Chapter 3

Random Number Generations

3.1 Introductions

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 lines 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:

(a) Autocorrelation between numbers.

(b) Numbers successively higher or lower than adjacent numbers.

(c) 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.

1. The routine should be fast. . The total cost can be managed by selecting a computationally efficient method of random-number generation.

2. 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.

3. 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.

4. 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.

5. 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 Lehmer [1951], produces a sequence of integers, X_1, X_2,... 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 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 Equation (7.1), 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 $X_0 = 27$, $a = 17$, $c = 43$, 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 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 (7.2) can only assume values from the set I = {0,1 /m, 2/m,..., (m — l)/m), since each *Xi* 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 $m = 10^2 = 100$, $a = 19$, $c = 0$, 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

.

.

When m is a power of 10, say $m = 10^b$, the modulo operation is accomplished by saving the b rightmost (decimal) digits.

EXAMPLE 4.4

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 ⁄ 231

X2 = 7^5(2,074,941,799) mod (2^{31} - 1) = 559,872,160

$R2$ = X2 ⁄ 231= 0.2607

$X3 = 7^5(559,872,160) \bmod (2^{31} - 1) = 1,645,535,613$

$R3$ = X3 ⁄ 231= 0.7662

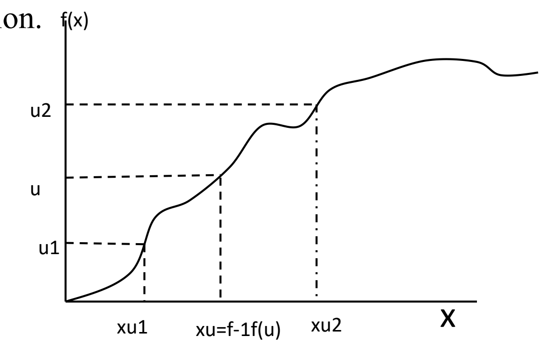
3.5.2 Inverse Transformation Method or

Probability Integral Transformation Method:

This Method requires a sequence of uniformly distributed random numbers. If ui (i=1,2,3,...) are independent uniformly distributed random numbers over the interval 0 to 1 and F-1(x) is the inverse of the cumulative distribution function for random variable X then the random variables generated using inverse transformation method will be xi=F^{-1}(ui). That is to produce random numbers from given probability function; the inverse cumulative distribution function must be evaluated with a sequence of uniformly distributed numbers in the interval 0 to 1.

Consider a probability distribution function f(x) which is continuous. Generate n random samples x1,x2,x3,............from f(x). The probability distribution function increase from 0 to 1 and the probability that a random sample x lies in the interval (x1,x2) is equal to f(x1)-f(x2) for all pairs of x1<=x2. Since f(x) is continuous, it takes all the values between 0 and 1. Therefore for any number u, where 0<=u<=1, there exist a unique x_u, such that $f(x_u) = u$. This value of u can be repeated by $f^{-1}(u)$ called the inverse function.

i. e. $f(x_u)= u$
$x_u=f^{-1}(u)$



Therefore to generate n samples, from only continuous probability distribution function, generate n uniform random numbers u_1, u_2, ...,u_n in the interval (0,1) and apply inverse transformation function $f^{-1}(u_i)$ to each.

Example:

Derive an equation to generate non-uniform random numbers from an exponential having pdf f(x) $=\lambda e^{-\lambda x}$ (x>=0) using inverse transformation method.

Sol^n:

1. Compute the cdf of the given random variable

$$\text{cdf}=\int_0^x f(x)dx = 1 - e^{-\lambda x}$$

2. set f(x) =u on the range of x i. e. $1 - e^{-\lambda x} = u, x \geq 0$

since x is random $1 - e^{-\lambda x}$ is also random over the interval 0 to 1.

3. Solve the equation f(x)= u for x in terms of u.

i. e. $1 - e^{-\lambda x} = u$

or, $e^{-\lambda x}$ =1-u

or, $-\lambda x$=ln(1-u)

or, $x= -1/\lambda \ln (1-u)$

this equation is called a random variable generate for the exponential distribution. this equation can be written as $x=f^{-1}(u)$

4. Generate uniform number $u_1,u_2,.....u_n$ and compute the desired random variables by using $x_i= f^{-1}(u)$; for exponential distribution $f^{-1}(u)= -1/\lambda \ln (1-u)$ for (i=1,2,3....)

since u_i and $1-u_i$ are uniformly distributed random numbers between 0 and 1, we can replace $(1-u_i)$ by ui. Therefore the equation becomes $x_i=-1/\lambda \ln (u_i)$.

3.5.3 Acceptance /rejection method

This method is applicable when the probability density function f(x) has a lower and upper limit to its range (A,B) and upper bound C. This method for obtaining samples from a given non uniform distribution basically works by generating uniform random numbers repeatedly and accepting only those numbers that meet a certain condition. for the rejection method to be applicable the pdf must be non zero only over a finite interval (A,B).

The steps involved in the acceptation/rejection procedure are:

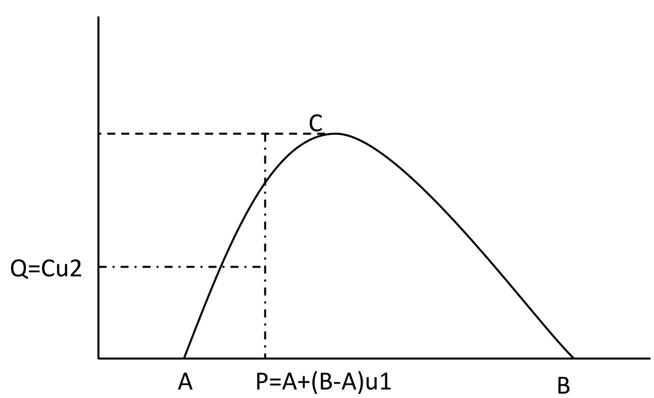
1. Generate a pair of independent uniformly distributed variables u_1 and u_2 in the interval (0,1).

2. Using u_1 compute a point P on the horizontal axis as $P=A+(B-A)u_1$.

3. Using u_2 compute a point Q on the vertical axis as $Q= C.u_2$

4. If $Q<=f(x)$ accept P as the value of a sample from the desired distribution, otherwise reject the pair and go to step 1. i. e. repeat the above process with a pair of new uniform variables.

In the above process steps 1,2,3 create a random points and the last step relates the points to the curve of the pdf. If the point P is accepted as the sample from the desired distribution else the point is rejected and the process as repeated.



3.5 Testing for Randomness

The desirable properties of random numbers — uniformity and independence To insure 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

a) Kolmogorov-Smirnov test

b) Chi- square test

Test for independence includes the five types of tests as given below

a) Autocorrelation test Tests the correlation between numbers and compares the sample correlation to the expected correlation of zero.

b) 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,

c) 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.

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

H0: Ri ~ U/[0,1]

H1: Ri ~U/[0,l]

The null hypothesis, H0 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 *a* must be stated. The level *a* is the probability of rejecting the null hypothesis given that the null hypothesis is true, or

a = *P* (reject H0 |H0 true)

The decision maker sets the value of & for any test. Frequently, *a* 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 R1 R2, ,• • •, RN, then the empirical cdf, $S_N(X)$,

is defined by

$S_N(X)$ =(number of R1 R2, ,• • •, Rn which are <= x)/N

As N becomes larger, $S_N(X)$ 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 $S_N(X)$ over the range of the random variable. I.e. it is based on the statistic

D = max | F(x) - $S_N(x)$| ……………………………..equation (3.1)

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) <= • • • <= R (N)

Step 2. Compute

$$D^+ = \text{MAX}\left\{\frac{i}{N} - R_i\right\}$$

$$D- = \text{Max}\left\{R_i - \frac{i-1}{N}\right\}$$

Step 3: Compute D=max{D+, D-}

Step 4. Determine the critical value, Dα, 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, Dα, 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, ,• • •, 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.

Ri	0.05	0.14	0.44	0.81	0.93
i/N	0.2	0.4	0.6	0.	1.0
i/N-Ri	0.15	0.26	0.16	------	0.07
Ri-(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 Ri –(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.

2. Chi- Square Test

The chi-square test uses the sample statistic

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

where Oi; is the observed number in the *i* th class, *Ei* is the expected number in the ith class, and *n* is the number of classes. For the uniform distribution, *Ei* the expected number in each class is given by Ei = N/n ; for equally spaced classes, where *N* is the total number of observations. It can be shown that the sampling distribution of χ_0^2 is approximately the chi-square distribution with *n* - 1 degrees of freedom.

Example 2.1

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

Solutions

The table for chi square statistics is

Class interval (i)	O_i	E_i	O_i-E_i	$(O_i-E_i)^2$	$(O_i-E_i)^2/E_i$
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	4	16	0.16
9	10	10	0	0	0.0
10	11	10	1	1	0.1
Total	N=100	N=100			Σ=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 χ_0^2 is 3.4.

Here degree of freedom is n-1=10-1=9 and α=0.05. The tabulated value of $\chi^2_{0.05, 9}$ =16.9.Since χ_0^2 is much smaller than the tabulated value of chi square, the null hypothesis of a uniform distribution is not rejected.

3. 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

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 number in a sequence. The test computes the auto correlation between every m numbers (m is also known as lag) starting with ith 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 ρ_{im} between Ri, Ri+m, Ri+2m, ……Ri+(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 **- za/2** <= **Z0** <= **za/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.eM<=22/5 4

hence M=4

Then,

ρ_{35} = 1/ 4 + 1**[** (0.23)(0.28) + (0.28)(0.33) + (0.33)(0.27) + (0.27)(0.05) + (0.05)(0.36) **]**

= -0.1945

And

$\sigma_{35} = \sqrt{(13(4) + 7) / 12(4 + 1)} = 0.1280$

Then, the test statistic assumes the value

$Z_0 = -0.1945/0.1280 = -1.516$

Now, the critical value is $Z_{0.025} = 1.96$ (Za/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 gap can occurs. The first gap is of length 9 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 lenth 10 for digit 3) can be calculated as

$$P(\text{gapof}\,10) = \frac{P(\text{not }3) \times p(\text{not }3) \ldots\ldots\ldots\ldots\ldots P(\text{not}3)\; P(3)}{10 \text{ of these terms}}$$

$$= 0.9 \times 0.9 \times \ldots\ldots\ldots\ldots\ldots \times 0.9 \times 0.1$$

$$= (0.9)^{10} \times 0.1$$

Hence in general the probability of length x is

P(t followed by exactly *x* non-r digits) = $(0.9)^x (0.1)$, X = 0,1,2…………equation 4.1

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 (4.1) 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 $S_N(X)$ as in K-S test.

Step 4. Determine the critical value, Da, from Table for the specified value of *a* and the sample size *N.*

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

EXAMPLE 4.13

Based on the frequency with which gaps occur, analyze the 110 digits Below to test whether they are independent. Use *a* = 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 number of gaps associated with the various digits are as follows:

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

The calculation for gap test is shown in following tables

Gap Length	Frequency	Relative frequency (freq/100)	Cumulative frequency $S_N(x)$	Theoretical frequencyF(x)	D=$\lvert F(x) - S(x)\rvert$
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	0.224
12-15	9	0.09	0.88	0.8147	0.0153
16-19	5	0.05	0.94	0.8784	0.0016
20-23	6	0.06	0.97	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	0.99	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 5.1

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 *a* = 0.05. Test these numbers using **poker test for three digit.**

The test is summarized in Table as:

Combination (i)	Observed Frequency(Oi)	Expected Frequency (Ei)	$\frac{(O_i - E_i)^2}{E_i}$
Three different digits	680	720	2.22
Three like digits	31	10	44.10
Exactly one pair	289	270	1.33
Total	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) class.

Example 5.2

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

a. All individual digits can be different

b. There can be one pair of like digit

c. There can be two pair of like digits

d. There can be three digits of a kind

e. There can be four digits of a kind

The probabilities associated with each of the possibilities is given by

P(four different digits)= $^4c_4 \times 10/10 \times 9/10 \times 8/10 \times 7/10 = 0.504$

P(one pair)= $^4c_2 \times 10/10 \times 1/10 \times 9/10 \times 8/10 = 0.432$

P(two pair)= $^4c_2 \times 10/10 \times 1/10 \times 9/10 \times 1/10 = 0.027$

P(three digits of a kind)= $^4c_3 \times 10/10 \times 1/10 \times 1/10 \times 9/10 = 0.036$

P(four digits of a kind)= $^4c_4 \times 10/10 \times 1/10 \times 1/10 \times 1/10 = 0.001$

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	0.504×1000=504	56	6.22
One pair	394	0.432×1000=432	-38	3.343
Two pair	32	0.027×1000=27	5	0.926
Three digits of a kind	13	0.036×1000=	-23	14.494
Four digit of a kind	1	0.0001×1000=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.