.

Combination, i	Observed Frequency, O,	Expected Frequency, £,	$\dfrac{(0, - E i)^2}{E,}$
Three different digits	680	720	2.22
Three like digits	31	10	44.10
Exactly one pair	289	270	1.33
	1000	1000	47.65

The appropriate degrees of freedom are one less than the number of class intervals. Since

47.65 > $X^2_{0.05, 2}$ = 5.99 (tabulated value), the independence of the numbers is rejected on the basis of this test. Here 2 or n-1 is the degree of freedom since there are only 3 (n) classes.

Example 2

Explain the independence test. 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

In four digit number, there are five different possibilities

- ✓ All individual digits can be different
- ✓ There can be one pair of like digit
- ✓ There can be two pair of like digits
- ✓ There can be three digits of a kind
- ✓ There can be four digits of a kind

The probabilities associated with each of the possibilities is given by

P (four different digits) = $4c_4$ x10/10x9/10x8/10x7/10=0.504
P (one pair) = $4c_2$ x10/10x1/10x9/10x8/10=0.432
P (two pair) = $4c_2$/2 x10/10x1/10x9/10x1/10=0.027
P (three digits of a kind) = $4c_3$ x10/10x1/10x1/10x9/10=0.036
P (four digits of a kind) = $4c_4$ x10/10x1/10x1/10x1/10=0.001

Now the calculation table for the Chi-square statistics is:

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

Here the calculated value of chi-square is 25.185 which is greater than the given value of chi-square so we reject the null hypothesis of independence between given numbers.

Exercise

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.

Chi-Square Table

	P(X ≤ x)							
	0.010	0.025	0.050	0.100	0.900	0.950	0.975	0.990
r	$\chi^2_{0.99}(r)$	$\chi^2_{0.975}(r)$	$\chi^2_{0.95}(r)$	$\chi^2_{0.90}(r)$	$\chi^2_{0.10}(r)$	$\chi^2_{0.05}(r)$	$\chi^2_{0.025}(r)$	$\chi^2_{0.01}(r)$
1	0.000	0.001	0.004	0.016	2.706	3.841	5.024	6.635
2	0.020	0.051	0.103	0.211	4.605	5.991	7.378	9.210
3	0.115	0.216	0.352	0.584	6.251	7.815	9.348	11.34
4	0.297	0.484	0.711	1.064	7.779	9.488	11.14	13.28
5	0.554	0.831	1.145	1.610	9.236	11.07	12.83	15.09
6	0.872	1.237	1.635	2.204	10.64	12.59	14.45	16.81
7	1.239	1.690	2.167	2.833	12.02	14.07	16.01	18.48
8	1.646	2.180	2.733	3.490	13.36	15.51	17.54	20.09
9	2.088	2.700	3.325	4.168	14.68	16.92	19.02	21.67
10	2.558	3.247	3.940	4.865	15.99	18.31	20.48	23.21

Kolmogorov-Smirnov Table

Sample Size, N	Level of Significance, α				
	0.20	0.15	0.10	0.05	0.01
3	0.565	0.597	0.642	0.708	0.828
4	0.494	0.525	0.564	0.624	0.733
5	0.446	0.474	0.474	0.565	0.669
10	0.322	0.342	0.368	0.410	0.490
15	0.266	0.283	0.304	0.338	0.404
20	0.231	0.246	0.264	0.294	0.356
25	0.21	0.22	0.24	0.27	0.32
30	0.19	0.20	0.22	0.24	0.29
35	0.18	0.19	0.21	0.23	0.27
40	0.17	0.18	0.19	0.21	0.25
45	0.16	0.17	0.18	0.20	0.24
50	0.15	0.16	0.17	0.19	0.23
over 50	$\dfrac{1.07}{\sqrt{N}}$	$\dfrac{1.14}{\sqrt{N}}$	$\dfrac{1.22}{\sqrt{N}}$	$\dfrac{1.36}{\sqrt{N}}$	$\dfrac{1.63}{\sqrt{N}}$

Normal Distribution Table

Standard Normal Probabilities
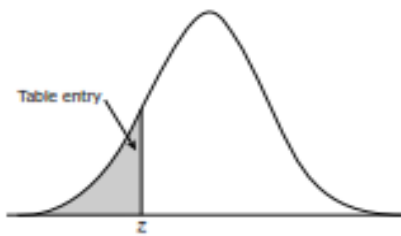
Table entry for z is the area under the standard normal curve to the left of z.

z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
−3.4	.0003	.0003	.0003	.0003	.0003	.0003	.0003	.0003	.0003	.0002
−3.3	.0005	.0005	.0005	.0004	.0004	.0004	.0004	.0004	.0004	.0003
−3.2	.0007	.0007	.0006	.0006	.0006	.0006	.0006	.0005	.0005	.0005
−3.1	.0010	.0009	.0009	.0009	.0008	.0008	.0008	.0008	.0007	.0007
−3.0	.0013	.0013	.0013	.0012	.0012	.0011	.0011	.0011	.0010	.0010
−2.9	.0019	.0018	.0018	.0017	.0016	.0016	.0015	.0015	.0014	.0014
−2.8	.0026	.0025	.0024	.0023	.0023	.0022	.0021	.0021	.0020	.0019
−2.7	.0035	.0034	.0033	.0032	.0031	.0030	.0029	.0028	.0027	.0026
−2.6	.0047	.0045	.0044	.0043	.0041	.0040	.0039	.0038	.0037	.0036
−2.5	.0062	.0060	.0059	.0057	.0055	.0054	.0052	.0051	.0049	.0048
−2.4	.0082	.0080	.0078	.0075	.0073	.0071	.0069	.0068	.0066	.0064
−2.3	.0107	.0104	.0102	.0099	.0096	.0094	.0091	.0089	.0087	.0084
−2.2	.0139	.0136	.0132	.0129	.0125	.0122	.0119	.0116	.0113	.0110
−2.1	.0179	.0174	.0170	.0166	.0162	.0158	.0154	.0150	.0146	.0143
−2.0	.0228	.0222	.0217	.0212	.0207	.0202	.0197	.0192	.0188	.0183
−1.9	.0287	.0281	.0274	.0268	.0262	.0256	.0250	.0244	.0239	.0233
−1.8	.0359	.0351	.0344	.0336	.0329	.0322	.0314	.0307	.0301	.0294
−1.7	.0446	.0436	.0427	.0418	.0409	.0401	.0392	.0384	.0375	.0367
−1.6	.0548	.0537	.0526	.0516	.0505	.0495	.0485	.0475	.0465	.0455
−1.5	.0668	.0655	.0643	.0630	.0618	.0606	.0594	.0582	.0571	.0559
−1.4	.0808	.0793	.0778	.0764	.0749	.0735	.0721	.0708	.0694	.0681
−1.3	.0968	.0951	.0934	.0918	.0901	.0885	.0869	.0853	.0838	.0823
−1.2	.1151	.1131	.1112	.1093	.1075	.1056	.1038	.1020	.1003	.0985
−1.1	.1357	.1335	.1314	.1292	.1271	.1251	.1230	.1210	.1190	.1170
−1.0	.1587	.1562	.1539	.1515	.1492	.1469	.1446	.1423	.1401	.1379
−0.9	.1841	.1814	.1788	.1762	.1736	.1711	.1685	.1660	.1635	.1611
−0.8	.2119	.2090	.2061	.2033	.2005	.1977	.1949	.1922	.1894	.1867
−0.7	.2420	.2389	.2358	.2327	.2296	.2266	.2236	.2206	.2177	.2148
−0.6	.2743	.2709	.2676	.2643	.2611	.2578	.2546	.2514	.2483	.2451
−0.5	.3085	.3050	.3015	.2981	.2946	.2912	.2877	.2843	.2810	.2776
−0.4	.3446	.3409	.3372	.3336	.3300	.3264	.3228	.3192	.3156	.3121
−0.3	.3821	.3783	.3745	.3707	.3669	.3632	.3594	.3557	.3520	.3483
−0.2	.4207	.4168	.4129	.4090	.4052	.4013	.3974	.3936	.3897	.3859
−0.1	.4602	.4562	.4522	.4483	.4443	.4404	.4364	.4325	.4286	.4247
−0.0	.5000	.4960	.4920	.4880	.4840	.4801	.4761	.4721	.4681	.4641

Standard Normal Probabilities

Table entry for z is the area under the standard normal curve to the left of z.

z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
0.0	.5000	.5040	.5080	.5120	.5160	.5199	.5239	.5279	.5319	.5359
0.1	.5398	.5438	.5478	.5517	.5557	.5596	.5636	.5675	.5714	.5753
0.2	.5793	.5832	.5871	.5910	.5948	.5987	.6026	.6064	.6103	.6141
0.3	.6179	.6217	.6255	.6293	.6331	.6368	.6406	.6443	.6480	.6517
0.4	.6554	.6591	.6628	.6664	.6700	.6736	.6772	.6808	.6844	.6879
0.5	.6915	.6950	.6985	.7019	.7054	.7088	.7123	.7157	.7190	.7224
0.6	.7257	.7291	.7324	.7357	.7389	.7422	.7454	.7486	.7517	.7549
0.7	.7580	.7611	.7642	.7673	.7704	.7734	.7764	.7794	.7823	.7852
0.8	.7881	.7910	.7939	.7967	.7995	.8023	.8051	.8078	.8106	.8133
0.9	.8159	.8186	.8212	.8238	.8264	.8289	.8315	.8340	.8365	.8389
1.0	.8413	.8438	.8461	.8485	.8508	.8531	.8554	.8577	.8599	.8621
1.1	.8643	.8665	.8686	.8708	.8729	.8749	.8770	.8790	.8810	.8830
1.2	.8849	.8869	.8888	.8907	.8925	.8944	.8962	.8980	.8997	.9015
1.3	.9032	.9049	.9066	.9082	.9099	.9115	.9131	.9147	.9162	.9177
1.4	.9192	.9207	.9222	.9236	.9251	.9265	.9279	.9292	.9306	.9319
1.5	.9332	.9345	.9357	.9370	.9382	.9394	.9406	.9418	.9429	.9441
1.6	.9452	.9463	.9474	.9484	.9495	.9505	.9515	.9525	.9535	.9545
1.7	.9554	.9564	.9573	.9582	.9591	.9599	.9608	.9616	.9625	.9633
1.8	.9641	.9649	.9656	.9664	.9671	.9678	.9686	.9693	.9699	.9706
1.9	.9713	.9719	.9726	.9732	.9738	.9744	.9750	.9756	.9761	.9767
2.0	.9772	.9778	.9783	.9788	.9793	.9798	.9803	.9808	.9812	.9817
2.1	.9821	.9826	.9830	.9834	.9838	.9842	.9846	.9850	.9854	.9857
2.2	.9861	.9864	.9868	.9871	.9875	.9878	.9881	.9884	.9887	.9890
2.3	.9893	.9896	.9898	.9901	.9904	.9906	.9909	.9911	.9913	.9916
2.4	.9918	.9920	.9922	.9925	.9927	.9929	.9931	.9932	.9934	.9936
2.5	.9938	.9940	.9941	.9943	.9945	.9946	.9948	.9949	.9951	.9952
2.6	.9953	.9955	.9956	.9957	.9959	.9960	.9961	.9962	.9963	.9964
2.7	.9965	.9966	.9967	.9968	.9969	.9970	.9971	.9972	.9973	.9974
2.8	.9974	.9975	.9976	.9977	.9977	.9978	.9979	.9979	.9980	.9981
2.9	.9981	.9982	.9982	.9983	.9984	.9984	.9985	.9985	.9986	.9986
3.0	.9987	.9987	.9987	.9988	.9988	.9989	.9989	.9989	.9990	.9990
3.1	.9990	.9991	.9991	.9991	.9992	.9992	.9992	.9992	.9993	.9993
3.2	.9993	.9993	.9994	.9994	.9994	.9994	.9994	.9995	.9995	.9995
3.3	.9995	.9995	.9995	.9996	.9996	.9996	.9996	.9996	.9996	.9997
3.4	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9998