UNIT 14 RANDOM NUMBER GENERATION FOR CONTINUOUS VARIABLES

Structure

14.1 INTRODUCTION

In Unit 13, we have given some methods of generating variates from some important discrete distributions. In this unit we shall describe generation of variates from some important continuous distributions. Sometimes we shall use the IPT method for generation of variates from a particular distribution but in some cases we shall use some other methods which are more convenient and efficient.

Method of generation of random variates from continuous Uniform and Exponential probability distribution is explained in Sections 14.2 and 14.3. In Section 14.4, the some useful methods of generation of random variates from Normal distribution i.e. Inverse Probability Transformation method, Central Limit Theorem approximation method, Box-Muller method and Marsaglia's Polar method are explored with examples. In Section 14.5, the methods of generating the variates from the Gamma, Chi-square and Beta distributions are explained whereas a brief discussion has been done about the Poisson process and Waiting Time distributions in Section 14.6.

Objectives

After studying this unit, you would be able to

- describe the generation of variables from continuous distributions;
- explain the method of generation of variables from Uniform distribution;
- explain the method of generation of variables from Exponential distribution;
- explain the method of generation of variables from Normal distribution;

- describe the method of generation of variables from Gamma distributions; and
- describe the method of generation of random variates from Poisson Process.

14.2 UNIFORM DISTRIBUTION

We have already described generation of Uniform random variable U(0,1) in Sub-section 13.3.5. Now if one wishes to generate variates from U(a,b) whose probability density function (p.d.f.), f(x), is given by

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

A simple transformation of u~ U (0, 1) will convert it to x ~ U (a, b)

$$u = (x - a)/ (b - a)$$

Hence $\quad x = a + (b - a)\, u$

Following is the algorithm:

1. Generate a uniform variate u ~ U(0,1)
2. Take x = a + (b - a) u

Repeat the steps (1) - (2) with a new u to generate more variates.

14.3 EXPONENTIAL DISTRIBUTION

The probability density function of Exponential distribution with parameter α is given by

$$f(x) = \alpha e^{-\alpha x}, \qquad \alpha > 0, x \geq 0$$

where, 1/ α is the mean of the distribution.

Distribution function F(x) of Exponential distribution is given by

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

Using IPT described in Sub-section 13.4.2 of Unit 13 and equating u~ U (0, 1) to F(x)

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

giving

$$x = -[\log (1 - u)]/ \alpha$$

or equivalently

$$x = -[\log (u)]/ \alpha$$

Algorithm is

1. Generate a uniform variate u ~ U (0, 1)
2. Take x = − [log (u)]/α, which has Exponential distribution with parameter α.

For example, if u = 0.5921. To generate an exponential variate with α = 2, take

$$x = -\log(u)/\alpha = 0.262$$

x is the desired exponential variate with α = 2.

14.4 NORMAL DISTRIBUTION

The probability density function of Normal distribution N (μ, σ^2) is given by

$$f(x) = \frac{1}{\sqrt{2\pi}\ \sigma^2} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \quad , -\infty < x < \infty$$

Transformed variable

$$z = \frac{(x-\mu)}{\sigma} \text{ is following N (0, 1)}$$

The cumulative distribution function (c.d.f.), of z denoted by Φ (z), is given by

$$\Phi(z) = P(Z \le z) = \int_{-\infty}^{Z} \frac{1}{\sqrt{2\pi}} e^{-\frac{u^2}{2}} du$$

14.4.1 Inverse Probability Transformation Method

We generate a u ~ U (0,1) and using inverse probability transformation, we equate

$$u = \Phi(z)$$

$$z = \Phi^{-1}(u)$$

Then we obtain z from standard normal cumulative probability table. Then X~ N (μ, σ^2) is given by equating

$$z = \frac{(x-\mu)}{\sigma}$$

giving, $x = \mu + \sigma z$

For example, if we wish to generate a X~ N (μ, σ^2), with μ = 5.0 and

$\sigma^2 = 2.0$.

Suppose, the generated u~U (0, 1) is 0.69.

Equating u to Φ (z), we obtain

$$\Phi(z) = 0.69$$

Using the normal c.d.f. table, we have

$$z = \Phi^{-1}(0.69) = 0.50$$

$$z = \frac{(x - 5.0)}{\sqrt{2.0}}$$

giving $$x = 5.0 + \sqrt{2.0}\ (0.50)$$

$$= 5.707$$

Hence, x is the desired normal variate.

As the integral Φ (z) is not in a closed form, one has to use normal probability table which is not very convenient when one has to generate a very large number of variates. We shall use an alternative method, which is more popular and efficient for generation of normal variates.

14.4.2 Central Limit Theorem Approximation

The slow but simple approach makes use of the Central limit theorem (CLT) from mathematical statistics. An often overlooked consequence of the CLT is its assertion that the sum of n independent samples from identical distributions, such as $x_1 + x_2 + \ldots + x_n$, is normally distributed with the mean equal to $n\mu_x$ and a variance of $n\sigma_x^2$ where μ_x and σ_x^2 are mean and variance of the variable X respectively. If the distribution of X is uniform on the (0, 1) interval, then a sum $y = x_1 + x_2 + \ldots + x_n$ of n such numbers is normally distributed with

$$\mu_y = n/2 \text{ and } \sigma_y^2 = n/12$$

so that

$$z = \frac{y - n/2}{\sqrt{n/12}}$$

is standard normal with $\mu_z = 0$ and $\sigma_z^2 = 1$.

The CLT informs us that the larger the value of n, the better the approximation to the Normal distribution. For most purposes, a value of 12 is convenient and leads to a reasonable approximation. In this case, z equation will be

$$z = \frac{y - 12/2}{\sqrt{12/12}} = y - 6$$

The CLT generator is not the best one if more than a few numbers are to be generated. It is very slow and it does not adequately sample the extreme tails of the Normal distribution.

14.4.3 Box-Muller Method

This method is very popular and used often to generate normal random variable. We shall give only algorithm here but for details see Ross (2002).

Algorithm

1. Generate two independent uniform U(0,1) variates u_1 and u_2.

2. $x = \sqrt{(-2\log u_1)}.\cos(2\pi u_2)$

   $y = \sqrt{(-2\log u_1)}.\sin(2\pi u_2)$

   where, $2\pi = 360^o$.

This method gives two normal independent N (0,1) variables x and y corresponding to two independent u_1 and u_2. x and y thus generated are desired N (0, 1) variables. They have to be transformed to obtain N (μ, σ^2). More variables can be generated by repeating the steps (1)-(2) with new independent u_i's.

14.4.4 Marsaglia's Polar Method

Complexity arises in generating Sines and Cosines of random angles in using above method. Marsaglia (1962), gave the following modification in the Box-Muller transformation that avoids evaluation of the trigonometric function:

1. Generate u_1 and u_2 from U (0, 1) and set $v_1 = 2\,u_1 - 1$ and $v_2 = 2\,Uu_2 - 1$

2. If $v_1^2 + v_2^2 > 1$ go to step 1

3. Deliver

   $$C = \left[\frac{-2\,\text{Ln}\left(v_1^2 + v_2^2\right)}{\left(v_1^2 + v_2^2\right)}\right]^{\frac{1}{2}}$$

   where, Ln is the natural logarithm at the base e.

4. $x = Cv_1$, $y = Cv_2$

E1) Six independent u_i~U (0,1) are given below:

$u_1 = 0.59$, $u_2 = 0.32$, $u_3 = 0.89$, $u_4 = 0.17$, $u_5 = 0.92$, $u_6 = 0.66$

Using Box-Muller method, generate 6 independent N (3, 1.5) variates.

14.5 GAMMA, CHI-SQUARE AND BETA DISTRIBUTIONS

14.5.1 Gamma Distribution

The probability distribution function of Gamma distribution G(n, α) is given by

$$f(x) = \frac{\alpha^n x^{n-1} e^{-\alpha x}}{\Gamma(n)}, \qquad 0 \le x < \infty, \alpha > 0, n > 0$$

The cumulative distribution function is not in closed form and thus IPT method cannot be applied. Therefore, some alternative methods have been used.

In case where n is an integer then G (n, α) can be obtained by summing n exponential variables with parameter α.

Thus, the algorithm is:

1. Generate n independent uniform U (0,1) variables $u_1, u_2, \ldots, u_n$.
2. Obtain n independent exponential variates y_i's with parameter α.

$$y_i = -\log u_i / \alpha, \qquad i = 1, 2, \ldots, n$$

3. $x = \sum_{i=1}^{n} y_i = \sum_{i=1}^{n} \frac{-\log u_i}{\alpha}$

This is also known as **Erlang** distribution.

When n is not an integer then the generation of Gamma variable is rather complicated. For generation of variables for non-integer values of n see Law and Kelton (1982).

For example, if we wish to generate two random variates from G(2, 0.5) and suppose we have generated two independent u_i's as

$$u_1 = 0.49, \qquad u_2 = 0.83,$$

Therefore, compute

$$x = \sum_{i=1}^{2} \frac{-\log u_i}{0.5} = 1.799$$

14.5.2 Chi-square Distribution

Chi-square distribution, denoted by χ^2_m, is a particular case of Gamma distribution with n = m/2, where m is even and α = 1/2. The probability distribution function is

$$f(x) = \frac{x^{(m/2)-1} e^{-x/2}}{2^{m/2}\Gamma(m/2)}, \qquad x > 0$$

Chi-square variate χ^2_m may also be defined as sum of square of m independent standard normal variates, so it can be generated by the sum of squares of m independent N (0, 1) variables. Algorithm is:

1. Generate m independent N (0,1) variates z_i's, (i = 1, 2,…, m)
2. Take $x = \sum_{i=1}^{m} z_i^2$
3. x is the desired χ^2_m variate on m degrees of freedom.

14.5.3 Beta Distribution

A random variable X has a Beta distribution if the probability distribution function is given by:

$$f(x) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1}(1-x)^{\beta-1} \; ; \alpha > 0, \beta > 0, 0 \le x \le 1$$

and is denoted by Be (α, β).

An easy method of generation of Beta random variate is the following:

If two variates y_1 and y_2 are independently distributed as $G(\alpha,1)$, $G(\beta,1)$, respectively, then

$$x = y_1/ (y_1+y_2)$$

has a $Be(\alpha, \beta)$ distribution.

E2) Using first four N (0, 1) variates generated in E1) obtain a χ^2 variate with four degrees of freedom.

14.6 POISSON PROCESS

A random variable which is often used for simulation in queuing theory is known to be distributed as Poisson Process.

If the inter-arrival times $x_1, x_2, \ldots$ are independent and identically distributed as exponential random variables with parameter α, then the times at which events occur is known to have a poisson process. In this case, number of events occurring in an interval (0, t) has a Poisson distribution with mean αt. Seeing the above relationship between Exponential distribution and poisson process one can easily generate poisson process variate.

Algorithm is given as follows:

1. Generate $u_1, u_2, \ldots$ from U (0, 1).
2. Find exponential variates y's by taking

   $y_i = -(\log u_i)/ \alpha, i = 1, 2, \ldots$
3. Take $t_1 = -(\log u_1)/\alpha = y_1$

$$t_2 = t_1 - (\log u_2) / \alpha = y_1+y_2$$

$$t_3 = t_2 - (\log u_3) / \alpha = y_1+y_2+y_3$$

$$\vdots$$

$$t_i = t_{i-1} - (\log u_i) / \alpha = y_1+y_2+\ldots+y_i$$

$t_1, t_2, t_3, \ldots$ are the times at which event occur and follow poisson process with rate α.

For example, if six U (0, 1) variables u_i's are given as:

$u_1 = 0.78, u_2 = 0.31, u_3 = 0.76, u_4 = 0.23, u_5 = 0.79, u_6 = 0.96$

so we generate timing (t_i) of event occurring from a poisson process when rate of events occurring per unit time is 0.3 i.e. $\alpha = 0.3$.

Then, the calculated $y_i = -(\log u_i)/ \alpha$ are:

$$y_1= 0.828,\ y_2 = 3.904,\ y_3 = 0.915,\ y_4 = 4.899,\ y_5= 0.786,\ y_6 = 0.136$$

Therefore, $\quad t_i = \sum_{j=1}^{i} y_j$

$t_1 = 0.828$, $t_2 = 4.732$, $t_3 = 5.647$, $t_4 = 10.565$, $t_5 = 11.332$, $t_6 = 11.468$

t_i's are the required timings at which the events occur.

E3) The inter-arrival times of patients arriving in a clinic has an Exponential distribution with rate $\alpha = 0.2$ per minute. Simulate the times of six patients arriving in the clinic. Also give the number of patients arriving in first 20 minutes.

(Hint: Use LCG $x_i = (1573\, x_{i-1} + 19) \bmod 10^3$, starting with $x_0 = 159$.)

E4) A train is expected to arrive in a station at 8:00 AM. However, it has been observed that it reaches station between 7:55 A.M. to 8:05 AM and the times are uniformly distributed between the above intervals. Using the following U (0, 1) random numbers simulate time for arrival on ten days:

U (0, 1): 0.579 0.052 0.312 0.307 0.645 0.945 0.645 0.956
0.394 0.110 0.854 0.476

E5) In a warehouse three trucks arrive per hour to be unloaded. Generate inter-arrival times of ten trucks using random numbers of E4), assuming that distribution of inter-arrival times is exponential. In how many cases it exceeds half an hour?

E6) The mileage (in thousand of miles) which car owners get with a certain kind of radial tyre is a random variable having an Exponential distribution with $\alpha = 0.025$. Generate the mileage of five such tyres (use random numbers of E4). Obtain the average mileage of these five tyres.

E7) The probability density of X is given by

$$f(x) = \begin{cases} \dfrac{x}{2} & ; \quad \text{for } 0 < x < 2 \\ 0 & , \quad \text{elsewhere} \end{cases}$$

Generate five random variables from f(x) using U (0, 1) from E4).

E8) If u_i's are independent U (0, 1) random variables then define y as sum of n "u_i's" as

$$y = \sum_{i=1}^{n} u_i$$

$$E(y) = n/2, \quad V(y) = n/12$$

If n is large then $z = [y - E(y)]/\sqrt{V(y)}$

is approximately distributed as N(0, 1). Using ten u_i's given in E4) generate a normal random variate N (0, 1) in this way.

E9) Give an algorithm using IPT method to generate variate from the following Beta probability density function:

$$f(x) = 6x\,(1 - x), \qquad 0 \le x \le 1$$

Find x when u = 0.5.

E10) Distribution function of Pareto random variable is given by

$$F(x) = 1 - \left(\frac{k}{x}\right)^a, \qquad a > 0,\ 0 < k \le x$$

Given a u~U (0, 1) generate x, when a = 2 and k = 1. Suppose u = 0.5 then find x.

E11) If x and y are distributed normally and independently as N (2, 2) and N (3, 5), respectively. Define a new variate z = x + y. Generate a new variable z using uniform variates given in E4).

E12) Describe an algorithm using IPT for generating a variate from Logistic distribution whose p.d.f. is given by:

$$f(x) = \frac{e^{-x}}{\left(1+e^{-x}\right)^2}, \qquad -\infty < x < \infty$$

Take u = 0.3 to generate a variate x from the above distribution.

E13) In a certain city, the daily consumption of water (in millions of litres), follows approximately a Gamma distribution, G (n, α) with n = 3 and α = 1/3. Using random numbers given in E4) generate daily consumption of water for three days from this distribution.

E14) If the annual proportion of new restaurants that fail in a given city may be looked upon as a random variable having a Beta distribution with α = 1 and β = 4, then generate proportion of restaurants that are expected to fail in one year, using random numbers given in E4).

E15) The life in years of a certain type of electrical switch has an Exponential distribution with α = 0.5. Generate life time (in years) for ten switches using uniform variates in E4). What is the proportion of bulbs with life more than one year in the sample obtained?

E16) Buses arrive at a sporting event according to Poisson process with rate of five per hour. Write an algorithm to simulate the arrival of buses by time t =1, using uniform variable in E4)

14.7 SUMMARY

In this unit, we have discussed:

1 Method for generation of random variables from Uniform distribution;

2 Method for generation of random variables from Exponential distribution;

3 Various methods for generation of random variables from Normal distribution;

4 Methods for generation of random variables from Gamma distribution;

5 Methods for generation of random variables from Chi-square distribution;

6 Methods for generation of random variables from Beta distribution; and

7 Poisson process which is useful in simulation of queuing problems.

14.8 SOLUTIONS / ANSWERS

E1) Taking pairs (u_1, u_2), (u_3, u_4), (u_5, u_6) and using Box-Muller transformation we obtain the following N (0, 1) variates:

$y_1 = -0.249$, $y_2 = 0.999$, $y_3 = 0.276$, $y_4 = 0.396$, $y_5 = -0.339$,

$y_6 = -0.227$

Corresponding N (3,1.5) variates are obtained by

$x_i = (\sqrt{1.5})\, y_i + 3.0$ giving

$x_1 = 2.695$, $x_2 = 4.223$, $x_3 = 3.338$, $x_4 = 3.485$, $x_5 = 2.585$, $x_6 = 2.722$

E2) N (0,1) variables y_i's generated in E1) are given as

$y_1 = -0.249$, $y_2 = 0.999$, $y_3 = 0.276$, $y_4 = 0.396$

χ^2_4 variable x = $\sum_{i=1}^{4} y_i^2 = 1.293$

E3) We have $u_1 = 0.126$, $u_2 = 0.217$, $u_3 = 0.577$, $u_4 = 0.640$, $u_5 = 0.739$,

$u_6 = 0.466$ then

$y_i = -(\log u_i)/\alpha, \qquad i = 1, 2, \ldots, 6$

$y_1 = 10.357$, $y_2 = 7.639$, $y_3 = 2.749$, $y_4 = 2.231$, $y_5 = 1.512$,

$y_6 = 3.818$

Therefore, $t_i = \sum_{j=1}^{i} y_j$

$t_1 = 10.357$, $t_2 = 17.996$, $t_3 = 20.745$, $t_4 = 22.976$, $t_5 = 24.488$,

$t_6 = 28.306$.

Number of patients in first twenty minutes = 2

E4) This is the case of uniform distribution U (a, b) with a = 7:55,

b = 8:05, therefore

x = 7:55 + 10 u

Using u's, we get x's as:

8:01, 7:55, 7:58, 7:58, 8:01, 8:04, 7:59, 7:56, 8:04, 8:00.

E5) Exponential variate is given by

$x_i = -(\log u_i)/\alpha = -(\log u_i)/3$

This gives x_i as

x: 0.182, 0.985, 0.388, 0.394, 0.146, 0.015, 0.310, 0.736, 0.052, 0.247

(where x is in hours)

In two cases it exceeds half an hour.

E6) We have $x_i = -(\log u_i)/\alpha = -(\log u_i)/0.025$

x: 21.86, 118.26, 46.59, 47.24, 17.54

Average = 50.30 (in thousand miles).

E7) Cumulative distribution function F(x) is given by

$$F(x) = \int_0^x \frac{u}{2}\, du$$

$$F(x) = \frac{x^2}{4}, \qquad 0 < x < 2$$

Equating uniform U (0, 1) random variable u, we have

$$u = \frac{x^2}{4} \Rightarrow x = 2\sqrt{u}$$

This gives 1.522, 0.456, 1.117, 1.108, 1.606.

E8) We have $y = \sum_{i=1}^{10} u_i = 4.685$

Then E (y) = n/2 = 10 / 2 = 5.0;

V(y) = n / 12 = 10/12 = 0.833

Therefore,

$$z = (4.685 - 5.0) / \sqrt{0.833}$$

$$= -0.315 / 0.913$$

$$= -0.345$$

z is the required N (0,1) variate.

E9) We have $f(x) = 6x(1-x) \qquad 0 \le x \le 1$

$$F(x) = \int_0^x 6u(1-u)\, du$$

$$= 3x^2 - 2x^3 \qquad 0 \le x \le 1$$

Generate a u ~ U (0, 1), Put

$$u = 3x^2 - 2x^3$$

and obtain $0 \le x \le 1$ which satisfies the above equation. This gives a random variate from f(x).

For u = 0.5 the value of x = 0.50.

E10) Equate

$$u = 1 - \left(\frac{k}{x}\right)^a$$

$$x = \frac{k}{(1-u)^{1/a}}$$

when u = 0.5 then

$$x = \frac{1}{(1-0.5)^{1/2}} = \frac{1}{\sqrt{0.5}}$$

$$= 1/0.707 = 1.414$$

Hence, x = 1.414 is the generated variate.

E11) Form E4), we have, $u_1 = 0.579$, $u_2 = 0.052$

Using Box-Muller transformation

$$x_1 = \sqrt{(-2\log u_1)}.\cos(2\pi u_2) = 1.045 \times 0.957 = 1.000$$

$$x_2 = \sqrt{(-2\log u_1)}.\sin(2\pi u_2) = 1.045 \times 0.290 = 0.303$$

If x, y are two independent standard normal variates then we know that

$$z = x + y$$

is also normally distributed with mean as

$$E(z) = E(x) + E(y)$$

$$= 2 + 3 = 5,$$

and variance V(z) as

$$V(z) = V(x) + V(y)$$

$$= 2 + 5 = 7$$

Now, z can be generated from x_1 by taking

$$x_1 = \frac{z-5}{\sqrt{7}}$$

giving $z = 5 + \sqrt{7}\, x_1 = 7.645$, which is the desired variate.

E12) Cumulative density function for a Logistic distribution is given by

$$F(x) = \frac{1}{\left(1 + e^{-x}\right)}$$

Suppose u is U (0, 1) random variate then taking

$$u = \frac{1}{\left(1 + e^{-x}\right)}$$

giving

$$x = -\log\left[(1-u)/u\right]$$

Put u = 0.3

$$x = -\log\left[0.7/0.3\right]$$

$$= -\log(2.333) = -0.847$$

Hence, the logistic variate generated is -0.847.

E13) If n is an integer then we can generate a Gamma (n, α) variate by summing n exponential variate with parameter α. Hence the algorithm is

$$x = -\left(\sum_{i=1}^{n} \log u_i\right) / \alpha$$

Using u_i's in E4) we obtain Gamma variate G (3, 1/3)

$$n = 3, \alpha = 1/3$$

$$x_1 = 3\sum_{i=1}^{3} \log u_i \,,\ x_2 = 3\sum_{i=4}^{6} \log u_i \,,\ x_3 = 3\sum_{i=7}^{9} \log u_i$$

$x_1 = 13.971$, $x_2 = 8.445$ and $x_3 = 4.242$

E14) Using uniform random numbers given in E4) we generate

y_1~G (1, 1) and y_2~ G (4, 1) and then take $x = y_1 / (y_1 + y_2)$.

Take

$$y_1 = -\log u_1/1, \quad \text{and} \quad y_2 = -\sum_{i=2}^{5} \log u_i /1$$

$$y_1 = 0.546, \quad \text{and } y_2 = 5.302$$

$$x = 0.546/5.848 = 0.093$$

Hence one Beta (1, 4) variate generated is 0.093.

E15) Exponential variate with parameters α can be generated from uniform random variables U as

$$x = -[\log u] / \alpha$$

using u's given in E4) we have the following:

1.09, 5.91, 2.33, 2.36, 0.88, 0.09, 1.86, 4.41, 0.32, 1.48

Proportion of bulbs with life more than one year is $\frac{7}{10} = 0.7$.

E16) Generate times of arrival of buses by Poisson process with $\alpha = 5$. Using u_i's from E4) we have times of arrival of buses t_i as

$$t_1 = -[\log u_1] / \alpha = 0.109$$

$$t_2 = t_1 - [\log u_2] / \alpha = 0.700$$

$$t_3 = t_2 - [\log u_3] / \alpha = 0.933$$

$$t_4 = t_3 - [\log u_4] / \alpha = 1.169$$

Hence from this simulation three buses arrive in the first one hour.