UNIT - 5 : RANDOM-NUMBER GENERATION, RANDOM-VARIATE GENERATION: Properties of random numbers; Generation of pseudo-random numbers; Techniques for generating random numbers; Tests for Random Numbers. Random-Variate Generation: Inverse transform technique; Acceptance-Rejection technique; Special properties.8 Hours

RANDOM-NUMBER GENERATION

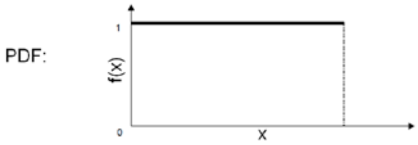
Random numbers are a necessary basic ingredient in the simulation of almost all discrete systems. Most computer languages have a subroutine, object, or function that will generate a random number. Similarly simulation languages generate random numbers that arc used to generate event times and other random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

5.1 Properties of Random Numbers

A sequence of random numbers, R1, R2... must have two important statistical properties, uniformity and independence. Each random number *Ri,* is an independent sample drawn from a continuous uniform distribution between zero and 1. That is, the pdf is given by

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

The density function is shown below:



The expected value of Ri, is

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

The variance is given by 0

$$V(R) = \int_0^1 x^2 dx - [E(R)]^2$$

$$= [x^3/3]_0^1 - (1/2)^2 = 1/3 - 1/4$$

$$= 1/12$$

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 m each interval ii N/n where A' is the total number of observations.
2. The probability of observing a value in a particular interval is of the previous values drawn.

5.2 Generation of 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 initiates, the ideal properties of uniform distribution and independence as closely as possible.

When generating 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 include 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.

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**

5.3 Techniques for Generating Random Numbers

The linear congruential method

It widely used technique, initially proposed by Lehmer [1951], produces a sequence of integers, X1, X2,... between zero and m — 1 according to the following recursive relationship:

$$X_{i+1} = (aX_i + c) \bmod m, \quad i = 0, 1, 2.... \quad (7.1)$$

The initial value **X0** is called the seed, **a** is called the constant multiplier, **c** is the increment, and **m** is the modulus.
If c ≠ 0 in Equation (7.1), 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 X0 drastically affects the statistical properties and the cycle length. 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 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

$$R_i = X_i/m, \quad i= 1,2,\ldots\ldots \quad (7.2)$$

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

X0 = 27
X_1 = (17*27 + 43) mod 100 = 502 mod 100 = 2
R_1=2/100=0. 02
X_2 = (17*2 + 43) mod 100 = 77 mod 100 = 77
R_2=77 /100=0. 77
X_3 = (17*77+ 43) mod 100 = 1352 mod 100 = 52
R_3=52 /100=0. 52

Second, to help achieve maximum density, and to 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_0.

The max period (P) is:

- For m a power of 2, say m = 2^b, and c ^{1}≠0, the longest possible period is P = m = 2^b, which is achieved provided that c is relatively prime to m (that is, 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 provided that the seed X_0 is odd and 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 provided that the multiplier, a, has the property that the smallest integer k such that $a^k - 1$ is divisible by m is k = m – 1.

Multiplicative Congruential Method:

Basic Relationship:

$$\mathbf{X_{i+1} = a\,X_i \ (\text{mod } m), \text{ where } a \geq 0 \text{ and } m \geq 0 \quad \ldots (7.3)}$$

Most natural choice for **m** is one that equals to the capacity of a computer word.
$m = 2^b$ (binary machine), where b is the number of bits in the computer word.
$m = 10^d$ (decimal machine), where d is the number of digits in the computer word.

Example 2: Using the multiplicative congruential method, find the period of the generator for a = 13, $m = 2^6 = 64$, and X_0 = 1, 2, 3, and 4. When the seed is 1 and 3, the sequence has period 16. However, a period of length eight is achieved when the seed is 2 and a period of length four occurs when the seed is 4.

Period Determination Using Various seeds

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

EXAMPLE 3 Let m = 102 = 100, a = 19, c = 0, and X0 = 63, and generate a sequence c random integers using Equation (7.1).

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 Let $a = 7^5$ = 16,807, $m = 2^{31}-1$ = 2,147,483,647 (a prime number), and c= 0. These choices satisfy the conditions that insure a period of P = m-1. Further, specify a seed, X_0 = 123,457.
The first few numbers generated are as follows:

$X_1 = 7^5(123{,}457) \text{ mod } (2^{31} - 1) = 2{,}074{,}941{,}799 \text{ mod } (2^{31} - 1)$
$X_1 = 2{,}074{,}941{,}799$ $R_1 = X_1 / 2^{31}$

$X_2 = 7^5(2{,}074{,}941{,}799) \bmod (2^{31} - 1) = 559{,}872{,}160$

$R_2 = X_2 / 2^{31} = 0.2607$

$X_3 = 7^5(559{,}872{,}160) \bmod (2^{31} - 1) = 1{,}645{,}535{,}613$

$R_3 = X_3 / 2^{31} = 0.7662$

...

Notice that this routine divides by m + 1 instead of m; however, for such a large value of m, the effect is negligible.

Combined Linear Congruential Generators

As computing power has increased, the complexity of the systems that we are able to simulate has also increased. One fruitful approach is to combine two or more multiplicative congruential generators in such a way that the combined generator has good statistical properties and a longer period.

The following result from L'Ecuyer [1988] suggests how this can be done:

If $W_{i,1}$, $W_{i,2}$,... , $W_{i,k}$ are any independent, discrete-valued random variables (not necessarily identically distributed), but one of them, say $W_{i,1}$, is uniformly distributed on the integers 0 to m_i — 2, then

$$W_i = \left(\sum_{j=1}^{k}(-1)^{j-1}W_{i,j}\right) \bmod m_1 - 1$$

is uniformly distributed on the integers 0 to m_i — 2.

To see how this result can be used to form combined generators, let $X_{i,1}$, $X_{i,2}$,..., $X_{i,k}$ be the *i* th output from k different multiplicative congruential generators, where the *j* th generator has prime modulus m_j, and the multiplier a_j is chosen so that the period is m_j — 1. Then the j'th generator is producing integers $X_{i,j}$ that are approximately uniformly distributed on 1 to m_j - 1, and $W_{i,j} = X_{i,j}$ — 1 is approximately uniformly distributed on 0 to m_j - 2. L'Ecuyer [1988] therefore suggests combined generators of the form

$$Xi = \left(\sum_{j=1}^{k}(-1)^{j-1}X_{i,j}\right) \bmod m_1 - 1$$

With

$$Ri = \begin{cases} \dfrac{X_i}{m_1}, X_i > 0 \\ \dfrac{m_1 - 1}{m_1}, X_i = 0 \end{cases}$$

Notice that the "$(-1)^{j-1}$"coefficient implicitly performs the subtraction $X_{i,1}$-1; for example, if k = 2, then

$$(-1)^0\left(X_{i,1} - 1\right) - (-1)^1\left(X_{i,2} - 1\right) = \sum_{j=1}^{2}(-1)^{j-1}X_{i,j}$$

The maximum possible period for such a generator is

$$p = \frac{(m_1 - 1)(m_2 - 1)\ldots(m_k - 1)}{2^{k-1}}$$

EXAMPLE 5 For 32-bit computers, L'Ecuyer [1988] suggests combining k = 2 generators with m1 = 2147483563, a1= 40014, m2 = 2147483399, and a2 = 40692.

This leads to the following algorithm:

1) Select seed $X_{1,0}$ in the range [1, 2,14,74,83,562] for the first generator, and seed $X_{2,0}$ in the range [1, 2,147,483,399]. Set j =0.
2) Evaluate each individual generator.
   i) $X_{1,j+1}$ = 40014 $X_{1,j}$ mod 2,147,483,563
   ii) $X_{2,j+i}$ = 40692 $X_{2,j}$ mod 2,147,483,399
3) Set X_{j+1} = ($X_{1,j+1}$ - $X_{2,j+1}$) mod 2,147,483,562.
4) Return

$$Ri = \begin{cases} \dfrac{X_{j+1}}{2{,}147{,}483{,}563}, X_{j+1} > 0 \\ \dfrac{2{,}147{,}483{,}562}{2{,}147{,}483{,}563}, X_{j+1} = 0 \end{cases}$$

5) Set *j* = *j* + 1 and go to step 2.

Tests For Random Numbers

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. ***Runs test***. Tests the runs up and down or the runs above and below the mean by comparing the actual values to expected values. The statistic for comparison is the chi-square.
3. ***Autocorrelation test***. Tests the correlation between numbers and compares the sample correlation to the expected correlation of zero.
4. ***Gap test***. Counts the number of digits that appear between repetitions of a particular digit and then uses the Kolmogorov-Smirnov test to compare with the expected number of gaps.
5. ***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.

   In testing for uniformity, the hypotheses are as follows:

   H_0: R_i ~ U[0,1]

   H_1: R_i ≁ U[0,1]

   The null hypothesis, H_0, reads that the numbers are distributed uniformly on the interval [0, 1].

   In testing for independence, the hypotheses are as follows;

   H_0: R_i ~ independently

   H_1: R_i ≁ independently

   This null hypothesis, H_0, reads that the numbers are independent. Failure to reject the null hypothesis means that no evidence of dependence has been detected on the basis of this test. This does not imply that further testing of the generator for independence is unnecessary.

   Level of significance a

   a = P (reject H_0 | H_0 true)

   Frequently, a is set to 0.01 or 0.05

   (Hypothesis)

	Actually True	Actually False
Accept	1 - α	β (Type II error)
Reject	α (Type I error)	1- β

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.

1. Kolmogorov-Smirnov and
2. 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.
- Both tests are on the null hypothesis of no significant difference between the sample distribution and the theoretical distribution.

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,

$$\mathbf{F(x) = x, 0 \le x \le 1}$$

If the sample from the random-number generator is R1 R2, ,..., RN, then the empirical cdf, SN(x), is defined by

$$S_n(x)\frac{\text{number of }\ \mathrm{R1, R2, \ldots, Rn}\text{ which are} \le \mathrm{x}}{N}$$

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

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

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

Step 1: Rank the data from smallest to largest. Let R (i) denote the i th smallest observation, so that

$$\mathbf{R_{(1)} \le R_{(2)} \le \ldots \le R_{(N)}}$$

Step 2: Compute

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

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

Step 3: Compute D = max (D+, D-).

Step 4: Determine the critical value, **D_α,** from **Table A.8** for the specified significance level α and the given sample size N.

Step 5: $D \le D_\alpha$ Accept: No Difference between **S_N(x) and F(x)**

$D > D_\alpha$ Reject: No Difference between **S_N(x) and F(x)**

We conclude that no difference has been detected between the true distribution of {R_1, R_2,... R_N} and the uniform distribution.

EXAMPLE 6: 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 **α of 0.05.**

Step 1: Rank the data from smallest to largest.

0.05, 0.14, 0.44, 0.81, 0.93

Step 2: Compute **D^+ and D^-**

	R_i	$\frac{i}{N}$	$D^{+} = \max_{1 \le i \le n} \left\{ \frac{i}{N} - R_{(i)} \right\}$	$D^{-} = \max_{1 \le i \le n} \left\{ R_{(i)} - \frac{i-1}{N} \right\}$
1	0.05	0.20	0.15	0.05
2	0.14	0.40	0.26	~
3	0.44	0.60	0.16	0.04
4	0.81	0.80	~	0.21
5	0.93	1.00	0.07	0.13

Step3: Compute D = max (D+, D-).

D=max (0.26, 0.21) = 0.26

Step 4: Determine the critical value, **D_α,** from Table A.8 for the specified significance level α and the given sample size N.

Here α=0.05, N=5 then value of **D_α = 0.565**

Step 5: 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. The chi-square test.

The chi-square test uses the sample statistic

$$\chi_0^2 = \sum_{i=0}^{n} \frac{(O_i - E_i)^2}{E_i}$$

Where,

O_i is observed number in the *i* th class

E_i is expected number in the *i* th class,

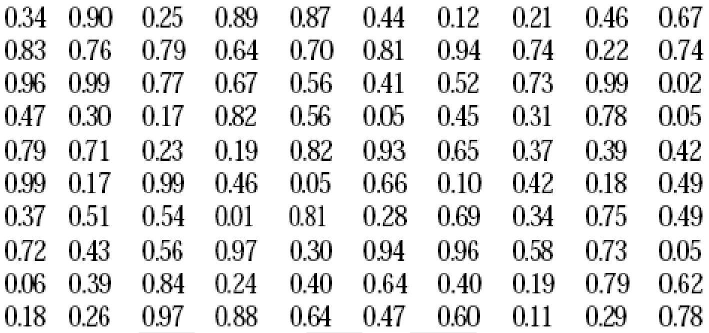
$$E_i = \frac{N}{n}$$

N – No. of observation

n – No. of classes

Note: sampling distribution of χ_0^2 is approximately the chi square has n-1 degrees of freedom

Example 7: Use the chi-square test with α = 0.05 to test whether the data shown below are uniformly distributed. The test uses n = 10 intervals of equal length, namely [0, 0.1), [0.1, 0.2)... [0.9, 1.0). **(REFER TABLE A.6)**



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

Computations for chi square test

Interval	Range	O_i	E_i	$O_i - E_i$	$(O_i - E_i)^2$	$\frac{(O_i - E_i)^2}{E_i}$
1	0.0-0.1	8	10	-2	4	0.4
2	0.1-0.2	8	10	-2	4	0.4
3	0.2-0.3	10	10	0	0	0.0
4	0.3-0.4	9	10	-1	1	0.1
5	0.4-0.5	12	10	2	4	0.4
6	0.5-0.6	8	10	-2	4	0.4
7	0.6-0.7	10	10	0	0	0.0
8	0.7-0.8	14	10	4	16	1.6
9	0.8-0.9	10	10	0	0	0.0
10	0.9-1.0	11	10	1	1	0.1
		100	100	0		3.4

The value of χ_0^2 is 3.4. This is compared with the critical value $\chi_{0.05,9}^2$ = 16.9.Since χ_0^2 is much smaller than the tabulated value of $\chi_{0.05,9}^2$, the null hypothesis of a uniform distribution is not rejected.

Run Tests (Up and Down)

The runs test examines the arrangement of numbers in a sequence to test the hypothesis of independence.

A run is defined as a succession of similar events preceded and followed by a different event.

E.g. in a sequence of tosses of a coin, we may have

H T T H H T T T H T

The first toss is preceded and the last toss is followed by a "no event". This sequence has six runs, first with a length of one, second and third with length two, fourth length three, fifth and sixth length one.

A few features of a run two characteristics:

1. number of runs and the length of run
2. an up run is a sequence of numbers each of which is succeeded by a larger number; a down run is a squence of numbers each of which is succeeded by a smaller number

Consider the 40 numbers; both the Kolmogorov-Smirnov and Chi-square would indicate that the numbers are uniformly distributed. But, not so

0.08	0.09	0.23	0.29	0.42	0.55	0.58	0.72	0.89	0.91
0.11	0.16	0.18	0.31	0.41	0.53	0.71	0.73	0.74	0.84
0.02	0.09	0.30	0.32	0.45	0.47	0.69	0.74	0.91	0.95
0.12	0.13	0.29	0.36	0.38	0.54	0.68	0.86	0.88	0.91

Now, rearrange and there is less reason to doubt independence.

0.41	0.68	0.89	0.84	0.74	0.91	0.55	0.71	0.36	0.30
0.09	0.72	0.86	0.08	0.54	0.02	0.11	0.29	0.16	0.18
0.88	0.91	0.95	0.69	0.09	0.38	0.23	0.32	0.91	0.53
0.31	0.42	0.73	0.12	0.74	0.45	0.13	0.47	0.58	0.29

Concerns:

- Number of runs
- Length of runs

Note: If N is the number of numbers in a sequence, the maximum number of runs is N-1, and the minimum number of runs is one.

If “a” is the total number of runs in a sequence, the mean and variance of “a” is given by

$$\mu_a = (2N - 1) / 3$$

$$\sigma_a^2 = (16N - 29) / 90$$

For N > 20, the distribution of “a” approximated by a normal distribution, N (μ_a, σ_a^2).

This approximation can be used to test the independence of numbers from a generator.

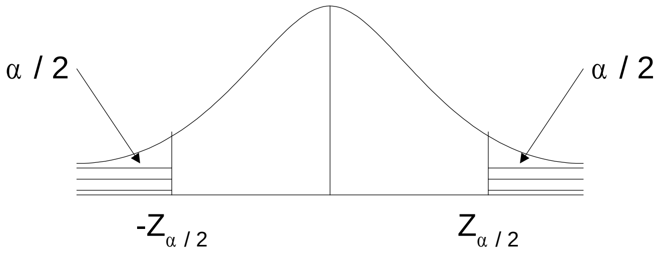
$$Z_0 = (a - \mu_a) / \sigma_a$$

Substituting for μ_a and σ_a ==>

$$Z_a = \frac{\{a - [(2N-1)/3]\}}{\{\sqrt{(16N-29)/90}\}}$$

where $Z_a \sim N(0,1)$

Acceptance region for hypothesis of independence $-Z_{\alpha/2} \le Z_0 \le Z_{\alpha/2}$



Example 8: Based on runs up and runs down, determine whether the following sequence of 40 numbers is such that the hypothesis of independence can be rejected where a = 0.05.

0.41 0.68 0.89 0.94 0.74 0.91 0.55 0.62 0.36 0.27
0.19 0.72 0.75 0.08 0.54 0.02 0.01 0.36 0.16 0.28
0.18 0.01 0.95 0.69 0.18 0.47 0.23 0.32 0.82 0.53
0.31 0.42 0.73 0.04 0.83 0.45 0.13 0.57 0.63 0.29

The sequence of runs up and down is as follows:

+ + + - + - + - - - + + - + - - + - + - - + - - + - + + - - + + - + - - + + -

There are 26 runs in this sequence. With N=40 and a=26,

μ_a = {2(40) - 1} / 3 = 26.33 and

σ_a^2 = {16(40) - 29} / 90 = 6.79

Then, $Z_0 = (26 - 26.33)/\sqrt{6.79} = -0.13$ Now, the critical value is $Z_{0.025}$ = 1.96 (Hint: $Z_{a/2}$= $Z_{0.5/2}$= $Z_{0.025}$) Refer Table A.3, so the independence of the numbers cannot be rejected on the basis of this test.

Poker Test

It based on the frequency with which certain digits are repeated.

Example 9:

0.255 0.577 0.331 0.414 0.828 0.909

Note: Pair of like digits appear in each number generated.

In 3-digit numbers, there are only 3 possibilities.

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

P(3 different digits) = (2nd diff. from 1st) * P(3rd diff. from 1st & 2nd)
= (0.9) (0.8) = 0.72

P(3 like digits) = (2nd digit same as 1st) * P(3rd digit same as 1st)
= (0.1) (0.1) = 0.01

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

Example 10:

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 next table.

Combination, i	Observed Frequency, O_i	Expected Frequency, E_i	$\frac{(O_i - E_i)^2}{E_i}$
Three different digits	680	720	2.24
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 $\chi^2_{0.05,2}$ = 5.99 < 47.65(Table A.6), the independence of the numbers is rejected on the basis of this test.

The Gap Test

- It measures the number of digits between successive occurrences of the same digit.
- A gap of length x occurs between the recurrence of some digit.

Steps involved in the test.

Step 1: Specify the cdf for the theoretical frequency distribution given by Equation below based on the selected class interval width.

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

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 $\mathbf{S_N(x)}$ as in Equation.

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

Step 4: Determine the critical value, D_α, from Table A.8 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: 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

Note: eighteen 3's in list ==> 17 gaps, the first gap is of length 10

We are interested in the frequency of gaps.

P (gap of 10) = P (not 3) $\times\times\times$ P (not 3) P (3),

Note: there are 10 terms of the type P (not 3)

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

The theoretical frequency distribution for randomly ordered digit is given by

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

Note: observed frequencies for all digits are compared to the theoretical frequency using the Kolmogorov-Smirnov test.

Example 10:

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 digits minus 10, or 100. The numbers of gaps associated with the various digits are as follows:

Digit	0	1	2	3	4	5	6	7	8	9
# of Gaps	7	8	8	17	10	13	7	8	9	13

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

The critical value of D is given by

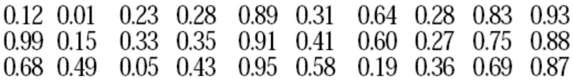
$$D_{0.05} = 1.36 \big/ \sqrt{100} = 0.136$$

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.

Tests for Auto-correlation

The tests for auto-correlation are concerned with the dependence between numbers in a sequence.

The list of the 30 numbers appears to have the effect that every 5th number has a very large value. If this is a regular pattern, we can't really say the sequence is random.



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

The test computes the auto-correlation between every m numbers (m is also known as the lag) starting with the ith number.

Thus the autocorrelation ρ_{im} between the following numbers would be of interest.

$$R_i, R_{i+m}, R_{i+2m}, .., R_{i+(M+1)m}$$

Form the test statistic $Z_0 = \dfrac{\rho_{\hat{i}m}}{\sigma_{\rho_{\hat{i}m}}}$ which is distributed normally with a mean of zero and a variance of one.

The actual formula for $\rho_{\hat{i}m}$ and the standard deviation is $\rho_{\hat{i}m} = \dfrac{1}{M+1}\left[\sum_{k=0}^{M} R_{i+km} R_{(k+1)m}\right] - 0.25$ and

$$\sigma_{\rho_{\hat{i}m}} = \frac{\sqrt{13M+7}}{12(M+1)}$$

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.

EXAMPLE 11: 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), and M = 4 (largest integer such that 3 + (M +1)5 < 30).

$$\rho_{\hat{i}m} = \frac{1}{4+1}\left[(0.23)(0.28)+(0.28)(0.33)+(0.33)(0.27)+(0.27)(0.05)+(0.05)(0.36)\right]-0.25$$
$$= -0.1945$$

And

$$\sigma_{\rho_{\hat{i}m}} = \frac{\sqrt{13(4)+7}}{12(4+1)} = 0.1280$$

Then, test for 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 can't be rejected on the basis of this test.

Random Variate Generation

TECHNIQUES:

- INVERSE TRANSFORMATION TECHNIQUE
- ACCEPTANCE-REJECTION TECHNIQUE

All these techniques assume that a source of uniform (0,1) random numbers is available R1,R2….. where each R1 has probability density function and cumulative distribution function.

$$\text{pdf: } f_R(x) = \begin{cases} 1, & 0 \le x \le 1 \\ 0, & \text{otherwise} \end{cases} \quad \text{and}$$
$$\text{cdf: } F_R(x) = \begin{cases} 0, & x < 0 \\ x, & 0 \le x \le 1 \\ 1, & x > 1 \end{cases}$$

Note: The random variable may be either discrete or continuous.

Inverse Transform Technique

The inverse transform technique can be used to sample from exponential, the uniform, the Weibull and the triangle distributions.

Exponential Distribution

The exponential distribution, has probability density function (pdf) given by

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

and cumulative distribution function (cdf) given by

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

The parameter λ can be interpreted as the mean number of occurrences per time unit. For example, if interarrival times X_1, X_2, X_3 . . . had an exponential distribution with rate, and then could be interpreted as the mean number of arrivals per time unit, or the arrival rate.

For any i,

$E(X_i)$= 1/λ

And so 1/λ is mean interarrival time. The goal here is to develop a procedure for generating values X_1, X_2, X_3 . . . which have an exponential distribution.

The inverse transform technique can be utilized, at least in principle, for any distribution. But it is most useful when the cdf. F(x), is of such simple form that its inverse, F^{-1}, can be easily computed. A step-by-step procedure for the inverse transform technique illustrated by me exponential distribution, is as follows:

Step 1: Compute the cdf of the desired random variable X.

For the exponential distribution, the cdf is **$F(x) = 1-e^{-\lambda x}$, x≥0.**

Step 2: Set F(X) = R on the range of X.

For the exponential distribution, it becomes **$1 - e^{-\lambda X} = R$ on the range x ≥ 0.**

Since X is a random variable (with the exponential distribution in this case), so $1-e^{-\lambda x}$ is also a random variable, here called R. As will be shown later, R has a uniform distribution over the interval (0,1).,

Step 3: Solve the equation F(X) = R for X in terms of R. For the exponential distribution, the solution proceeds as follows:

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

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

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

$$x=-1/\lambda \ln(1 - R) \qquad \text{…(5.1)}$$

Equation (5.1) is called a random-variate generator for the exponential distribution.

In general, Equation (5.1) is written as $X=F^{-1}(R)$. Generating a sequence of values is accomplished through steps 4.

Step 4: Generate (as needed) uniform random numbers R1, R2, R3,... and compute the desired random variates by

$$X_i = F^{-1} (R_i)$$

For the exponential case, **$F^{-1} (R) = (-1/\lambda)\ln(1- R)$** by Equation (5.1), so that

$$X_i = -1/\lambda \ln (1 - R_i) \qquad \text{…(5.2)}$$

for i = 1,2,3,.... One simplification that is usually employed in Equation (5.2) is to replace $1 - R_i$ by R_i to yield

$$X_i = -1/\lambda \ln R_i \quad \text{…(5.3)}$$

which is justified since both Ri and 1- Ri are uniformly distributed on (0,1).

Table 5.1 Generation of Exponential Variates X, with Mean 1, Given Random Numbers R_i,

I	1	2	3	4	5
R_i	0.1306	0.0422	0.6597	0.7965	0.7696
X_i	0.1400	0.0431	1.078	1.592	1.468

Table 5.1 gives a sequence of random numbers from Table A.1 and the computed exponential variates, X_i, given by Equation (5.2) with a value of = 1. Figure 5.1(a) is a histogram of 200 values, R_1, R_2,…R_{200} from the uniform distribution and figure 5.1(b)

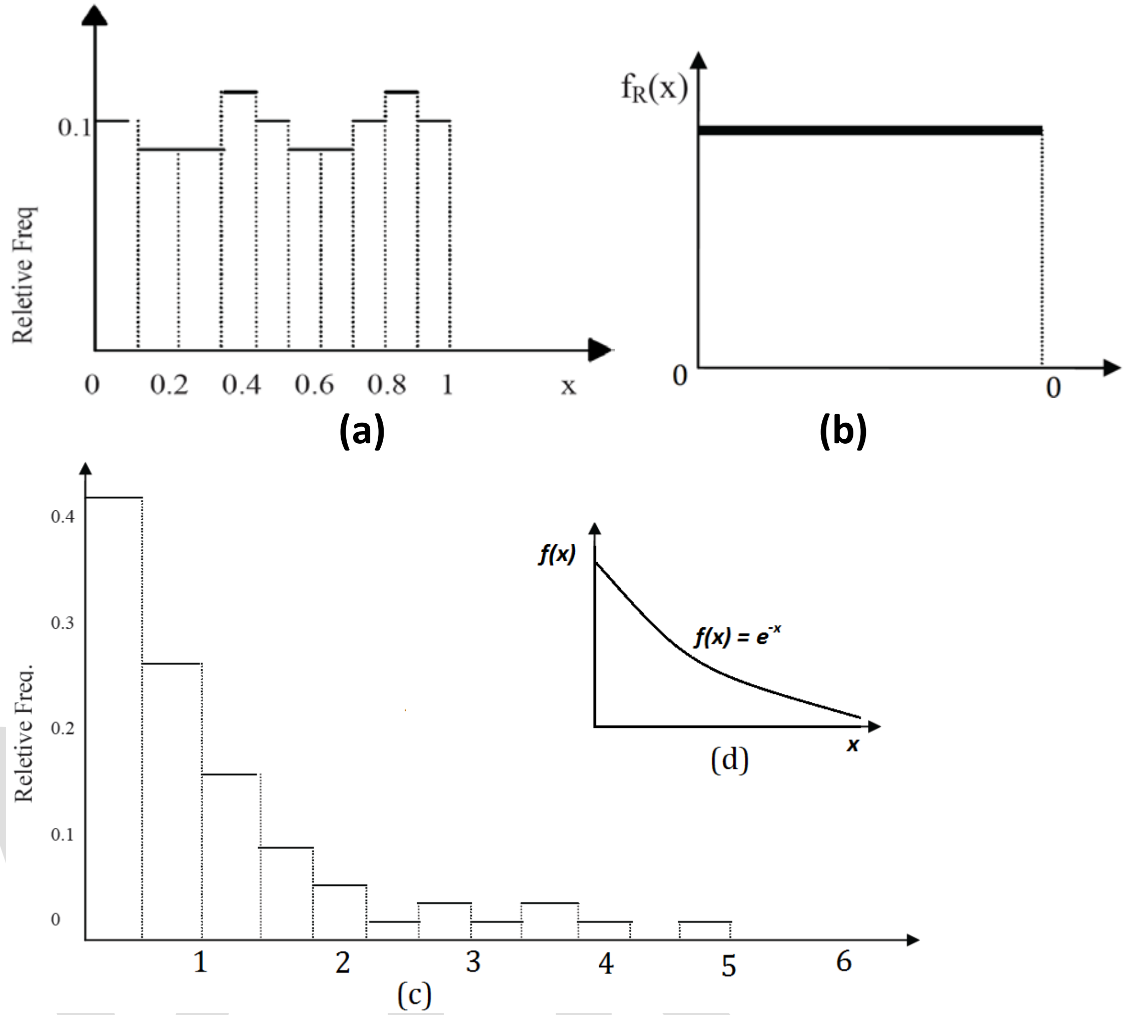


Figure 5.1: (a) Empirical histogram of 200 uniform random numbers; (b) empirical histogram of 200 exponential variates; (c) theoretical uniform density on (0, 1); (d) theoretical exponential density with mean 1.

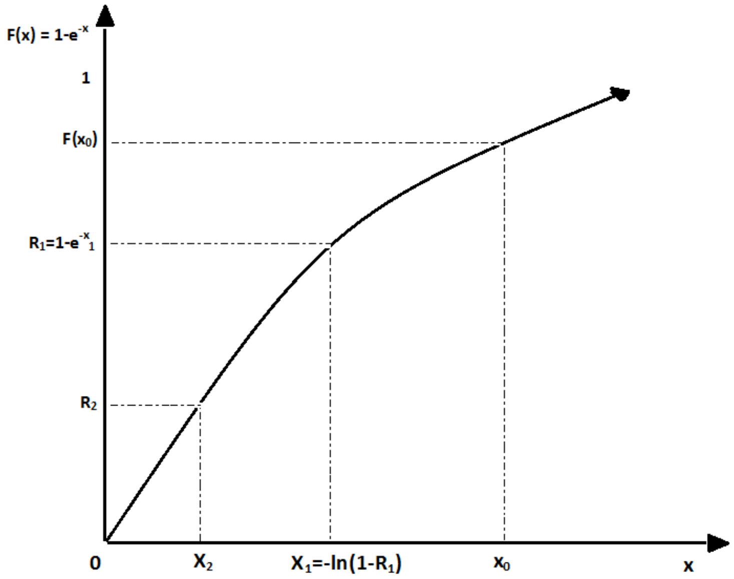


Figure 5.2. Graphical view of the inverse-transform technique

Acceptance-Rejection technique

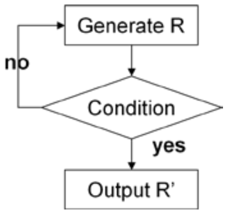
- Useful particularly when inverse cdf does not exist in closed form
- Illustration: To generate random variates, $X \sim U(1/4, 1)$
- Procedures:

Step 1: Generate a random number 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* ³ *¼*} does.
- Efficiency: Depends heavily on the ability to minimize the number of rejections.



Poisson Distribution

A Poisson random variable, N, with mean a > 0 has pmf

$$p(n) = P(N = n) = \frac{e^{-\alpha}\alpha^n}{n!}, \ \ n = 0,1,2,...$$

- N can be interpreted as number of arrivals from a Poisson arrival process during one unit of time
- Then time between the arrivals in the process are exponentially distributed with rate α.
- Thus there is a relationship between the (discrete) Poisson distribution and the (continuous) exponential distribution, namely

$$N = n \ \Leftrightarrow \ \sum_{i=1}^{n} A_i \le 1 < \sum_{i=1}^{n+1} A_i$$

$$\sum_{i=1}^{n} A_i \le 1 < \sum_{i=1}^{n+1} A_i \Leftrightarrow \sum_{i=1}^{n} -\frac{1}{\alpha}\ln R_i \le 1 < \sum_{i=1}^{n+1} -\frac{1}{\alpha}\ln R_i$$

$$\Leftrightarrow \prod_{i=1}^{n} R_i \ge e^{-\alpha} > \prod_{i=1}^{n+1} R_i$$

The procedure for generating a Poisson random variate, N, is given by the following steps:

Step 1: Set n = 0, and P = 1

Step 2: Generate a random number R_{n+1} and let P = P. R_{n+1}

Step 3: If P < $e^{-\alpha}$, then accept N = n. Otherwise, reject current n, increase n by one, and return to step 2

Nonstationary Poisson Process (NSPP):

- A Poisson arrival process whose arrival rate (l_i) changes over time.
- Think "fast food". Arrival rates at the lunch and dinner hour much greater than arrival rates during "off hours".
- Thinning Process:
- Generates Poisson arrivals at the fastest rate, but "accept" only a portion of the arrivals, in effect thinning out just enough to get the desired time-varying rate.

Nonstationary Poisson Process (NSPP) – Thinning Algorithm:

To generate successive arrival time (Ti) when rates vary:

Step 1 – Let $\lambda^* = \max_{0\le t\le T} \lambda(t)$ be the maximum arrival rate, and set t = 0and i = 1.

Step 2 – Generate E from the exponential distribution with rate λ*, and let t = t+E (the arrival time of the next arrival using max rate).

Step 3 – Generate random number R from U [0, 1). If R < λ (t) / λ* then Ti = t and i = i + 1.

Step 4 – Go to step 2.

Note: Some part of random variate generation chapters are not covered please refer to text book.