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 arc used to generate event limes and other random variables.

PROPERTIES OF RANDOM NUMBERS

A sequence of random numbers, $R_1, R_2, \ldots\ldots$, must have two important statistical properties: uniformity and independence. Each random number R_i must be an independent sample drawn from a continuous uniform distribution between zero and 1 – that is, the pdf is given by

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

This density function is shown in Figure 5.1. The expected value of each R_i is given by

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

and the variance is given by

$$V(R) = \int_0^1 x^2 dx - [E(R)]^2 = \frac{x^3}{3}\Big|_0^1 - \left(\frac{1}{2}\right)^2 = \frac{1}{3} - \frac{1}{4} = \frac{1}{12}$$

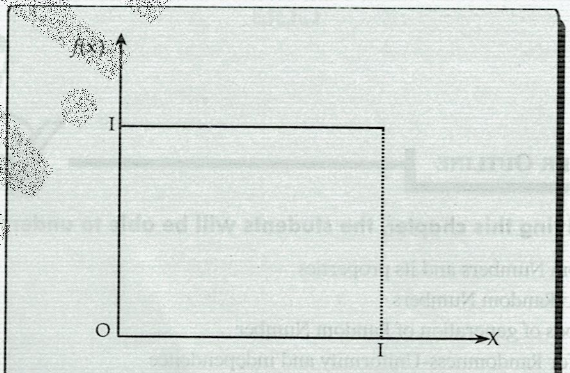


Fig: 5.1: pdf for random numbers

Some consequences of the uniformity and independence properties are the following:

1. If the interval [0, 1] is divided into n classes, or subintervals of equal length, the expected number of observations in each interval is $\frac{N}{n}$, where N is the total number of observations.
2. The probability of observing a value in a particular interval is independent of the previous values drawn.

PSEUDO RANDOM NUMBERS

"Pseudo" means false, so false random numbers are being generated. In this instance, "pseudo' is used to imply that the very act of generating random numbers by a known method removes the potential for true randomness. If the method is known, the set of random numbers can be replicated. Then an argument can be made that the numbers are not truly random. The goal of any generation scheme, however, is to produce a sequence of numbers between 0 and 1 that simulates, or imitates, the ideal properties of uniform distribution and independence as closely as possible.

To be sure, in the generation of pseudo-random numbers, certain problems or errors can occur. These errors, or departures from ideal randomness, are all related to the properties stated previously. Some examples of such departures include the following:

1. The generated numbers might not be uniformly distributed.
2. The generated numbers might be discrete-valued instead of continuous-valued.
3. The mean of the generated numbers might be too high or too low.
4. The variance of the generated numbers might be too high or too low.

5. There might be dependence. The following are examples:
   (a) Auto correlation between numbers.
   (b) Numbers successively higher or lower than adjacent numbers;
   (c) Several numbers above the mean followed by several numbers below the mean.

Departures from uniformity and independence for a particular generation scheme often can be detected by different uniformity and independent tests. If such departures are detected, the generation scheme should be dropped in favor of an acceptable generator. Usually, random numbers are generated by a digital computer, as part of the simulation. There are numerous methods that can be used to generate the values. Before we describe some of these methods, or routines, we should mention the following important considerations.

1. The routine should be fast. Individual computations are inexpensive, but simulation could require many millions of random numbers. The total cost can be managed by selecting a computationally efficient method of random-numbers generation.
2. The routine should be portable to different computers – and, ideally, to different programming languages. This is desirable so that the simulation program will produce the same results wherever it is executed.
3. The routine should have 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 of 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. Those is helpful for debugging purposes and is a means of facilitating comparisons between systems. For the same reasons, it should be possible to easily specify different starting points, widely separated, within the sequence.

Inventing techniques that seem to generate random numbers is easy: inventing techniques that really do produce sequences that appear to be independent, uniformly distributed random numbers is incredibly difficult. There is now a vast literature and rich theory on the topic, and many hours of testing have been devoted to establishing the properties of various generators. Even when a technique is known to be theoretically sound, it is seldom easy to implement it in a way that will be fast and portable.

METHODS OF GENERATING RANDOM NUMBERS

1. Linear Congruential Generator/ Method

The linear congruential method of is the most widely used techniques for generating random numbers. This method, initially proposed by Lehmer [1951], produces a sequence of integers, $X_1, X_2, \ldots\ldots$ between zero and $m - 1$ by following a recursive relationship:

$$X_{i+1} = (aX_i + c) \bmod m,\ i = 0, 1, 2, \ldots\ldots n$$

The initial value X_0 is called the seed, a is called the multiplier, c is the increment, and m is the modulus. If $c \neq o$ in Equation (5.3), then the form is called the mixed Congruential method. When $c = 0$, the form is known as the multiplicative congruential method. The selection of the values for a, c, m, and X_0 drastically affects the statistical properties and the cycle length. Variations of Equation (5.3) are quite common in the computer generation of random numbers. An example will illustrate how this technique operates.

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 value generated will all be between zero and 99 because of the value of the modulus. Also notice that random integers are being generated rather than random numbers. These random integers should appear to be uniformly distributed on the integers zero to 99. Random numbers between zero and 1 can be generated by

$$R_i = \frac{X_i}{m},\ i = 1, 2, \ldots$$

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

$$X_0 = 27$$

$$X_1 = (17*27 + 43) \bmod 100 = 502 \bmod 100 = 2$$

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

$$X_2 = (17*2 + 43) \bmod 100 = 77 \bmod 100 = 77$$

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

$$X_3 = (17*77 + 43) \bmod 100 = 1352 \bmod 100 = 52$$

$$R_3 = \frac{52}{100} = 0.52$$

Recall that $a = b \bmod m$ provided that $(b - a)$ is divisible by m with no remainder. Thus, $X_1 = 502 \bmod 100$, but $\frac{502}{100}$ equals 5 with a remainder of 3, so that $X_1 = 2$. In other words, $(502 - 2)$ is evenly divisible by $m = 100$, so X_1 502 "reduces" to $X_1 = 2$ mod 100.

The ultimate test of the linear congruential method, as of any generation scheme, is how closely the generated numbers $R_1, R_2, \ldots$ approximate uniformity and independence. There are, however, several secondary properties that must be considered. These include maximum density and maximum period.

First, notice that the numbers generated from Equation (5.3) assume values only from the set I = {0, 1/m, 2/m, ..., (m – 1)/m}, because each X_i is an integer in the set {0, 1, 2, ...m –1}. Thus, each R_i is discrete on I, instead of continuous on the interval [0, 1]. This approximation appears to be of little consequence if the modulus m is a very large integer. (Values such as m = 23^{31} – 1 and m = 2^{48} are in common use in generators appearing in many simulation languages.) By maximum density is meant that the values assumed by R_i, i = 1, 2 ...leave no large gaps on [0, 1].

Second, to help achieve maximum density, and avoid cycling (i.e. recurrence of the same sequence of generated numbers) in practical applications, the generator should have the largest possible period. Maximal period can be achieved by the proper choice of a, c, m, and X_o [Fishman, 1978; Law and Kelton, 2000].

- For m a power of 2, say m = 2^b, and c ≠ 0, the longest possible period is p = m = 2^b, which is achieved whenever c is relatively prime to m (that i, the greatest common factor of c and m is 1) and a = 1 + 4k, Where k is an integer
- For m a power of 2, say m = 2^b, and c = 0, the longest possible period is p = m/4 = 2 b – 2, which is achieved if the seed X_0 is odd and if the multiplier, a, is given by a = 3 + 8k or a = 5 + 8k, for some k = 0, 1, ...
- For m a prime number and c = 0, the longest possible period is p = m – 1, which is achieved whenever the multiplier, a, has property that the smallest integer k such that a^{k-1} is divisible m is k = m – 1.

Example 2

Let m = 10^2 = 100, a = 19, c = 0, and X_0 = 63, and generate a sequence of random integers using LCG

X_0 = 63

X_1 = (19)(63) mod 100 = 1197 mod 100 = 97

X_2 = (19) (97) mod 100 = 1843 mod 100 = 43

X_3 = (19) (43) mod 100 = 817 mod 100 = 17

.....

When m is power of 10, say $m = 10^b$, the modulo operation is accomplished by saving the right most (decimal) digits. By analogy, the modulo operation is most efficient for binary computers when $m = 2^b$ for some $b > 0$.

2. Mid Square Method

Mid square method is also used to generate pseudo random numbers. For this we follow the following steps:

- Select a seed number (X) with n-digits. $(X_1X_2.....X_n)$. X is the initial random number. Note: <u>n is even</u>
- Square X to obtain number (Y) with m-digits. $Y_1Y_2.....Y_m = (X_1X_2.....X_n)^2$
- Add zeros to the left of Y to form Z number with (2n) digits,

  $Z_1Z_2.....Z_{2n} = 00.....0\ Y_1Y_2.....Y_m$
- Extract the middle n-digits of Z, which represent by number R,

  $R_1R_2.....R_n = Z_{(n/2)}Z_{(n/2)+1}.....Z_{(3n/2)}$
- Here R is random number and if more random number are needed repeat steps 1 to 4.

Example:

Calculate the first five random numbers using mid-square method for the cases:

a. Seed=15

15,	$(15)^2 = 225$,	0225,	then $R_1 = 22$
22,	$(22)^2 = 484$,	0484,	then $R_2 = 48$
48,	$(48)^2 = 2304$,	2304,	then $R_3 = 30$
30,	$(30)^2 = 900$,	0900,	then $R_4 = 90$
90,	$(90)^2 = 8100$,	8100,	then $R_5 = 10$

b. Seed=1920

1920,	$(1920)^2 = 3686400$,	03686400,	then $R_1 = 6864$
6864,	$(6864)^2 = 47114496$,	47114496,	then $R_2 = 1144$
1144,	$(1144)^2 = 1308736$,	01308736,	then $R_3 = 3087$
3087,	$(3087)^2 = 9529569$,	09529569,	then $R_4 = 5295$
5295,	$(5295)^2 = 28037025$,	28037025,	then $R_5 = 0370$

Tests for Random Numbers

The desirable properties of random numbers are uniformity and independence. To check on whether these desirable properties have been achieved, a number of tests can be performed. The tests can be placed in two categories, according to the properties of interest: uniformity, and independence. A brief description of two types of tests is given in this chapter.

1. Frequency test. Uses the Kolmogorov – Smirnov or the chi-square test to compare the distribution of the set of numbers generated to a uniform distribution.
2. Autocorrelation test. Tests the correlation between numbers and compares the sample correlation to the expected correlation zero.

In testing for uniformity, the hypotheses are as follows:

$H_0 : R_i \sim U[0, 1]$

$H_1 : R_i \nsim U[0, 1]$

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 evidence of non uniformity has not been detected by this test. This does not imply that further testing of the generator for uniformity is unnecessary.

In testing for independence, the hypotheses are as following

$H_0 : R_i \sim$ independently

$H_1 : R_i \nsim$ independently

The null hypothesis, H_0, reads that the numbers are independent. Failure to reject the null hypothesis means that evidence of dependence has not been detected by this test This does not imply that further testing of the generator for independence is unnecessary.

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

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

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

If several tests are conducted on the same set of numbers the probability of rejecting the null hypothesis on at least one test, by chance alone [i.e. making a Type I(α) error], increases. Say that α = 0.065 and that five different tests are conducted on a sequence of numbers. The probability of rejecting the null hypothesis on at least one test, by chance alone, could be as large as 0.25.

Similarly, if one test is conducted on many sets of numbers from a generator, the probability of rejecting the null hypothesis on at least one test by chance alone [i.e. making a Type I (α) error], increases as more sets of numbers are tested. For instance, if 100 sets of numbers were subjected to the test, with α = 0.05, it would be expected that five of those tests would be rejected by chance alone. If the number of rejections in 100 tests is close to 100α, then there is no compelling reason to discard the generator. The concept discussed in this and the preceding paragraph is discussed further at the conclusion of Example 7.8.

If one of the well-known simulation languages or random-number generators is used, it is probably unnecessary to apply the tests. However, random- number generators frequently are added to software that is not specifically developed for simulation, such as spreadsheet programs, symbolic/numerical calculators, and programming languages. If the generator that is at hand is not explicitly known or documented, then the tests should be applied to many samples of numbers from the generator. Some additional tests that are commonly used, but are not covered here, are Good's serial test for sampling numbers [1953, 1967], the median-spectrum test [Cox and Lewis, 1966; Durbin, 1967]. The runs test [Law and Kelton 2000] and a variance heterogeneity test [Cox and Lewis, 1966]. Even if a set of numbers passes all the tests, there is no guarantee of randomness; it is always possible that some underlying pattern has gone undetected.

Although few simulation analyses will need to perform these tests, every simulation user shculd be aware of the qualities of a good random-number generator.

FREQUENCY TESTS

A basic test that should always be performed to validate a new generator is the test of uniformity. Two different methods of testing are available. They are the Kolmogorov-Smirnov and the chi-square test. Both of these tests measure the degree of agreement between the distribution of a sample of generated random numbers and the theoretical uniform distribution and the theoretical distribution.

1. The Kolmogorov-Simirnov test. This test compares the continuous cdf. F(x), of the uniform distribution with the empirical cdf, $S_N(x)$, of the sample of N observations. By definition.

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

If the sample from the random-number generator is R_1, R_2.........R_N, then the empirical cdf, $S_N(x)$, is defined by

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

As becomes larger, $S_N(x)$ should become a better approximation to F(x), provided that the null hypothesis is true.

The cdf of an empirical distribution is a step function with jumps at each observed value.

The Kolmogorv-Smirnov test is based on the largest absolute deviation between F(x) and $S_N(x)$ over the range of the random variable -that is, it is based on the statistic

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

The sampling distribution of D is known; it is tabulated as a function of N in Table A1 (Appendix). For testing against a uniform cdf, the test procedure follows these steps:

Step1. Rank the data from smallest to largest. Let $R_{(i)}$, denote the *i*th smallest observation, so that

$$R_{(1)} \le R_{(2)} \le \cdots R_{(N)}$$

Step2. Compute $D^+ = \max_{1 \le i \le N} \left\{ \frac{i}{N} - R_{(i)} \right\}$

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

Step3. Compute $D = \max(D^+, D^-)$.

Step4. Locate in Table A1 (Appendix) the critical value, D_α, 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_α, the null hypothesis that the data are a sample from a uniform distribution is rejected. If $D \le D_\alpha$, conclude that no difference has been detected between the true distribution of $\{R_1, R_2, \ldots, R_N\}$ and the uniform distribution.

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 by using the Kolmogorov- Smirnov test with the level of significance $\alpha = 0.05$. First, the numbers must be ranked from smallest to largest. The calculations can be facilitated by use of Table 5.1. The top row lists the numbers from smallest ($R_{(1)}$). The computations for D^+, namely $i/N - R_{(i)}$, and for D^-, namely $R_{(i)} -, (i-1)/N$, are easily accomplished by using Table A2 (Appendix) The statistics are computed as $D^+ = 0.26$ and $D^- = 0.21$. Therefore, $D = \max\{0.26, 0.21\} = 0.26$. The critical value of D, obtained from Table A.8 for $\alpha = 0.05$ and B = 5, is 0.565. Since the computed value, 026. is less than the tabulated critical value, 0.565, the hypothesis that the distribution of the generated numbers is the uniform distribution is not rejected.

Table 5.1 Calculation for Kologorv–Smirnov Test

$R_{(i)}$	0.05	0.14	0.44	0.81	0.93
i/N	0.20	0.40	0.60	0.80	1.00
$i/N - R_{(i)}$	0.15	0.26	0.16	–	0.07
$R_{(i)} - (i-1)/N$	0.05	–	0.04	0.21	0.13

2. The chi-square test. The chi-square test uses the sample statistic

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

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 lasses. For the uniform distribution, E_i, the expected number in each class is given by

$$E_i = \frac{N}{n}$$

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.

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 degree of freedom.

Example: Use the chi-square test with $a = 0.05$ to test whether the data shown below are uniformly distributed. Table 7.3 contains the essential computations. The test uses n = 10 intervals of equal length, namely [0, 0.1), [0.1, 0.2), … , [0.9, 1.0).

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

Interval	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	4	16	1.6
9	10	10	0	0	0.0
10	11	10	1	1	0.1

$$\chi_0^2 = \sum_{i=1}^{n} \frac{(O_i - E_i)^2}{E_i} = 3.4$$

The value of X_0^2 is 3.4. This is compared with the critical value $X^2_{0.05,9}$ = 16.9. Since X_0^2 is much smaller than the tabulated value of $X^2_{0.05,9}$, the null hypothesis of a uniform distribution is not rejected.

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 on which 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 discovering 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 could alternate from very high to very low.

The test to be described shortly requires the computation of the autocorrelation between every m numbers (*m* is also known as the lag), starting with the *i*th number. Thus, the autocorrelation ρ_{im} between the following numbers would be of interest: R_i, R_{i+2m} ...., $R_i + (M+1)m$. The value M is largest integer such that $i + (M+1)m \le N$, where N is the total number of values in the sequence. (Thus, a subsequence of length M + 2 is being tested).

A nonzero autocorrelation implies a lack of independence, so the following two-tailed test is appropriate:

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

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

For large value of M, the distribution of the estimator of ρ_{im}, denoted $\hat{r}_{im}$, is approximately normal if the values $R_i, R_i + m, R_i + 2m$ ...., $R_i + (M+1)m$ are uncorrelated. Then the test statistic can be formed as follows:

$$Z_0 = \frac{\hat{\rho}_{im}}{\sigma_{\hat{r}im}}$$

Which is distributed normally with a mean of zero and a variance of 1, under the assumption of independence, for large M.

The formula for $\hat{r}_{im}$, in slightly different form,, and the standard deviation of the estimator, $\sigma_{\hat{r}im}$ are given by Schmidt and Taylor [1970] as follows:

and

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

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

After computing Z_0, do not reject the null hypothesis of independence if $-Z_{\alpha/2} \le Z_0 \le Z_{\alpha/2}$ where α is the level of significance and $Z_{\alpha/2}$ is obtained from Table A3 (Appendix) Figure 5.2 illustrations this test.

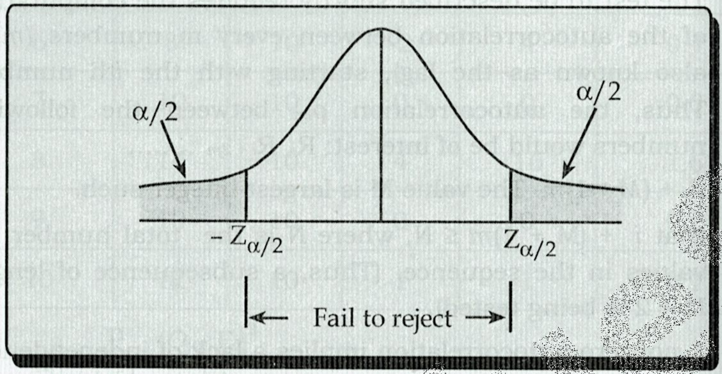


Fig: 5.2: Failure to reject hypothesis

If $\rho_{im} > 0$, the subsequence is said to exhibit positive autocorrelation. In this cash, successive values at lag m have a higher probability than expected of being close in value (i.e. high random numbers in the subsequence followed by high and low followed by low). On the other hand, if $\rho_{im} < 0$, the subsequence is exhibiting negative autocorrelation, which means that low random numbers tend to be followed by high ones, and vice versa. The desired property, independence (which implies zero autocorrelation), means that there is no discernible relationship of the nature discussed here between successive random numbers at lag m.

Example

Test for whether the 3rd, 8th, 13th, and so on, numbers in the sequence at the beginning of this section are auto correlated using $\alpha = 0.05$. Here, $i = 3$ (beginning with the third number), $m = 5$ (every five numbers), $N = 30$ (30 numbers in the sequence), and $M = 4$ (largest integer such that $3 + (M+1)\,5 \le 30$. then.,

$$\hat{\rho}_{35} = \frac{1}{4+1}\,[\ (0.23)(0.28) + (0.28)(0.33) + (0.33)(0.27) + (0.27)(0.05) + (0.05)(0.36)] - 0.25$$

and

$$\sigma_{\hat{\rho}_{35}} = \frac{\sqrt{13(4)+7}}{12(4+1)} = 0.1280$$

Then, the test statistic assumes the value

$$Z_0 = \frac{0.1945}{0.1280} = -1.516$$

Now, the critical value from Table A.3 is

$Z_{0.025} = 1.96$

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

It can be observed that this test is not very sensitive for small values of M, Particularly when the numbers being tested are on the low size. Imagine what would happen if each of the entries in the foregoing computation of $\hat{\rho}_{im}$ were equal to zero. Then $\hat{\rho}_{im}$ would be equal to – 025 and the calculated Z would have the value of – 1.95, not quite enough to reject the hypothesis of independence.

There are many sequences that can be formed in a set of data, given a large value of N. For example, beginning with the first number in the sequence, possibilities include (i) the sequence of all numbers, (2) the numbers, and so on. If $\alpha = 0.05$, there is a probability of 0.05 of rejecting a true hypothesis. If 10 independent or 0.60. Thus, 40% of the time significant autocorrelation would be detected when it does not exist. If α is 0.10 and 10 tests are conducted, there is a 65% chance of finding autocorrelation by change alone. In conclusion, in "fishing" for autocorrelation by performing numerous tests, autocorrelation might eventually be detected, perhaps by chance alone, even when there is no autocorrelation present.

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	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

To facilitate the analysis, the digit 3 has been underlined. There are eighteen 3's in the list. Thus, only 17 gaps can occur. The first gap is of length 10. The second gap is of length 7, and so on. The frequency of the gaps is of interest. The probability of the first gap is determined as follows:

10 of these terms

P(gapof 10) = P(no3)---P(no3)P(3) = $(0.9)^{10}$ (0.1)

Since the probability that any digit is not a 3 is 0.9, and the probability that any digit is a 3 is 0.1.

In general,

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

The theoretical frequency distribution for randomly ordered digits is given by

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

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 (7.14)based on the selected class interval width.

Step 2. 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 Equation (7.3).

Step 4. Determine the critical value, D_a, from Table A.8 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: Based on the frequency with which gaps occur, analyze the 110 digits above to test whether they are independent. Use a = 0.05. 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
Number of gaps	7	8	8	17	10	13	7	8	9	13

The gap test is presented in Table .The critical value of D is given by

Gap Length	Frequency	Relative Frequency	Cumulative Frequency (Snx)	Function $Fx=1-0.9^{x+1}$	$\lvert F(x)-S_N(x)\rvert$
0-3	35	0.35	0.35	0.3439	0.0061
4-1	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	0.97	0.9657	0.0043
32-35	0	0.0	0.97	0.9775	0.0075
36-39	2	0.02	0.99	0.9852	0.0043
40-43	0	0.0	0.99	0.9903	0.0003

The gap test is presented in Table 7.6. The critical value of D is given by

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

$$= 0.136$$

Since D = max $\lvert F(x) - S_N(x)\rvert$ = 0.0224 is less than $D_{0.05}$ do not reject the hypothesis of independence on the basis of this test.

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, ...

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

Alternatively, the last result can be obtained as follows:

P(exactly one pair) = (32) (0.1)(0.9) = 0.27

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 a = 0.05. The test is summarized in following Table:

Combinations	Observed Frequency Oi	Expected Frequency Ei (Total N * Probability)	$(Oi-Ei)^2/Ei$
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 = 5.99. The independence of the numbers is rejected on the basis of this test.

For four and five digit Poker test students can perform as a class task.

Random - Variate Generation

The chapter deals with procedures for sampling from a variety of widely-used continuous and discrete distributions. Previous discussions and examples indicated the usefulness of statistical distributions in modeling activates that are generally unpredictable or uncertain. For example, inter arrival times and service times and service times at queues and demands for a product are quite often unpredictable in nature, at least to a certain extent. Usually, such variables are modeled as random variables with some specified statistical distribution, and standard statistical procedures exist for estimating the parameters of the hypothesized distribution and for testing the validity of the assumed statistical model.

In this chapter, it is assumed that a distribution has been completely specified, and ways are sought to generate samples from this distribution to be used as input to a simulation model. The purpose of the chapter is to explain and illustrate some widely-used techniques for generating random variates, not to give a state-of-the-art survey of the most efficient techniques. In Practice, most simulation modelers will use either existing routines available in programming libraries do not have built into the simulation language being used. However, some programming languages do not have built-in routines for all of the regularly used distributions, and some computer installations do not have random-variate-generation libraries; in such cases the modeler must construct an acceptable routine generation occurs.

This chapter discuses the inverse = transform techniques and, more briefly, the acceptance-rejection technique and special properties. Another techniques, the composition method, is discussed by Devroye [1986], Dagpunar [1980], Fishman [1998], and Law and Kelton [2000]. All the techniques in this chapter

assume that a source of uniform [0, 1] random numbers, R_1 , R_2 .....is readily available, where each R_i has pdf and cdf

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

$$F_R(x) = \begin{cases} 0, & x < 0 \\ x, & 0 \le x \le 1 \\ 1, & x > 1 \end{cases}$$

Throughout this chapter R and R_1, R_2......represent random numbers uniformly distributed on [0,1] and generated by one of the techniques discussed in previous sections.

Inverse-Transform Technique

The inverse-transform techniques can be used to sample from the exponential, the uniform, the Weibull, the triangular distributions and from empirical distributions. Additionally, it is the underlying principle for sampling from a wide variety of discrete distributions. The techniques will be explained in detail for the exponential distribution and then applied to other distributions. Computational, it is the straightest forward, but not always the most efficient, techniques.

Uniform Distribution

Consider a random variable X that is uniformly distributed on the interval [a,b]. A reasonable guess for generating X is given by

$$X = a + (b - a) R \qquad 8.5$$

[Recall that R is always a random number on [0, 1]. The pdf of X is given by

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

The derivation of Equation (8.5) follows Steps 1 through 3 of Section 8.1.1:

Step 1. The cdf is given by

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

Step 2. Set $F(x) = (x - a)/(b - a) = R$

Step 3. Solving for X in terms of R yields $X = a + (b - a) R$, which agrees with Equation (8.5).

Acceptance–Rejection Technique

Suppose that an analyst needed to devise a method for generating random variates, X uniformly distributed between ¼ and 1. One way to proceed would be to follow these steps:

Step 1 Generate a random number R.

Step 2a. If $R \geq 1/4$, accept X = R, then go to step3.

Step 2b. If $R < 1/4$, reject R, and return to Step 1.

Step 3. If another uniform random variate on [1/4, 1] is needed, repeat the procedure beginning at Step 1. If not, stop.

Each time Step 1 is executed, a new random number R must be generated. Step 2a is an "acceptance" and **Step 2b**. is a "rejection" in this acceptance - rejection technique. To summarize the technique, random variates (R) with some distribution (here uniform on [0,1]) are generated until some condition $\left(R > \frac{1}{4}\right)$ is satisfied. When the condition is finally satisfied, the desired random variate, X (here uniform on $\left(\frac{1}{4}, 1\right)$, can be computed (X = R). This procedure can be shown to be correct by recognizing that the accepted values of R are conditioned value; that is, R itself does not have the desired distribution, but R conditioned on the event $\left\{R \geq \frac{1}{4}\right\}$ does have the desired distribution. To show this, take $\frac{1}{4} \leq a < b \leq 1$; then

$$P\left(a < R \leq b \,|\, \frac{1}{4} \leq R \leq 1\right) = \frac{P(a < R \leq b)}{P\left(\frac{1}{4} \leq R \leq 1\right)}$$

Which is the correct probability for a uniform distribution on $\left[\frac{1}{4}, 1\right]$. Equation (8.21) says that h the probability distribution of R, given that R is between ¼ and 1 (all other values of R are thrown out), is the desired is distribution. Therefore, if $\frac{1}{4} \leq R \leq 1$, set X = R.

The efficiency of an acceptance - rejection technique depend heavily on being able to minimize the number of rejections. In this example, the probability of a rejection is $P\left(R > \frac{1}{4}\right) = \frac{1}{4}$, so that the number of rejections is a geometrically distributed random variable with probability of "Success" being p 3/4 and mean umber of rejection $\left(\frac{1}{p} - 1\right) = \frac{4}{3} - 1 = \frac{1}{3}$. (Example 8.6 discussed the geometric distribution.) The mean number of random numbers R required to generate one variate X is one more than the number of rejections; hence, it is $\frac{4}{3} = 1.33$. In other words to generate 1--- values of X would require approximately 1222 random numbers R.

In the present situation, an alternative procedure exists for generating a uniform variate on [1/4, 1] – namely, Equation (8.5), which reduces to $X = 1/4 + (3/4)R$. Whether the acceptance - rejection technique or an alternative producer, such the inverse-transform techniques [Equation 8.5], is the more efficient depends on several considerations. The computer being used, the skills of the programmer and the relative inefficiency of generating the additional (rejected) random numbers needed by acceptance-rejection should be compared to the computations required by the alternative procedure. In practice, concern with generation efficiency is left to specialists who conduct extensive tests comparing alternative methods. (i.e.. , until a simulation model begins to require excessive computer runtime due to the generator being used).

For the uniform distribution on [1/4, 1], h the inverse-transform techniques of Equation (8.5) is undoubtedly much easier to apply and more efficient than the acceptance-rejection techniques. The main purpose of this example was to explain and motivate the basic concept of the acceptance -rejection technique. However, for some importance distributions, such as the normal, gamma and beta, the inverse cdf does not exist in closed form and therefore the inverse-transform technique is difficult.

DISCUSSION EXERCISE

1. Explain the properties of random number & its consequences.
2. Explain the generation of Pseudo-random Numbers.
3. Explain the linear congruential method for random number generation?
4. Explain the combined linear congruential random number generation method?
5. What is the role of maximum density and maximum period in random number generation?

6. Generate a sequence of 15 random numbers for which seed is 342, constant multiplier is 20,increment is 45 and modulus is 30
7. Explain with an example the Kolmogorov-Smirnov test for random numbers.
8. Explain with an example the chi-square test for random numbers?
9. Explain auto correlation Test for random numbers.
10. Using the principles learnt, develop your own combined linear congruential random number generator.
11. What is inverse transform technique? Explain how it is used for producing random variants for exponential distribution and uniform distribution.
12. 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	555
One pair	389
Two pairs	36
Three digits of a kind	18
Four digits of a kind	2
Total	1000

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

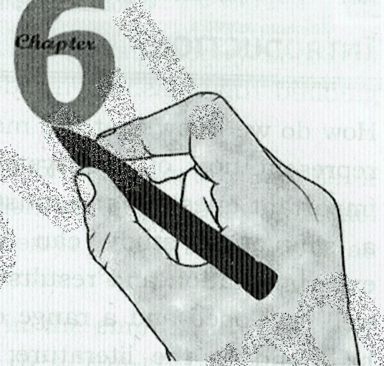
13. Explain inverse transform method for random number generation.
14. Why testing of random number is necessary. Describe the algorithm for Gap testing with the example of your own.

15. What is the main difference between real random number and pseudo random number? Use the multiplicative congruential method to generate a sequence of four digit random numbers. Let X0 = 315, a = 3 and m = 1000.

16. What are the properties of random number? The sequence of numbers 0.44, 0.83, 0.78, 0.10 and 0.58 has been generated. Use the Kolmogorov–Smirnov test α = 0.05 to determine if the hypothesis that the numbers are uniformly distributed on the interval 0 to 1 can be rejected. (Note that the critical value of D for α = 0.05 and N = 5 is 0.565)

17. Explain the process of testing for auto-correlation test.

❑❑❑



VERIFICATION AND VALIDATION

CHAPTER OUTLINE

After studying this chapter, the students will be able to understand the

- Design of Simulation Models
- Verification of Simulation Models
- Calibration and Validation of the models
- Three-Step Approach for Validation of Simulation Models
- Accreditation of Model