Random Number

By
Prof. S.Shakya

Random Number

- Random numbers are samples drawn from a uniformly distributed random variable between some satisfied intervals, they have equal probability of occurrence.
- Properties of random number has two important statistical properties.

1. Uniformity and
2. Independence

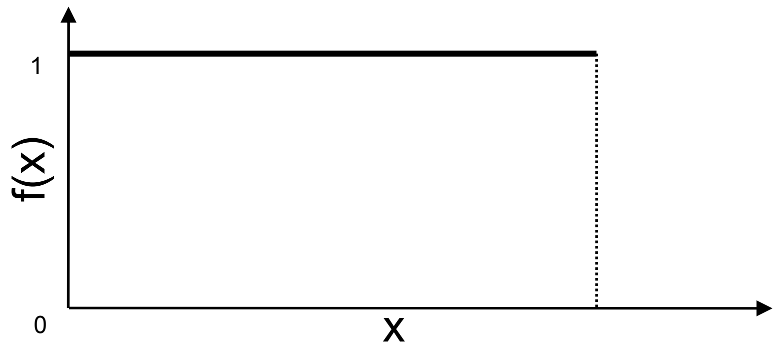
Random Number Generation (cont.)

Each random number R_t is an independent sample drawn from a continuous uniform distribution between 0 and 1

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

Random Number Generation (cont.)

PDF:



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

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

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

$$= 1/12$$

Random Number

- If the interval between 0 and 1 is divided into n equal parts or classes of equal length, then
  - The probability of observing a value in a specified interval is independent of previous value drawn
  - If a total of m observations are taken, then the expected number of observations in each interval in m/n, for uniform distribution.

Pseudo Random Numbers

- The pseudo means false.
- Pseudo implies that the random numbers are generated by using some known arithmetic operation.
- Since, the arithmetic operation is known and the sequence of random numbers can be repeated obtained, the numbers cannot be called truly random.
- However, the pseudo random numbers generated by many computer routines , very closely fulfill the requirement of desired randomness.

Pseudo Random Numbers

- If the method of random number generation that is the random number generator is defective, the generated pseudo random numbers may have following departures from ideal randomness.

- The generated numbers may not be uniformly distributed

- The generated numbers may not be continuous

- The mean of the generated numbers may be too high or too low

- The variance may be too high or too low.

Pseudo Random Numbers

- There may be cyclic patterns in the generated numbers, like;

  a) Auto correction between numbers

  b) a group of numbers continuously above the mean, followed by group continuously below of mean.
- Thus, before employing a pseudo random number generator, it should be properly validated, by testing the generated random numbers for randomness.

Generation of random number

- In computer simulation, where a very large number of random numbers is generally required, the random numbers can be obtained by the following methods.

1. Random numbers may be drawn from the random number tables stored in the memory of the computer.
2. Using electronics devices-Very expensive
3. Using arithmetic operation

Techniques for Generating Random Number (cont.)

Note: Cannot choose a seed that guarantees that the sequence will not degenerate and will have a long period. Also, zeros, once they appear, are carried in subsequent numbers.

Ex1: X_0 = 5197 (seed) X_0^2 = 27$\underline{0088}$09

==> R_1 = 0.0088 X_1^2 = 00$\underline{0077}$44

==> R_2 = 0.0077

Ex2: X_0 = 4500 (seed) X_0^2 = 20$\underline{2500}$00

==> R_1 = 0.2500 X_1^2 = 06$\underline{2500}$00

==> R_2 = 0.2500

Techniques for Generating Random Number (cont.)

- Multiplicative Congruential Method:

  Basic Relationship

  $X_{i+1} = a X_i \pmod{m}$, where $a \geq 0$ and $m \geq 0$

  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.

Techniques for Generating Random Number (cont.)

The max period(P) is:

- For m a power of 2, say $m = 2^b$, and $c \neq 0$, 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 $k = 0, 1, \ldots$

Techniques for Generating Random Number (cont.)

- 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,

Techniques for Generating Random Number (cont.)

(Example)

Using the multiplicative congruential method, find the period of the generator for a = 13, m = 2^6, and X_0 = 1, 2, 3, and 4. The solution is given in next slide. 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.

Techniques for Generating Random Number (cont.)

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	

Techniques for Generating Random Number (cont.)

```
SUBROUTINE RAN(IX, IY, RN)
   IY = IX * 1220703125
   IF (IY) 3,4,4
3: IY = IY + 214783647 + 1
4: RN = IY
   RN = RN * 0.4656613E-9
   IX = IY
   RETURN
END
```

Techniques for Generating Random Number (cont.)

- Linear Congruential Method:
  $X_{i+1} = (aX_i + c) \bmod m, i = 0, 1, 2....$
  (Example)
  let $X_0 = 27$, $a = 17$, $c = 43$, and $m = 100$, then
  $X_1 = (17*27 + 43) \bmod 100 = 2$
  $R_1 = 2 / 100 = 0.02$
  $X_2 = (17*2 + 43) \bmod 100 = 77$
  $R_2 = 77 / 100 = 0.77$
  ....

Summary of requirements for a good pseudo-random generator

1. The sequence of random numbers generated must follow the uniform (0, 1) distribution.
2. The sequence of random numbers generated must be statistically independent.
3. The sequence of random numbers generated must be reproducible. This allows replication of the simulation experiment.
4. The sequence must be non-repeating for any desired length. Although not theoretically possible, a long repeatability cycle is adequate for practical purposes.

Summary of requirements for a good pseudo-random generator

5. Generation of the random numbers must be fast because in simulation studies because a large number of random numbers are required. A slow generator will greatly increase the time and cost of the simulation studies/experiments.
6. The technique used in generating random numbers should require little computer memory.

Test 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.

Test for Random Numbers (cont.)

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.

Steps in Chi-Square

1. Determine the appropriate test
2. Establish the level of significance: α
3. Formulate the statistical hypothesis
4. Calculate the test statistic
5. Determine the degree of freedom
6. Compare computed test statistic against a tabled/critical value

Chi-Square Test (Example)

- The two Digit random numbers generated by a multiplicative congruential method are given below. Determine Chi-Square. Is it acceptable at 95% confidence level?
- 36, 91, 51, 02, 54, 06, 58, 06,58,02, 54, 01, 48, 97, 43, 22, 83, 25, 79, 95, 42, 87, 73, 17, 02, 42, 95, 38, 79, 29, 65, 09, 55, 97, 39, 83, 31, 77,17, 62, 03, 49, 90, 37, 13, 17, 58, 11, 51, 92, 33, 78, 21, 66, 09, 54, 49, 90, 35, 84, 26, 74, 22, 62, 12, 90,36, 83, 32, 75, 31, 94, 34, 87, 40, 07, 58, 05, 56,22, 58,77, 71, 10, 73,23,57,13, 36,89,22,68,02,44,99,27,81,26,85, 22

Chi-Square Test(example)

Class	Frequency	Diff(Oi-Ei)	Diff (Oi-Ei)^2
0<r≤ 10	13	3	9
10<r≤20	7	-3	9
20<r≤30	12	2	4
30<r≤40	13	3	9
40<r≤50	7	-3	9
50<r≤60	13	3	9
60<r≤70	5	-5	25
70<r≤80	10	0	0
80 <r≤90	12	2	4
90<r≤100	8	-2	4
			82

Gap Test

- The gap test is used to determine the significance of the interval between recurrence of the same digit.
- A gap of length *x* occurs between the recurrence of some digit.
- The probability of a particular gap length can be determined by a Bernoulli trail.

$$P(\text{gap of } n) = P(x \neq 9)P(x \neq 9) \ldots P(x \neq 9)P(x = 9)$$

- If we are only concerned with digits between 0 and 9, then $P(\text{gap of } n) = 0.9^n 0.1$
- The theoretical frequency distribution for randomly ordered digits is given by
- $P(\text{gap} \leq x) = F(x) = 0.1 \sum_{n=0}^{x} (0.9)^n = 1 - 0.9^{x+1}$ …….(1)

Algorithm for GAP Test

- **Step 1.** Specify the cdf for the theoretical frequency distribution given by Equation (1) 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 S_N (x)as in Equation

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

- **Step 4.** Determine the critical value, Dα, from Table(K-S critical value) 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 Dα , the null hypothesis of independence is rejected

Test for Random Numbers (cont.)

- The *Gap Test* measures the number of digits between successive occurrences of the same digit.

(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

Test for Random Numbers (cont.)

We are interested in the frequency of gaps.

$$P(\text{gap of } 10) = P(\text{not } 3) \times \times \times P(\text{not } 3) \; P(3),$$

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.

Test for Random Numbers (cont.)

(Example)

Based on the frequency with which gaps occur, analyze the 110 digits above to test whether they are independent. Use $\alpha = 0.05$. The number of gaps is given by the number of digits minus 10, or 100. The number 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

Test for Random Numbers (cont.)

Gap Test Example

Gap Length	Frequency	Relative Frequency	Cum. Relative Frequency	F(x)	$\|F(x) - S_N(x)\|$
0-3	35	0.35	0.35	0.3439	0.0061
4-7	22	0.22	0.57	0.5695	0.0005
8-11	17	0.17	0.74	0.7176	0.0224 ←
12-15	9	0.09	0.83	0.8147	0.0153
16-19	5	0.05	0.88	0.8784	0.0016
20-23	6	0.06	0.94	0.9202	0.0198
24-27	3	0.03	0.97	0.9497	0.0223
28-31	0	0.00	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

Example Explanation

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

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

Example Explanation

- Number of Gaps: Number of data values – Number of Distinct Digits = 110 – 10 = 100
- The gap length and the frequency give the total number of occurrences of gaps of the lengths in the class. For example, the gap length 0-3 has a frequency of 35 means for all the digits from 0 to 9, the total number of gaps of length 0, 1, 2 or 3 are 35. Similarly the second class 4-7 tells us that there are 22 gaps in total in the table that are of length 4, 5, 6 or 7.

Example Explanation

- The **relative frequency** is given by

  Relative frequency = Frequency / Number of gaps
- For first class, **relative frequency = 35/100 = 0.35 a**nd so on.

Example Explanation

- The value of the **theoretical frequency distribution F(x)** has been calculated using the formula:

  $1-0.9^{x+1}$ where x is the maximum length of the gap in that class.

- For example, in the table, the first gap length is 0-3. So, taking the maximum gap length of 3, we have,

  $F(3) = 1-0.9^{3+1} = 1-0.9^{4} = 0.3439$ and so on for the remaining rows.

Example Explanation

- The value of the **observed frequency distribution $S_N(x)$** has been calculated using the formula:

$$S_N(x) = \frac{\text{Number of gaps} \le x}{\text{Total no of gaps (N)}}$$

- This value is equal for the **cumulative relative frequency** for each of the gap class.

Example Explanation

- For the first gap length of 0-3, we have **cumulative relative frequency of 0.35.** So $\mathbf{S_N(x)}$ is also 0.35. Thus,

$| F(x) - S_N(x) | = | 0.3439 – 0.35 | = | -0.0061 | = 0.0061$

and so on for all other classes.

Test for Random Numbers (cont.)

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 $D_{0.05}$, do not reject the hypothesis of independence on the basis of this test.

Test for Random Numbers (cont.)

- *Poker Test* - based on the frequency with which certain digits are repeated.

  Example:

  0.255 0.577 0.331 0.414 0.828 0.909

  Note: a pair of like digits appear in each number generated.

Test for Random Numbers (cont.)

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

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

Test for Random Numbers (cont.)

(Example)

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

Let α = 0.05.

The test is summarized in next table.

Test for Random Numbers (cont.)

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 $\chi^2_{0.05, 2} = 5.99 < 47.65$, the independence of the numbers is rejected on the basis of this test.

Methods of generating non-uniform Variables: Generating discrete distributions

Generating discrete distributions

- When the discrete distribution is uniform, the requirement is to pick one of N alternatives with equal probability given to each.
- Given a random number U($0 \le U < 1$), the process of multiplying by N and taking the integral portion of the product, which is denoted mathematically by the expression [UN], gives N different outputs.
- The output are the numbers 0,1,2,….,(N-1).
- The result can be changed to the range of values C to N + C-1 by adding C.

Generating discrete distributions

- Alternately, the next highest integer of the product UN can be taken.
- In that case, the outputs are 1, 2,3 ,…..,N.
- Note that the rounded value of the product U.N is not satisfactory.
- It produces N+1 numbers as output, since it includes 0 and N and these two numbers have only half the probability of occurring as the intermediate numbers.
- Generally, the requirement is for a discrete distribution that is not uniform, so that a different probability is associated with each output.

Table 1: Number of Items Bought by customers

Number of Items Xi	Number of Customers Ni	Probability Distribution P(Xi)	Cumulative Distribution P(Xi)
1	25	0.10	0.10
2	128	0.51	0.61
3	47	0.19	0.80
4	38	0.15	0.95
5	12	0.05	1.00

Generating discrete distributions

- Suppose, for example , it is necessary to generate a random variable representing the number of items bought by a shopper at store , where the probability function is the discrete distribution, where the probability function is the discrete distribution given previous in table 1.

- A table is formed to list the number of items, x , and the cumulative probability , y, as shown in table 2.

- Taking the output of a uniform random number generator, U, the value is compared with the values of y.

- If the value falls in an interval $y_i < U \leq y_{i+1}$(i=0,1,…….,4), the corresponding value of x_{i+1} is taken as desired output.

Table 2: Generating a Non-Uniform Discrete Distribution

No of Items X	Probability P(X)	Cumulative Probability y
0	0	0
1	0.10	0.10
2	0.51	0.61
3	0.19	0.80
4	0.15	0.95
5	0.05	1.00

Generating discrete distributions

- It is not necessary that the intervals be in any particular order.
- A computer routine will usually search the table from the first entry .
- The amount of searching can be minimized by selecting the intervals in decreasing order of probability.
- Arranged as a table for a computer routine, the data of Table 2, would then appear as shown in table 3,
- With this arrangement , 51% of the searches will only need to go to the first entry , 70% to the first or second and so on.
- With the original ordering, only 10% are satisfied with the first entry and only 61% with the first two

Table 3: Non-Uniform Discrete Distribution Data Arranged for a Computer Program

Cumulative Probability U	No. of Items X
0.51	2
0.70	3
0.85	4
0.95	1
1.00	5

Generating discrete distributions

- Random numbers generators, rather than distribution functions, the input of table 3 is being denoted by U to indicate a random number between 0 and 1.
- The output continuous to be denoted by x, however, to maintain the connection to the distribution from which the numbers defining the generator have been derived.

Inversion, Rejection and Composition

Inversion

- In the simplest case of inversion, we have a continuous random variable X with a strictly increasing distribution function F. Then F has an inverse F^{-1} defined on the open interval (0,1): for $0<u<1$, $F^{-1}(u)$ is the unique real number x such that $F(x)=u$ i.e. $F(F^{-1}(u))=u$, and $F^{-1}(F(x))=x$.

$$P(F^{-1}(U)\le x)=P(U\le F(x))=F(x)$$

- Let U~unif(0,1) denote a uniform random variable on (0,1). Then
so $F^{-1}(U)$ has distribution function F.

inversion

- To extend this result to a general distribution function F the generalized inverse of F is:

$$F^{-}(u) = \inf\{ x : F(x) \geq u \} \quad 0<u<1$$

- $F^{-}(u)$ is the u-quantile so that,

$$F(F^{-}(u)) \geq u, \quad F^{-}(F(x) \leq x,$$

And hence that $F^{-}(u) \leq x \Leftrightarrow F(x) \geq u$

Therefore,

$$P(F^{-}(U) \leq x) = P(U \leq F(x)) = F(x)$$

rejection

- The rejection method is applied when the probability density function, f(x), has a lower and upper limit to its range, a and b, respectively, and an upper bound c.
- The method can be specified as follows:
  - Compute the values of two, independent uniformly distributed variates U_1 and U_2.
  - Compute X_0=a+U1(b-a).
  - Compute Y_0=cU_2.
  - If accept X_0 as the desired output; otherwise repeat the process with two new uniform variates.

rejection

- This method is closely related to the process of evaluating an integral using the Monte Carlo technique.
- The probability density function is enclosed in a rectangle with sides of lengths b-a and c.
- The first three steps of rejection method creates just a random point and the last step relates the point to the curve of the probability density function.

rejection

- In the Monte Carlo method, acceptance means that the point is added to a count, n, and, after many trials, the ratio of that count to the total number of trials, N, is taken as an estimate of the ratio of the area under the curve to the area or the rectangle.
- In the rejection method the curve is probability density function so that the area under curve must be 1 i.e. c(b-a)=1.
- The following figure shows a probability density function with limits a and b, and upper bound c.

rejection

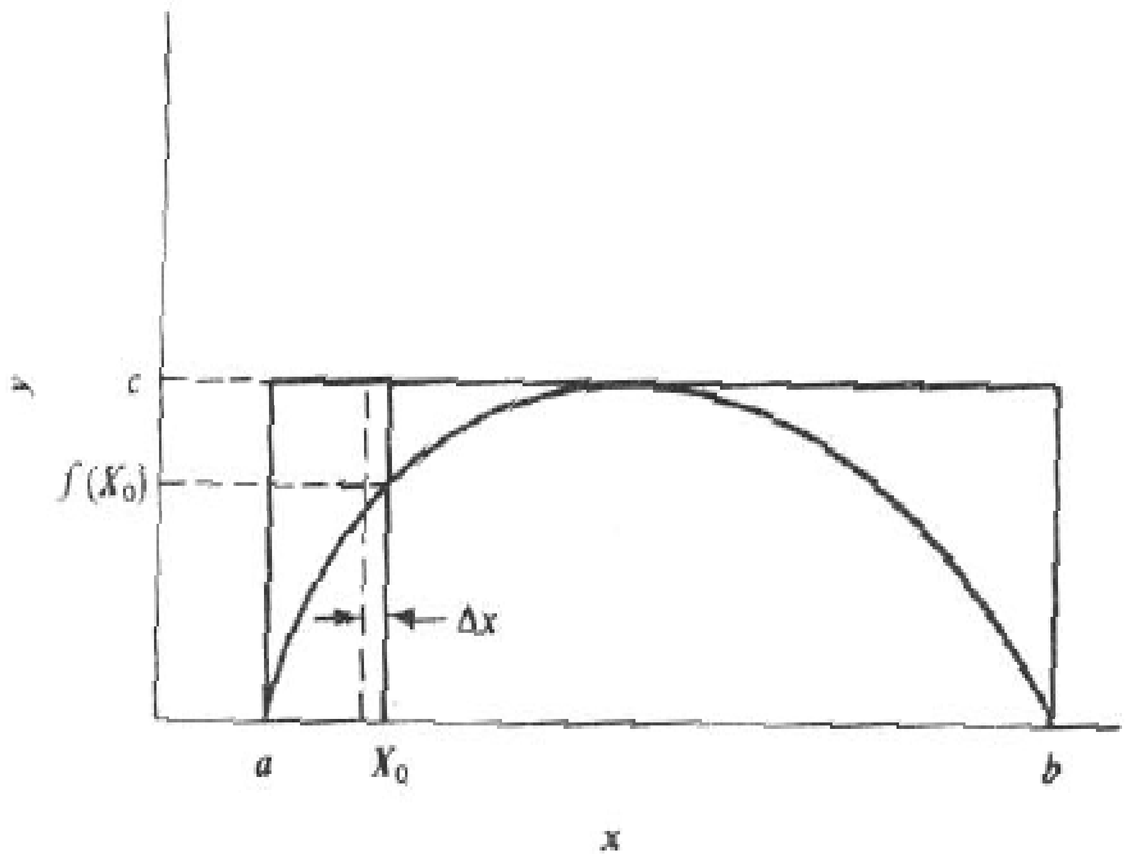


Figure 6-10. The rejection method.

rejection

- The probability that the output X is less than or equal to X_0 is equal to the probability that Y_0 falls under the curve to the left of X_0, given $X \le X_0$.
- The probability of X being less than or equal to X_0 is by definition, $F(X_0)$ so,

$$F(X_0) = \frac{\int_a^{X_0} f(x)dx}{c(X_0 - a)} \cdot \frac{(X_0 - a)}{(b - a)}$$

rejection

- Since c(b-a)=1, it follows that

$$F(X_0) = \int_a^{X_0} f(x)dx$$

which shows that X_0 has the desired distribution.

composition

- Sometimes the random variables X of interest involves the sum of n>1 independent random variables: $X=Y_1+Y_2+Y_3+\ldots+Y_n$.
- To generate a value for X, we can generate a value for each of the random variables Y_1, Y_2, Y_3, Y_n and add them together. This is called composition.
- Composition can also be used to generate random numbers that are approximately normally distributed.

COMPOSITION

- The normal distributed is one of the most important and frequently used continuous distributions.
- The notion N(μ , σ) refers to the normal distribution with mean μ and variance σ^2.
- The central limit theorem in probability says that if Y_1, Y_2, Y_3, Y_n are independent and identically distributed random variables with mean μ and positive variance σ^2, then is random variable.

$$Z = \frac{\sqrt{n}\left(\left(\frac{x}{n}\right) - \mu\right)}{\sigma}$$

Convolution Method

- The probability distribution of a sum of two or more independent random variables is called a convolution of the distributions of the original variables.
- The convolution method thus refers to adding together two or more random variables to obtain a new random variable with the desired distribution.

Convolution Method

- Technique can be used for all random variables *X that can be expressed as the* sum of *n random variables* $X = Y1 + Y2 + Y3 + . . . + Yn$
- In this case, one can generate a random variate *X by generating n random* variates, one from each of the *Yi, and summing them.*
- Examples of random variables:

– Sum of *n Bernoulli random variables is a binomial random variable.*

– Sum of *n exponential random variables is an n-Erlang random variable.*

References

- Jerry Banks, John S. Carson, II Barry L. Nelson , David M. Nocol, “Discrete Event system simulation”
- Geoffrey Gordon, System Simulation