UNIT 11 CONTINUOUS PROBABILITY DISTRIBUTIONS

Objectives

After reading this unit, you should be able to:

- identify situations where continuous probability distributions can be applied
- appreciate the usefulness of continuous probability distributions in decision-making.
- analyse situations involving the Exponential and the Normal distributions.

Structure

11.1 INTRODUCTION

In the last unit, we have examined situations involving discrete random variables and the resulting probability distributions. Let us now consider a situation, where the variable of interest may take any value within a given range. Suppose that we are planning for release of water for hydropower generation and irrigation. Depending on how much water we have in the reservoir viz. whether it is above or below the "normal" level, we decide on the amount and time of release. The variable indicating the difference between the actual reservoir level and the normal level, can take positive or negative values, integer or otherwise. Moreover, this value is contingent upon the inflow to the reservoir, which in turn is uncertain. This type of random variable which can take an infinite number of values is called a continuous random variable, and the probability distribution of such a variable is called a continuous probability distribution. The concepts and assumptions inherent in the treatment of such distributions are quite different from those used in the context of a discrete distribution. The objective of this unit is to study the properties and usefulness of continuous probability distributions. Accordingly, after a presentation of the basic concepts, we discuss some important continuous probability distributions, which are applicable to many real-life processes. In the final section, we discuss some possible applications of these distributions in decision-making.

Activity A

Give two examples of a continuous random variables. Note down the difficulties you face in writing down the probability distributions of these variables by proceeding in the manner explained in the last unit.

…………………………………………………………………………………………
…………………………………………………………………………………………
…………………………………………………………………………………………
…………………………………………………………………………………………
…………………………………………………………………………………………
…………………………………………………………………………………………
…………………………………………………………………………………………
…………………………………………………………………………………………...

11.2 BASIC CONCEPTS

We have seen that a probability distribution is basically a convenient representation of the different values a random variable may take, together with their respective probabilities of occurrence. The random variables considered in the last unit were discrete, in the sense that they could be listed in a sequence, finite or infinite. Consider the following random variables that we have taken up in Unit 10 :

1 Demand for Newspaper (D)

2 Number of Trials (N) required to get r successes, given that the probability of a success in any trial is P.

In the first case, D could take only finite number of integer values, 30, 31,…..35; whereas in the second case, N could take an infinite number of integer values r, r + 1, r + 2 …………… ∞. In contrast to these situations, let us now examine the example cited in the introduction of this unit. Let us denote the variable, "Difference between normal and actual water level", by X. We find that X can take any one of innumerable decimal values within a given range, with each of these values having a very small chance of occurrence. This marks the difference between the continuous variable X and the discrete variables D and N. Thus, in case of a continuous variable, the chance of occurrence of the variable taking a particular value is so small that a totally different representation of the probability function is called for. This representation is achieved through a function known as "probability density function" (p.d.f.). Just as a p.m.f. represents the probability distribution of a discrete random variable, a p.d.f. represents the distribution of a continuous random variable. Instead of specifying the probability that the variable X will take a particular value, we now specify the probability that the variable X will lie within an interval. Before discussing the properties of a p.d.f., let us study the following example.

Example 1

Consider the experiment of picking a value at random from all available values between the integers 0 and 1. We are interested in finding out the p.d.f. of this value X. (Alternatively, you may consider the line segment 0-1, with the origin at 0. Then, a point picked up at random will have a distance X from the origin. X is continuous random variable, and we are interested in the distribution of X.)

Solution

Let us first try to find the probability that X takes any particular value, say, .32.

The Probability (X = .32), written as P(X = .32) can he found by noting that the 1st digit of X has to be 3, the 2nd digit of X has to be 2 and the rest of the digits have to be zero. The event of the 1st digit having a particular value is independent of the 2nd digit having a particular value, or any other digit having a particular value.

Now, the probability that first digit of X is 3 = $\frac{1}{10}$ (As there are 10 possible numbers 0 to 9).

Similarly the probabilities of the other digits taking values of 2, 0, 0 ...etc. are $\frac{1}{10}$ each.

$$P(X = .32) = \frac{1}{10} \times \frac{1}{10} \times \frac{1}{10} \approx 0 \ldots\ldots\ldots\ldots\ldots\ldots(1)$$

Thus, we find that for a continuous random variable the probability of occurrence of any particular value is very small. Therefore we have to look for some other meaningful representation.

We now try to find the probability of X taking less than a particular value, say .32. Then P(X < .32) is found by noting the following events :

A) The first digit has to be less than 3, or

B) The first digit is 3 but the second digit is less than 2.

$$P(X < .32) = \frac{3}{10} + \frac{1}{10} \times \frac{2}{10} = .32 \ldots\ldots\ldots(2)$$

Combining (1) & (2) we have :

$$P(X \leq .32) = .32$$

Similarly, we can find the probability that X will lie between any two values a and b, i.e., $P(a \leq x \leq b)$; this is the type of representation that is meaningful in the context of a continuous random variable.

Properties of a p.d.f.

The properties of p.d.f. follow directly from the axioms of probability discussed in Unit 9. By definition, any probability function has to be non-negative and the sum of the probabilities of all possible values that the random variable can take, has to be 1. The summation for continuous variables is made possible through `integration'.

If f(X) denotes the pdf of a continuous random variable X, then

1) f(X) 0, and
2) $\int_R f(X)\, dX = 1$, where " $\int_R$ " denotes the integration over the entire range {R) of values of X.

The probability that X will lie between two values a and b, will be given by : $\int_a^b f(X)\, dx$.

The cumulative density function (c.d.f.) is found by integrating the p.d.f. from the lowest value in the range upto an arbitrary level X. Denoting the c.d.f. by F(X), and the lowest value the variable can take by a, we have :

$$F(X) = \int_a^x f(X)\, dx$$

Once the p.d.f. of a continuous random variable is known, the corresponding c.d.f. can be found. You may once again note, that as the variable may take any value in a specified interval on a real line, the probabilities are expressed for intervals rather than for individual values, and are obtained by integrating the p.d.f. over the relevant interval.

Example 2

Suppose that you have been told that the following p.d.f. describes the probability of different weights of a "1kg tea pack" of your company :

$$f(x) = \begin{cases} 100(x-1), & 1 \leq x \leq 1.1 \\ 0 \text{ otherwise} & \text{otherwise} \end{cases}$$

Verify whether the above is a valid p.d.f.

Solution

As, $f(x) = 100\,(x - 1)$ for $1 \leq x \leq 1.1$

$= 0$ otherwise.

The relevant limits for integration are 1 and 1.1; for all other values below 1 and above 1.1, the probability being zero.

In order that f(x) is a valid p.d.f., two conditions need to be satisfied. We test them one by one.

1 Check $f(x) \geq 0$

i.e. to show that $100\,(x - 1) \geq 0$ for $1 \leq x \leq 1.1$

It is easy to see that this is true, for all other values of x, f(x) is given to be 0. So this condition is satisfied.

2 Check f(x) dx = 1

i.e. To show that $\int_1^{1.1} 100\,(x-1)\,dx = 1$

Left Hand side $= 100\,\frac{x^2}{2}\Big|_1^{1.1} - 100\,x\,\Big|_1^{1.1}$ **(By integration)**

$$= \frac{100}{2}[1.1^2 - 1^2] - 100\,[1.1 - 1]$$

$$= 50 \times 2.1 \times .1 - 100 \times .1 = 10.5 - 10 = .5$$

As this is not equal to 1, this is not a valid p.d.f.

Example 3

The p.d.f. of the different weights of a "1kg tea pack" of your company is given by : $f(x) = 200 (x-1)$ for $1 \leq x \leq 1.1$

$= 0$, otherwise.

You may note that the packing process is such that even if you set the machine to a value, you will only get packs around that value. The p.d.f. shows that there are chances of only exceeding the 1 kg value and there is no chance of packing less than 1kg. This is normally achieved by setting the machine to a relatively high value to overcome the government regulation on packing standard weights.)

Verify that the given p.d.f. is a valid one. Find the probability that the weight of any pack will lie between 1.05 and 1.10.

Solution

Proceeding in the same way as in the earlier example, we can show that

$$\int_{1}^{1.1} 200(x-1)dx = 1$$

Now, we find the probability that x will lie between 1.05 and 1.10 :

$$\begin{aligned} P(1.05 \leq x \leq 1.10) &= \int_{1.05}^{1.10} 200\,(x-1)\,dx \\ &= 100\,x^2 \Big|_{1.05}^{1.10} - 200\,x \Big|_{1.05}^{1.10} \\ &= 100\,(1.1^2 - 1.05^2) - 200\,(1.1 - 1.05) \\ &= 100 \times 2.15 \times .05 - 200 \times .05 = 15 \times .05 = .75 \end{aligned}$$

Alternatively, we could have found the above as follows :

$$\begin{aligned} P(1.05 \leq x \leq 1.10) &= P(1 \leq x \leq 1.1) - P(1 \leq x \leq 1.05) \\ &= 1 - [100 \times 2.05 \times .05 - 208 - .05] \\ &= 1 - .25 = .75 \end{aligned}$$

Example 4

find the cdf for the pdf given in Example 3.

Solution

$$\begin{aligned} F(X) &= \int_{1}^{X} 200(x-1)dx \\ &= (100\,X^2 - 100) - 200(X-1) \\ &= 100\,(X^2 - 2X + 1) \end{aligned}$$

(Here , 1 is the lowest possible value that x can take).

In this section we have elaborated on the concept of a continuous random variable and have finally shown how to arrive at a representation of the probability function of such a variable. We have used "integration" for our purpose. Those of you who are not familiar with the concept of integration, may note that this is similar to the summation sign (I) used in the context of a discrete variable. Also, if f(x) vs x is plotted on a graph, we will have a curve. The integration between two values a and b of x then signifies the area under the curve, and as we have already seen, this is nothing but the probability that x will lie between a and b. This idea will be useful again when we discuss some important theoretical probability distributions for continuous variables in the next section.

Activity B

Suppose that you are told that the time to service a car at your friend's petrol station is uncertain with the p.d.f. given as :

$$\begin{aligned} f(x) &= 3x - 2\,x^2 + 1 \text{ for } 0 \leq x \leq 2 \\ &= 0 \qquad \text{otherwise.} \end{aligned}$$

Examine whether this is a valid p.d.f.

(You may need to brush up Integration from any elementary Calculus book.)

……………………………………………………………………………………………
……………………………………………………………………………………………
……………………………………………………………………………………………

Activity C

The life in Hours of an electric bulb is known to be uncertain. A particular manufacturer of bulbs has estimated the p.d.f. of "life" (the total time for which the bulb will burn before getting fused) as :

$f(x) = 0$, for $x < 0$

$$= \frac{1}{100} e^{-(x/100),} \quad \text{for } x \geq 0$$

Check whether the above is a valid p.d.f.

If it is a valid p.d.f., find the probability that a bulb will have a life of more than 100 hours.

…………………………………………………………………………………………
…………………………………………………………………………………………
…………………………………………………………………………………………
…………………………………………………………………………………………

11.3 SOME IMPORTANT CONTINUOUS PROBABILITY DISTRIBUTIONS

The knowledge of the probability density function (p.d.f.) of a continuous random variable is helpful in many ways. The p.d.f. allows us to calculate the probability that a variable will lie within a certain range. The usefulness of such calculations are illustrated with the help of the following two situations.

Situation 1

Mr. X manufactures tea and sells it in packets of 1kg. He knows that the packing process is imperfect, so that there is always a chance that any packet that is finally sold will have a tea content exceeding 1kg or less than 1 kg. In the current process, it is possible to set the packing machine, so that the packet weighs within a certain range. As the government regulation forbids packets with weights lesser than what is specified on the packets, Mr. X has set the machine at a higher value, so that only packets with weights exceeding 1kg. will be produced. This has created a problem for him. He feels that currently he is losing a lot of money in the way of excess material being packed. He has got an option to go for a more sophisticated packing machine at a certain cost that will reduce the variability. He wants to find out whether it is worthwhile going for the new machine. Say, the new process will produce packets with weight ranging from 1 to 1.05 kg., if set in the same manner.

A knowledge of p.d.f. of the weights produced by the current process will help Mr. X to calculate the probability that any packet will weigh more than, say, 1.05 kg. , or that any packet will weigh between 1.01 to 1.05 kg. These probabilities are helpful in his decision. A high probability of the weight exceeding 1.05 kg.is an indicator of a high percentage of packets having more than 1.05 kg.weight. These probabilities may help him calculate the expected loss due to the current process. This expected loss may be traded off then with the cost of buying the machine to arrive at the final decision.

Situation 2

Mr. T, a manufacturer of Electric bulbs, feels that the desired life of a bulb should be 100 hrs. , i.e. a new bulb should bum for 100 hrs. before the filament breaks. He realises that a high cost is associated with having a process that will manufacture all bulbs with life of more than 100 hrs. He is ready to make a trade off between the quality level and the cost.

In this case, if he knows the p.d.fs. of "the life (in hours)" of bulbs manufactured through different processes, then for different processes he can find out the probabilities that the life will exceed or equal 100 hrs. Suppose, he found the following for two processes

P(Iife $\leq$ 100 hrs.) = .8 for process 1

P(life $\geq$ 100 hrs.) = .9 for process 2

The above formula indicates that the process 2 is a better process, so far as quality is concerned. One may note that the cost for process 2 is higher than that of process 1. Mr X may now try to decide whether it is worthwhile paying extra cost for this quality.

The above formula shows how the information on p.d.f. can be helpful in decision making. This brings us to the question of assessing a p.d.f. Like we have seen in the cast-of discrete variables, for continuous variables also may real life situations can be approximated by certain theoretical distribution functions. Knowledge about the process of interest, and the past data, on the variable help us to find out what type of standard (theoretical) p.d.f. is to be applied in a particular situation.

We now present two important theoretical probability density functions, viz., the Exponential and the Normal. A study of the properties of these functions will be helpful in characterising the probability distributions in a variety of situations.

Exponential Distribution

Time between breakdown of machines, duration of telephone calls, life of an electric bulb are examples of situations where the Exponential distribution has been found useful. In the previous unit, while discussing the discrete probability distributions, we have examined the Poisson process and the resulting Poisson distribution. In the Poisson process, we were interested in the random variable of number of occurrences of an event within a specific time or space. Thus, using the knowledge of Poisson process, we have calculated the probability that 0, 1, 2 .....accidents will occur in any month. Quite often, another type of random variable assumes importance in the context of a Poisson process. We may be interested in the random variable of the lapse of time before the first occurrence of the event. Thus, for a machine, we note that the first failure or breakdown of the machine may occur after 1 month or 1.5 months etc. The random variable of the number of failures within a specific time, as we have already seen, is discrete and follows the Poisson distribution. The variable, time of first failure, is continuous and the Exponential p.d.f. characterises the uncertainty.

If any situation is found to satisfy the conditions of a Poisson process, and if the average occurrence of the event of interest is m per unit time, then the number of occurrences in a given length of time t has a Poisson distribution with parameter mt, and the time between any two consecutive occurrences will be Exponential with parameter m. This can be used to derive the p.d.f. of the Exponential distribution.

Let f(t) denote the p.d.f. of the time between occurrence of the event

F(t) denote the c.d.f. of the time between occurrence of the event (say, $t > 0$).

Let A be the event that time between occurrence is less than or equal to t.

and B be the event that time between occurrence is greater than t.

By definition, as A and B are mutually exclusive and collectively exhaustive : P(A) + P(B) = 1 .... (1)

From the definition of c.d.f. and the description of event A,

P(A) = F(t) ....(2)

From the definition of event B, as the time between occurrence is greater than t, it implies that the number of occurrences in the interval (0, t) is zero. Taking the distribution of number of occurrences in time t as Poisson, we can write':

P(B) = Probability that zero occurrences are there in time t, given that the average number of occurrences are mt.

From Poisson formula, P(B) can be written as :

$$P(B) = \frac{e^{-mt} x (mt)^0}{0!} = e^{-mt} \ldots\ldots(3)$$

From (1), (2) and (3), we have :

$$F(t) + e^{-mt} = 1$$

or, $$F(t) = 1 - e^{-mt}$$

Differentiating, we arrive at the p.d.f. :

$$f(t) = \begin{cases} m\,e^{-mt}, & t > 0 \\ 0, & \text{otherwise.} \end{cases}$$

The above formula gives the pdf of the Exponential Distribution. We can now verify as to whether this is a valid pdf.

Wefind f(t) $\geq$ for all t as m>0

also $\int_0^\infty f(t)\,dt = \int_0^\infty m\,e^{-mt}dt = 1$

Hence this is a valid p.d.f.

If we assume that the occurrence of an event corresponds to customers arriving for servicing, then the time between the occurrence would correspond to the inter-arrival time (IAT), and m would correspond to the arrival rate. Exponential has been used widely to characterise the IAT distribution. The Exponential p.d.f. is also used for characterising service time distributions. The parameter `m' in that case, corresponds to the service rate. We take up an example to show the probability calculations using the Exponential p.d.f. In the final section of this unit, we will be illustrating through an example, the use of the Exponential distribution in decision-making.

Example 5

A highway petrol pump can serve on an average 15 cars per hour. What is the probability that for a particular car, the time taken will be less than 3 minutes?

Solution

Here, Exponential applies with m = 15 (service rate). We are interested in finding the probability that t <3 minutes i.e. $t < \frac{3}{60}$ hrs

From definition of c.d.f., we want to Find $F\left(\frac{3}{60}\right) = F\left(\frac{1}{20}\right)$

we have seen that $F(t) = 1 - e^{-mt}$

$$F\left(\frac{1}{20}\right) = 1\text{-}e^{-15\times 1/20} = 1\text{-}e^{-3/4} = .5276.$$

Example 6

The distribution of the total time a light bulb will burn from the moment it is first put into service is known to be exponential with mean time between failure of the bulbs equal to 1000 hrs. What is the probability that a bulb will burn more than 1000 hrs.

Solution

$$\text{Here, } m = \frac{1}{1000}$$

$$\text{and } f(t) = \begin{cases} \frac{1}{1000}\, e^{-t/1000} & \text{for } t \geqslant 0 \\ 0 & \text{otherwise} \end{cases}$$

We are interested in finding the probability that t >1000 hrs.

$$P(t > 1000) = 1 - p(t \leqslant 1000) = 1 - F(1000)$$
$$F(1000) = 1 - e^{-1000 \times 1/1000} = 1 - e^{-1}$$

$\therefore$ The required probability = e^{-1} = 0.368.

Activity D

In Example 5, find the probability that for any car, the time taken to service will be more than 10 minutes. Discuss how this probability and the probability you have found in Example 5, can be useful for the petrol pump owner.

…………………………………………………………………………………………
…………………………………………………………………………………………
…………………………………………………………………………………………
…………………………………………………………………………………………
…………………………………………………………………………………………
…………………………………………………………………………………………
…………………………………………………………………………………………

Activity E

In Example 6, find the probability that the life of any bulb will lie between 100 hrs. and 120 hrs. Elaborate as to how this information may be useful to the manufacturer of the bulb.

…………………………………………………………………………………………
…………………………………………………………………………………………
…………………………………………………………………………………………
…………………………………………………………………………………………

Normal Distribution

The Normal Distribution is the most versatile of all the continuous probability distributions. It is found to be useful in Statistical inferences, in characterising uncertainties in many real-life processes, and in approximating other probability distributions.

Quite often, we face the problem of making inferences about processes based on limited data. Limited data is basically a sample from the full body of data on the process. Irrespective of how the full body of data is distributed, it has been found that the Normal Distribution can be used to characterise the sampling distribution. This helps considerably in Statistical Inferences.

Heights, weight and dimensions of a product are some of the continuous random variables which are found to be normally distributed. This knowledge helps us in calculating the probabilities of different events in varied situations, which in turn is useful for decision-making.

Finally, the Normal Distribution can be used to approximate certain probability distributions. This helps considerably in simplifying the probability calculations.

In the next few paragraphs we examine the properties of the Normal Distribution, and explain the method of calculating the probabilities of different events using the distribution. We then show the Normal approximation to Binomial distribution to '" illustrate how the probability calculations are simplified by using the approximation. An application of the Normal Distribution in decision-making is presented in the last section of the unit. The use of this distribution is Statistical Inferences is taken up in a later Block.

Properties of the Normal Distribution

The p.d.f. of the Normal Distribution is given by :

$$F(x) = \frac{1}{\sigma\sqrt{2\pi}}\ e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2 dy} \quad -\infty < x < \infty \text{ …………………..(1)}$$

where, π and e are two constants with values 3.14 and 2.718 respectively. The μ and σ are the two parameters of the distribution, and x is a real number denoting the continuous random variable of interest.

The c.d.f. is given by:

$$F(X) = \int_{-\infty}^{x} \frac{1}{\sigma\sqrt{2\pi}}\ e^{-\frac{1}{2}\left(\frac{y-\mu}{\sigma}\right)^2 dy}$$

It is apparent from the above that f is a positive function, $e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$ being positive for any real number x. It can be shown that

$\int_{-\infty}^{+\infty} f(x)\,dx = 1$, so that f(x) is a valid p.d.f, The interested reader may look up the book by Gangolli et. al. for proof,

The mean and the standard deviation are respectively denoted by μ and σ. Thus, different values of these two parameters lead to different `nominal curves'

The inherent similarity in all the `normal curves' can be seen by examining the `Standardised curve'. The Standard Curve with μ = 0 and σ = 1 is obtained by using $Z = \frac{x - \mu}{\sigma}$, so that we get the p.d.f.

$$f(z) = \frac{1}{\sqrt{2\pi}}\ e^{-\frac{1}{2}z^2} \quad -\infty < x < \infty \text{ …………………………..(2)}$$

The p.d.f. (1) is referred to as the regular form, while the p.d.f. (2) is known as the standard form. Normal Distribution with mean μ and standard deviation a is generally denoted by N(m, σ).

For large value of n, it is possible to derive the above p.d.f. as an approximation to the Binomial Distribution. The p.d.f. cannot be integrated analytically. The c.d.f. is tabulated for N(0,1) and the probabilities are calculate with the help of this table.

The plot of f(x) vs. x gives the Normal curve, and the area under the curve gives the probability. The Normal Distribution is symmetric about the mean; the area on each side of the mean is 0.5. The area between $\mu + K_1\sigma$ and $\mu + K_2\sigma$ is the same for all Normal curves irrespective of the values of μ and σ.

Though the range of the variable is specified from - ∞ to ∞, 99.7% of the values of the random variable fall within $\pm 3\sigma$ limits, that is, $P(\mu - 3\sigma \le x \le \mu + 3\sigma) = .997$. Moreover, it is known that 95.4% and 68.6% of the values of the random variable lie between $\pm 2\sigma$ and $\pm 1\sigma$ limits respectively.

Because of the symmetry, and the points of inflexion at $\pm 1\sigma$ distance, the Normal curve has a bell shape. The right and left tails of the curve extend indefinitely without touching the horizontal line.

Probability Calculation

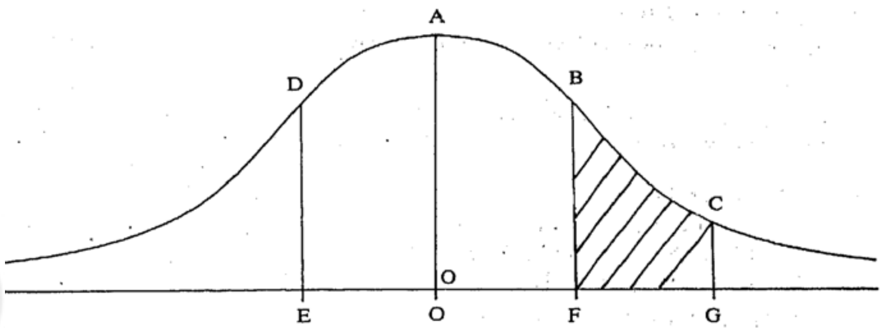
Suppose, it has been found that the duration of a particular project is normally distributed with a mean of 12 days and a standard deviation of 3. We are interested in finding the probability that the project will be completed in 15 days..

Given the μ and σ of the random variable of interest, we first find

$$Z = \frac{x - \mu}{\sigma} \qquad \text{Hence, } \mu = 12,\ \sigma = 3 \text{ and } x = 15,\ \therefore\ Z = \frac{15 - 12}{3} = 1$$

The values of the probabilities corresponding to Z are tabulated and can be found from the table. The Standard Normal being a symmetrical distribution, the table for one half (the right half) of the curve is sufficient for our purpose. The table gives the probability of Z being less than equal to a particular value.

Consider the following diagram depicting the Standardised Normal curve, denoted by N(0,1). The probability of Z lying between 1 and 2 can be represented by the area under the curve between Z values of 1 and 2; that is, the area represented by FBCG in the diagram given below.



Because of the symmetry, the area on the right of OA = area on the left of OA = 0.5. If you now look up a `normal table' in any basic Statistics text book, you will find that corresponding to Z = 1.0, the probability is given as 0.3413. This only implies that the area OABF = 0.3413, so that,

$P(Z \le 1) = 0.5 + 0.3413 = 0.8413$, the area to the left of OA being 0.5.

Similarly, corresponding to Z = 2.0, we find the value 0.4772 (area OACG = 0.4772). This implies,

$P(Z \le 2) = 0.5 + 0.4772 = 0.9772$

$\therefore$ If we are interested in the shaded area FBCG, we find that, FBCG = Area OACG - Area OABF = 0.4772 - 0.3413 = 0.1359.

$\therefore$ $P(1 \le Z \le 2) = 0.1359$.

The area, hence the probability, corresponding to a negative value of Z can be found from symmetry. Thus, we have the area OADE = the area OABF = 0.3413.

$\therefore$ $P(Z \leq -1) = 0.5 - 0.3413 = 0.1587$.

Returning to our example, we are interested in finding the probability that the project duration is less than or equal to 15 days. Denoting the random variable by T, we know that T is N(12, 3).

$$\therefore P(T \leq 15) = P\left(\frac{T-12}{3} \leq \frac{15-12}{3}\right) = P(Z \leq 1) = 0.5 + 0.3413 = 0.8413.$$

Similarly, if we were interested in finding out the chances that the project duration will be between 9 and 15 days, we can proceed in a similar way.

$$\therefore P(9 \leq T \leq 15) = P\left(\frac{9-12}{3} \leq \frac{T-12}{3} \leq \frac{15-12}{3}\right) =$$

$$P(-1 \leq Z \leq 1) = 0.3413 + 0.3413 = 0.6813.$$

(Note that this confirms our earlier statement that 68% of the values lie between $\pm 1\sigma$ limit.)

Normal as an Approximation to Binomial

For large n and with p value around 0.5, the Normal is a good approximation for the Binomial. The corresponding p, and a for the Normal are np and $\sqrt{npq}$ respectively. Suppose, we want to find the probability that the number of heads in a toss of 12 coins will lie between 6 and 9. From the previous unit, we know that this probability is equal to :

$$\sum_{r=6}^{9} {}^{12}C_r \left(\tfrac{1}{2}\right)^r \left(\tfrac{1}{2}\right)^{12-r}$$

As such, this tedious calculation can be obviated by assuming that the random variable, number of heads (H), is Normal with mean = np and $\sigma = \sqrt{npq}$. Here μ = 12 x 0.5 = 6 and $\sigma = \sqrt{12 \times 0.5 \times 0.5} = \sqrt{3} = 1.732$

$\because$ assuming H is N (6, 1.732), we can find the probability that H lies between 6 and 9. The following continuity correction helps in better approximation. Instead of looking for the area under the Normal curve between 6 and 9, we look up the area between 5.5 and 9.5, i.e. 0.5 is included on either side.

$$\therefore P(5.5 \leq H \leq 9.5) = P\left(\frac{5.5-6}{1.732} \leq \frac{H-6}{1.732} \leq \frac{9.5-6}{1.732}\right)$$

$$\therefore P(-.289 \leq Z \leq 2.02).$$

From the table, corresponding to Z = 0.289 and 2.02 we find the values 0.114 and 0.4783.

$\therefore$ the required probability = 0.114 + 0.4783 = 0.5923, Now you may check that by using the Binomial distribution, the same probability can be calculated as 0.5934.

Fractile of a Normal Distribution

The concept of Fractile as applied to Normal Distribution is often found to be useful. The kth fractile of $N(\mu, \sigma)$ can be found as follows. First we find the k^{th} fractile of the N(0,1). Let Z_k be the K^{th} fractile of N(0,1).

By definition, $F(Z_k) = K$, $(0 < K < 1)$.

Say, if Z_k is the $.975^{th}$ fractile of N(0,1), then

$F(Z_k) = 0.975$, $P(Z \leq Z_k) = 0.975 = 0.5 + 0.475$.

From the table, we find that corresponding to Z = 1.96, the probability is 0.475. Hence $Z_k = 1.96$. Now suppose that we are interested in the 0.975^{th} fractile of N(50,6). If X_k be the required fractile,

then $$\frac{X_k - \mu}{\sigma} = Z_k$$

$$\therefore X_k = \mu + Z_k\sigma = 50 + 1.96 \times 6 = 61.76$$

From symmetry, the $.025^{th}$ fractile of N(50,6) can be seen to be = 50 - 1.96 x 6 = 38.24.

Activity F

A ball-bearing is manufactured with a mean diameter of 0.5 inch and a standard deviation in diameters of .002 inch. The distribution of the diameter can be considered to be normal. The bearing with less than .498 inch and more than .0502 inch are considered to be defective. What is the probability that a ball - bearing manufactured through this process will be defective ?

…………………………………………………………………………………………
…………………………………………………………………………………………
…………………………………………………………………………………………
…………………………………………………………………………………………
…………………………………………………………………………………………
…………………………………………………………………………………………
…………………………………………………………………………………………

Activity G

Suppose from the above exercise, you have found that the probability of a defective is 0.32. If the bearing are packed in lots of 100 units and sent to the supplier, what is the probability that in any such lot, the number of defectives will be less than 27? (The probability corresponding to Z value of 1.07 is 0.358.)

…………………………………………………………………………………………
…………………………………………………………………………………………
…………………………………………………………………………………………
…………………………………………………………………………………………
…………………………………………………………………………………………
…………………………………………………………………………………………
…………………………………………………………………………………………
…………………………………………………………………………………………
…………………………………………………………………………………………

11.4 APPLICATIONS OF CONTINUOUS DISTRIBUTIONS

The following two examples illustrate the use of the Exponential and the Normal Distribution in decision-making.

Example 7

A TV manufacturer is facing the problem of selecting a supplier of Cathode-ray tube which is the most vital component of a TV. Three foreign suppliers, all equally dependable, have agreed to supply the tubes. The price per tube and the expected life of a tube for the three suppliers are as follows :

	Price/tube	Expected life per tube
Supplier 1	Rs. 800	1500 hrs.
Supplier 2	Rs. 1000	2000 hrs.
Supplier 3	Rs. 1500	4000 hrs.

The manufacturer guarantees its customers that it will replace the TV set if the tube fails earlier than 1000 hrs. Such a replacement will cost him **Rs.** 1000 per tube, over and above the price of the tube.

Can you help the manufacturer to select a supplier?

Solution

The Expected cost per tube for each supplier can be found as follows :

Expected cost per tube = price per tube + expected replacement cost per tube.

Expected replacement cost per tube is given by the product of the cost of replacement and the probability that a replacement is needed. Both the cost of replacement and the probability vary from supplier to supplier. We note that, a replacement is called for if the tube fails before 1000 hrs., so that, for each supplier we can calculate the P(life of tube $\leq$ 1000 hrs.). This probability can be calculated by assuming that the time between failure is exponential. Thus, $p(t \leq 1000)$ is basically exponential with

$$m = \frac{1}{1500}, \frac{1}{2000}, \text{ and } \frac{1}{4000} \text{ for the three suppliers}$$

$$p(t \leq 1000/m = \frac{1}{1500}) = F_1(1000) = 1 - e^{-1000/1500} = .4866$$

$$p(t \leq 1000/m = \frac{1}{2000}) = F_2(1000) = 1 - e^{-1000/2000} = .3935$$

$$\text{and } p(t \leq 1000/m = \frac{1}{4000}) = F_3(1000) = 1 - e^{-1000/4000} = .2212$$

Once the expected costs for each supplier are known, we can take a decision based on the cost. The calculations are shown in the table below :

Supplier Number	Price per tube	Cost per Replacement	P(life 1000 hrs.)	Expected cost per tube
	P	C	P	E=(P+Cp)
1	800	1800	.4886	1679.48
2	1000	2000	.3935	1787
3	1500	2500	.2212	2053

We find that for the supplier 1, the expected cost per tube is the minimum. Hence the decision is to select 1.

Example 8

A supplier of machined parts has got an order to supply piston rods to a big car manufacturer. The client has specified that the rod diameter should lie between 2.541 and 2.548 cms. Accordingly, the supplier has been looking for the right kind of machine. He has identified two machines, both of which can produce a mean diameter of 2.545 cms. Like any other machine, these machines are also not perfect. The standard deviations of the diameters produced from the machine 1 and 2 are 0.003 and 0.005 cm. respectively, i.e. machine 1 is better than machine 2. This is reflected in the prices of the machines, and machine 1 costs Rs. 3.3 lakhs more than machine 2. The supplier is confident of making a profit of Rs. 100 per piston rod; however, a rod rejected will mean a loss of Rs. 40.

The supplier wants to know whether he should go for the better machine at an extra cost.

Solution

Assuming that the diameters of the piston rods produced by the machining process is normally distributed, we can find the probability of acceptance of a part if produced in a particular machine.

For machine 1, we find that the diameter is N(2.545,.003), and for machine 2, we find that the diameter is N(2.545,.005)

If D denote the diameter, then :

$2.541 \leq D \leq 2.548$, implies the rod is accepted.

Probability of acceptance if a rod is produced in machine 1

$$= P(2.541 \leq 2.548)$$

$$= P\left(\frac{2.541 - 2.545}{.003} \leq Z \leq \frac{2.548 - 2.545}{.003}\right)$$

$$= P(-1.33 \leq Z \leq 1)$$

$$= .4066 + .3413 = .7479 \text{[from N(0,1) table]}$$

Hence probability of rejection = 1 - .7479 = .2521

Expected profit per rod if machine 1 is used

= 100 x .7479 - 40 x .2521 = Rs. 64.706 ....(1)

Similarly, if machine 2 is used, we can find the expected profit per rod

Probability of acceptance here

$$= p\left(\frac{2.541 - 2.545}{.005} \leq Z \leq \frac{2.548 - 2.545}{.005}\right)$$

$= p(-.8 \leq D \leq .6)$

$= .2881 + .2257 = .5138$

Probability of rejection = 1 - .5138 = .4862

Expected profit per rod if machine 2 is used

$= 100 \times .5138 - 40 \times .4862 = \text{Rs. } 31.932$ .... (2)

Thus, from (1) and (2), we find that the expected profit per part is more if machine 1 is used. As machine 1 costs 3.3 lakh more than machine 2, it will be profitable to use machine 1 only if the production is more.

We can find the breakeven production level as follows.

Let N be the number of rods produced, for which both the machines are equally profitable.

Then $N \times (64.706 - 31.932) = 3,30,000$

or; $N \approx 10,069$

This implies that it is advisable to go in for machine 1, only if the production level is higher than 10,070. (Note that we assume that there is enough demand for the rods.)

Activity H

Suppose in Example 8, you have decided that machine 1 should be used for production. Assume now, that this machine has got a facility by which one can set the mean diameter, i.e., one can set the machine to produce any one mean diameter ranging from 2.500 to 2.570 cm. Once the machine is set to a particular value, the rods are produced with mean diameter equal to that value and standard deviation equal to 0.003 cm. If the profit per rod and loss per rejection is same as in example 8, what is the optimal machine setting?

……………………………………………………………………………………
……………………………………………………………………………………
……………………………………………………………………………………
……………………………………………………………………………………
……………………………………………………………………………………
……………………………………………………………………………………
……………………………………………………………………………………
……………………………………………………………………………………

11.5 SUMMARY

The function that specifies the probability distribution of a continuous random variable is called the probability density function (p.d.f.). The cumulative density function (c.d.f.) is found by integrating the p.d.f. from the lowest value in the range upto an arbitrary level x. As a continuous random variable can take innumerable values in a specified interval on a real line, the probabilities are expressed for interval rather than for individual values. In this unit, we have examined the basic concepts and assumptions involved in the treatment of continuous probability distributions. Two such important distributions, viz., the Exponential and the Normal have been presented. Exponential distribution is found to be useful for characterising uncertainty in machine life, length of telephone call etc., while dimensions of machined parts, heights, weights etc. found to be Normally distributed. We have examined the properties of these p.d.fs. and have seen how probability calculations can be done for these distributions. In the final section, two examples are presented to illustrate the use of these distributions in decision-making.

11.6 FURTHER READINGS

Chance, W., 1969. *Statistical Methods for Decision Making, R.* Irwin Inc.: Homewood.

Feller, W., 1957. *An Introduction to Probability Theory and Its Applications,* John Wiley & Sons Inc.: New York.

Gangolli, R.A. and *D. Ylvisaker. Discrete Probability,* Harcourt Brace and World. Inc.: New York.

Levin, R., 1984. *Statistics for Management,* Prentice-Hall Inc.: New York.

Parzen, E., 1960. *Modern Probability Theory and Its Applications,* Wiley: New York.