UNIT 11 CONTINUOUS PROBABILITY DISTRIBUTIONS*

Structure

11.0 OBJECTIVES

After going through this unit you should be able to

- explain and use the normal distribution;
- explain the concept of the degrees of freedom; and
- form some elementary ideas about the chi-squared distribution, the student's-*t* distribution and the *F* distribution.

11.1 INTRODUCTION

In the previous Unit, we made a distinction between a discrete random variable and a continuous random variable. In that unit, we introduced the concept of a probability distribution and found that it is essentially a statement regarding the values taken by a random variable with their associated probabilities. We studied two important discrete probability distributions namely, the binomial distribution and the Poisson distribution. In the present Unit we will continue with the topic and study a very important continuous probability distribution called the normal distribution. It may be mentioned that the normal distribution plays an important role in the statistical inference and tests of hypotheses and these are going to be our subject matter of Block-7.

In fact, we will consider in Unit 13 the topic of sampling distribution that forms the foundation of statistical inferences and tests of hypotheses. However,

* Dr. Anup Chatterjee (retd.), ARSD College, University of Delhi

sampling distribution can be properly appreciated only if we have some rudimentary ideas about three other continuous probability distributions besides the normal distribution, viz., the chi-squared distribution, the student's-*t* distribution and the *F* distribution. We will discuss about these probability distributions below.

11.2 NORMAL DISTRIBUTION

Normal distribution is perhaps the most widely used distribution in Statistics and related subjects. It has found applications in inquiries concerning heights and weights of people, IQ scores, errors in measurement, rainfall studies and so on. Abraham de Moivre gave the mathematical equation for the normal distribution in 1733. Karl Friedrich Gauss also independently derived its equation from a study of errors in repeated measurements of the same quantity. Accordingly, sometimes it is also referred to as the Gaussian distribution. The distribution has provided the foundation for much of the subsequent development of mathematical statistics.

We have seen in the previous Unit that for a continuous random variable, the counterpart of a probability mass function is the *probability density function.* We shall denote the probability density function also by $p(x)$. The probability density function of a continuous random variable that follows the normal distribution is given by

$$p(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$$

where $-\infty < x < \infty$, and

$\pi = 3.17141$ (approximately)

$e = 2.71828$ (approximately).

It is clear that the normal density function is completely determined by the parameters μ and σ. It means that given the values of μ and σ, we can trace out the normal curve by obtaining the values of $p(x)$ for different values of x. In fact, it can be shown that μ and σ are respectively the mean and the standard deviation of the normal distribution. When a random variable X follows normal distribution with mean μ and standard deviation σ we write it in symbols as $X \sim N(\mu, \sigma)$ and read as 'X follows normal distribution with mean μ and standard deviation σ. The normal curve is a symmetrical bell-shaped curve as shown in Fig. 11.1.

The important features of the normal distribution are as follows:

1) The normal curve stretches from $-\infty$ to $+\infty$. This means that a normal random variable (X) assumes values between $-\infty$ to $+\infty$.

2) The curve is symmetric about its mean, i.e., $\overline{x} = \mu$. This means that corresponding to $x = \mu + a$ and $x = \mu - a$, the values of $p(x)$ are the same (for any arbitrarily chosen 'a').

3) The median and the mode of the distribution coincide with the mean. Thus mean = median = mode = μ.

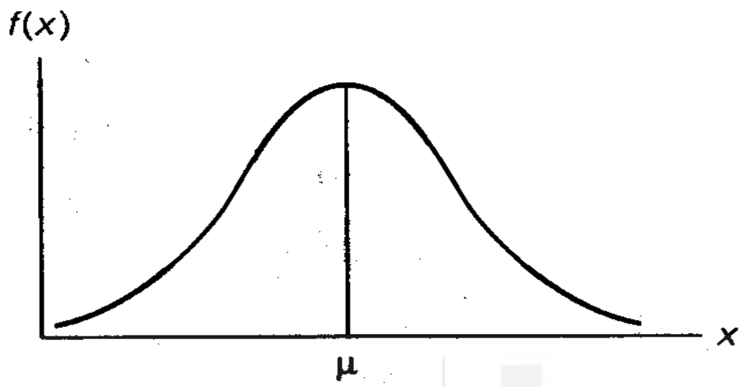

Fig 11.1: Normal Curve

4) The maximum of the normal curve occurs at $x = \mu$. Thus $p(x)$ is maximum when x= μ.

5) The points of inflexion of the normal curve occurs at $x = \mu + \sigma$ and $x = \mu - \sigma$. At the points of inflexion, the normal curve changes its curvature.

The following area-properties hold for a normal distribution. In Fig.11.2 below we plot a normal curve with mean $\mu = 50$ and standard deviation $\sigma = 4$.

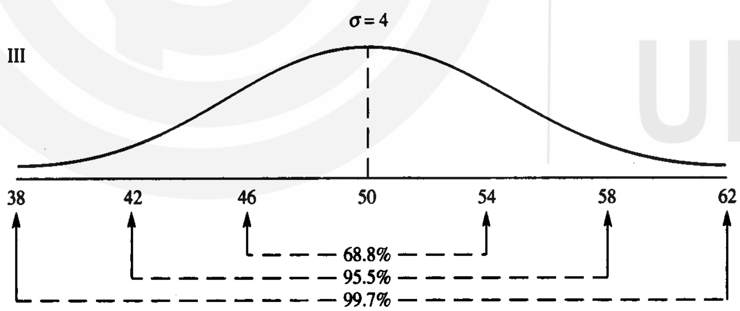

Fig. 11.2: Area under Normal Curve

(a) 68.8 % of the area under the normal curve lies between the ordinates at $\mu - \sigma$ and $\mu + \sigma$. Thus in Fig.11.2, 68.8% area is covered when x ranges between 46 and 54.

(b) 95.5% of the area under the normal curve lies between the ordinates at $\mu - 2\sigma$ and $\mu + 2\sigma$. In Fig. 11.2, 95.5% area is covered when $42 \leq X \leq 58$.

(c) 99.7% of the area (i.e., almost the whole of the distribution) under the normal curve lies between the ordinates at $\mu - 3\sigma$ and $\mu + 3\sigma$. In Fig.11.2 we find that 99.7% area is covered when $38 \le X \le 62$.

If we have different values of μ and σ, the range of x mentioned in Fig. 11.2 will change.

11.2.1 Standard Normal Curve

We have seen in the previous unit that the curve for any continuous probability distribution or probability density function is so traced out that the area under the curve bounded by the two ordinates corresponding to $x = x_1$ and $x = x_2$ gives the probability that the random variable assumes a value between $x = x_1$ and $x = x_2$. Thus, for a normal curve

$$P(x_1 \le X \le x_2) = \int_{x_1}^{x_2} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} dx = \frac{1}{\sqrt{2\pi}\sigma}\int_{x_1}^{x_2} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} dx$$

Obviously, this probability depends upon the values of the two parameters μ and σ. However, it is very difficult to solve the above-mentioned integral of the normal distribution. This has necessitated the tabulation of normal curve areas for quickly obtaining the probabilities of the normal variable assuming values in different intervals. But, it is really meaningless to attempt to construct a separate table for every conceivable combination of values for μ and σ. Fortunately, the solution to an apparently hopeless task has been achieved by the application of a standard result in statistics that we have seen and proved in the last unit. Let us recapitulate the result. We have seen that

For any variable with a given mean and standard deviation, whenever the mean is subtracted from the variable and the result is divided by the standard deviation; the resultant variable has a mean equal to zero and a standard deviation equal to one.

Thus if X is a variable with mean (expectation) μ and standard deviation σ then $z = \frac{X-\mu}{\sigma}$ has a mean equal to zero and standard deviation equal to one.

It means that normal variables with different combinations of μ and σ can all be transformed into a unique normal variable with mean 0 and standard deviation 1.

Thus if X is a normal variable with mean (expectation) μ and standard deviation σ, then $z = \frac{X-\mu}{\sigma}$ for any combination of μ and σ, is always a normal variable with mean 0 and standard deviation 1.

Symbolically,

If $X \sim N(\mu, \sigma)$

then,

$$z = \frac{X-\mu}{\sigma} \sim N(0,1)$$

for any μ and σ.

Such a transformed normal variable is called a *standard normal variate*. The probability density function of the standard normal variate z is given by

$$p(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2} \qquad -\infty < z < \infty$$

Once we obtain a standard normal variate, our seemingly hopeless task of obtaining probability areas for different combinations of μ and σ becomes elegantly simple. Let us see how. We should note that a standard normal variate has a unique mean of 0 and a unique standard deviation of 1. It means, if we can construct a table for probability areas of such a unique standard normal variate, it can be used for obtaining probability for any normal variable with any combination of mean and standard deviation. The only thing is that the given normal variable is to be transformed into the standard normal variate. In fact, such a table for areas (or probability) has been compiled for a standard normal variate (see Appendix Table A1: Area under the Standard Normal Curve given at the end of this book) and is very much in use in statistics. Thus, for the computation of the required probability for any normal variable with some mean and standard deviation, the upper and the lower limits say, $x = x_1$ and $x = x_2$ of the given interval are converted into the corresponding z-values say, $z = z_1$ and $z = z_2$ and the relevant area is obtained from Appendix Table A1 given at the end of this book.

Remember that the standard normal curve is symmetrical and it covers an area of 1.0 Since the value of z ranges between $-\infty$ and ∞, we find that the area between $0 < z < \infty$ is 0.5 (half the area under standard normal curve). Similarly, the area between $-\infty < z < 0$ is 0.5. Since the standard normal curve is symmetric we have one advantage; the area under the curve is the same on both sides. In Appendix Table A1 the area for different positive values of z are given.

If we look into column 1 of Appendix Table A1 we find that values assumed by z ranges from 0.0 to 3.0. Corresponding to each value there are 10 columns marked .00, .01, ..., .09. These columns represent the second digit after decimal. For example, if $z = 0.45$, then we look for the row corresponding to 0.4. On this row we move to the right and look for the column representing .05. In Table A1 we find that when z=0.45 the area covered is 0.1736. Note that when $z = -0.45$ the area under standard normal curve again is 0.1736. As another example, the area under the standard normal curve when $z = 1.31$ is 0.4049.

Theoretically z can assume any value between $-\infty$ and ∞. However, when z = 3.09 the area covered is 0.4990. Therefore, in Table A1 areas for $z > 3.09$ are not given.

Let us now consider some examples to see the applications of the normal area table.

Example 11.1

Find the area under the standard normal curve in each of the following cases by using Appendix Table A1 for areas under the standard normal curve.

a) Between $z = 0$ and $z = 1.8$.

b) Between $z = -0.25$ and $z = 0$.

c) Between $z = -0.52$ and $z = 2.25$.

Solution:

a) In Table A1, let us move downward under the column marked z until we reach the entry 1.8. Now, let us turn right to the column marked 0. We find here an entry equal to 0.4641. This is the required area.

b) Since the standard normal curve is symmetric about the mean, the required probability between $z = -0.25$ and $z = 0$ can be obtained by finding the area between $z = 0$ and $z = 0.25$ from the table. So, let us move downward under the column marked z until we reach the entry 0.2. Then we turn right to the column marked .05. We find here an entry equal to 0.0987. This is the required area.

c) It is clear that the required area is

(area between $z = -0.52$ and $z = 0$) + (area between $z = 0$ and $z = 2.25$)

= (area between $z = 0$ and $z = 0.52$) (by symmetry) + (area between $z = 0$ and $z = 2.25$)

= 0.1985 + 0.4878

= 0.6863.

Example 11.2

In a sample of 120 workers in a factory the mean and standard deviation of daily wages are Rs. 11.35 and Rs. 3.03 respectively. Find the percentage of workers getting wages between Rs. 9 and Rs. 17 in the whole factory assuming that the wages are normally distributed.

Solution: Let x be a random variable denoting wages. Then, x is a normal variable with mean $\mu = 11.35$ and the standard deviation $\sigma = 3.03$. The corresponding standard normal variate

$$z = \frac{x - 11.35}{3.03}$$

For $x = 9,\ z = \frac{9 - 11.35}{3.03} = \frac{-2.35}{3.03} = -0.78$

For $x = 17,\ z = \frac{17 - 11.35}{3.03} = \frac{5.65}{3.03} = 1.86$

$\therefore\ P(9 \le x \le 17) = P(-0.78 \le z \le 1.86)$

$= P(-0.78 \le z \le 0) + P(0 \le z \le 1.86)$

$= P(0 \le z \le 0.78) + P(0 \le z \le 1.86)$

$= 0.2823 + 0.4686$

$= 0.7509.$

Thus, 75.09% of workers get wages between Rs. 9 and Rs. 17.

11.2.2 Normal Approximation to the Binomial Distribution

Sometimes in statistics, one distribution is obtained as the limiting form of another distribution. For example, in Unit 10, we have learnt that at times the probability of success, *p*, in a binomial distribution is very small and the number of trials, *n*, is so large that the expectation, $\mu = np$, is a finite quantity. In such cases the binomial distribution tends to Poisson distribution. You can recall that both the binomial distribution and the Poisson distribution are discrete distributions. However, the limiting form of the discrete binomial distribution need not be uniquely the discrete Poisson distribution. In fact, when *n* is very large and *p* is not extremely close to 0 or 1; the binomial distribution approaches the continuous normal distribution. As a result, in real world situations, the normal distribution is often used for approximating the binomial distribution. We have already seen that the standard normal table comes very handy in obtaining probabilities for a normal variable falling within a specified interval. We may now state a result that helps us to use areas under the standard normal curve to approximate the binomial properties of a random variable:

If *X* is a binomial variable with mean $\mu = np$ and variance $\sigma^2 = npq$ then

$$z = \frac{X - np}{\sqrt{npq}}$$

tends to a normal distribution with mean zero and standard deviation one, as *n* tends to infinity

In symbols,

$$\lim_{n \to \infty} \frac{X - np}{\sqrt{npq}} \sim N(0,1)$$

It has been observed that the normal distribution provides a good approximation for a binomial distribution even when *n* is not very large and *p* is reasonably close to 0.5.

Let us consider an example.

Example 11.3

Suppose the probability of a particular kind of machine being defective is 0.4. A quality control inspector examines a lot of 15 such machines to identify the defective machines. Find the probability that the inspector will find 4 machines to be defective.

Solution: We know from our discussion on binomial distribution that the distribution of defective machines in a given lot is a binomial variable. Here

$n = 15$, $p = 0.4$ and $q = 1 - p = 0.6$

The probability of 4 defective machines is given by

$$^{15}c_4 (0.4)^4 (0.6)^{11} = \frac{15!}{4!11!}(0.4)^4 (0.6)^{11} = 1365 \times 0.0256 \times 0.0036279 = 0.1268$$

Now suppose, we want to find the required probability by the normal approximation. Then

$np = 15 \times 0.4 = 6$, $npq = np(1-p) = 15 \times 0.4 \times 0.6 = 3.6$ and $\sqrt{npq} = \sqrt{3.6} = 1.897$

We have

$$z = \frac{X - np}{\sqrt{npq}}$$

which is approximately a standard normal variate. But the standard normal variate is a continuous random variable. We know that for such a random variable, the probability of its taking a particular value cannot be determined. It is only the probability of the random variable lying within an interval that can be obtained. Thus the probability of the binomial variable taking a value 4 has to be translated into the probability of the corresponding normal variable falling in an interval around the value 4 for the required normal approximation. Since, a binomial variable assumes zero and positive integer values; the value immediately preceding 4 and the value immediately succeeding 4 that the binomial variable under question can take are 3 and 5 respectively. As a result, the probability of the binomial variable taking a value equal to 4 can be reasonably approximated by the probability of the corresponding normal variable falling within an interval (3.5, 4.5). The required probability is thus approximately equal to the area under the normal curve between the ordinates x_1 = 3.5 and x_2 = 4.5. Converting to z-values, we have

$$z_1 = \frac{x_1 - np}{\sqrt{npq}} = \frac{3.5 - 6}{1.897} = -1.32 \text{ and } z_2 = \frac{x_2 - np}{\sqrt{npq}} = \frac{4.5 - 6}{1.897} = -0.79.$$

If X is the binomial variable and z is the corresponding standard normal variate, then $P(X = 4) = P(-1.32 \le z \le -0.79) = 0.1214$ (From the area under standard normal curve given in Appendix Table A1)

We can see that the normal approximation of 0.1214 agrees quite closely with the actual probability of 0.1268 for 4 defective machines obtained from the binomial distribution.

Check Your Progress 1

1) Find the area under the normal curve in the following cases.
   a) Between $z = 1.55$ and $z = 2.55$.
   b) To the left of $z = -1.5$.
   c) To the right of $z = 2.5$.

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

2) The mean height of 1000 men is 68 inches and the standard deviation is 5 inches. If the heights are normally distributed, find how many men have heights between 67 inches and 69 inches.

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

3) A company, that sells 5000 batteries in a year, guarantees them for a life of 24 months. The life of the batteries is estimated to be approximately normal with mean equal to 34 months and standard deviation equal to 5 months. Find the number of batteries that will have to be replaced under the guarantee.

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

11.3 SOME OTHER CONTINUOUS DISTRIBUTIONS

There are some other continuous probability distributions that play important roles in various branches of statistics. In Block 4, we will study statistical inference. In that Block, in addition to normal distribution, we will often use concepts of three more continuous distributions, C*hi-Squared* (Pronounced as *kai-squared* and denoted by the symbol χ^2), the *Student's-t distribution* and the *F-distribution.* In this section, we are going to discuss these distributions in brief. We begin with a general concept, the 'degrees of freedom', which finds applications in all these distributions.

11.3.1 Degrees of Freedom

In connection with these distributions, we shall often come across a term: degrees of freedom. Let us get some idea about the term now. In a general sense, *the degrees of freedom refers to the number of pieces of independent information* that are required to compute some characteristic (say, variance) of a given set of observations. We consider an example here.

Suppose there are 5 observations: 4, 7, 12, 3 and 15. Hence, the arithmetic mean $\bar{X}$ is 8. The computation of variance $\frac{1}{n}\sum(x_i - \bar{x})^2$ involves obtaining the squares of the deviations of the values of the observations from their arithmetic mean and adding them as shown below:

$$(4-8)^2 + (7-8)^2 + (12-8)^2 + (3-8)^2 + (14-8)^2$$
$$= (-4)^2 + (-1)^2 + (4)^2 + (-5)^2 + (6)^2$$

From the properties of arithmetic mean ($\bar{X}$), we know that the summation of the values inside the brackets must be equal to zero, i.e., in general $\sum_{i=1}^{n}(x_i - \bar{X}) = 0$. It means that in the computation of the variance, if the first four terms inside the brackets happen to be –4, –1, 4 and –5 respectively, the fifth term cannot be any other term but 6. Thus in this example, we do not have 5 independent pieces of information inside the brackets. The fifth piece of information inside the bracket, i.e., '6' depends upon the first four pieces of information inside their respective brackets.

We can fix the idea better if we think of a person who does not have any idea about the individual observations. She is only told that there are 5 observations and the first 4 deviations of the values from the mean of the five observations are –4, –1, 4 and –5 respectively. She is then asked to calculate the variance of the 5 observations. If she knows the law that the sum of the deviations of the values of a variable from their arithmetic mean is always equal to zero, she will at once be able to figure out that the deviation of the 5th value from its arithmetic mean is 6. She will then proceed to take the squares of these deviations and add them together to arrive at the last step for the computation of the required variance. The last step involves a division by the number of observations to get an idea about the average dispersion of the values about their arithmetic mean (we take variance as the required measure here). What is the appropriate value that she should take for the number of observations? Is it 5? Let us probe a little into it. We have seen that the fifth deviation is determined by the first four deviations, which means that there are 4 independent pieces of information that produce the dispersions of the given 5 values about their arithmetic mean. Therefore, for measuring the average dispersion, i.e., the variance, the sum of the squares of the five deviations should be divided by 4 and not by 5.

Thus, in this example, the degrees of freedom are 4. Generalising the above example, for obtaining the variance of *n* observations, there are n–1 degrees of freedom because of the restriction $\sum_{i=1}^{N}(x_i - \overline{X}) = 0$. As a result, for variance, $\sum_{i=1}^{N}(x_i - \overline{X})^2$ is divided by n–1, i.e., $S^2 = \frac{1}{n-1}\sum_{i=1}^{n-1}(x_i - \overline{X})^2$, where, S^2 is the variance. From the above discussion, we can say that the degrees of freedom that we have for the computation of any characteristic is equal to the number of observations minus the number of restrictions put on the computation of the required characteristic. In symbols, *d.f.* = *n* – *r* where, *d.f.* is the degrees of freedom, *n* is the number of observations and *r* is the number of restrictions.

We should note here that when *n* is quite large, for calculating the variance or its positive square root, i.e., the standard deviation, we often divide $\sum_{i=1}^{N}(x_i - \overline{X})^2$ by *n* and not by *n –1*. However, strictly speaking, we should divide $\sum_{i=1}^{N}(x_i - \overline{X})^2$ by *n –1*, particularly when *n* is small.

11.3.2 The χ^2 (Chi-Squared) Distribution

Suppose, X is a normal variable with mean (expectation μ and standard deviationσ, than $z = \frac{X-\mu}{\sigma}$ is a standard normal variate, i.e., $z \sim \mathrm{N}(0,1)$ If we take the square of z, i.e., $z^2 = \left(\frac{X-\mu}{\sigma}\right)^2$, then z^2 is said to be distributed as a χ^2 variable with one degree of freedom and expressed as $\chi_1^{\,2}$.

It is clear that since $\chi_1^{\,2}$ is a squared term; for z laying between $-\infty$ and $+\infty$, $\chi_1^{\,2}$ will lie between 0 and $+\infty$ (because a squared term cannot take negative values). Again since, z has a mean equal to zero, most of the values taken by z will be close to zero. As a result, the probability density of χ^2 variable will be maximum near zero.

Generalizing the result mentioned above, if $z_1, z_2, \ldots, z_k$ are independent standard normal variates (i.e., normal variables with zero mean and unit variance), then the variable

$$z = \sum_{i=1}^{ki} z_1^2$$

is said to be a χ^2 variable with k degrees of freedom and is denoted by χ_k^2 Fig, 15.3 given below shows the probability curves for χ^2 variables with different degree of freedom.

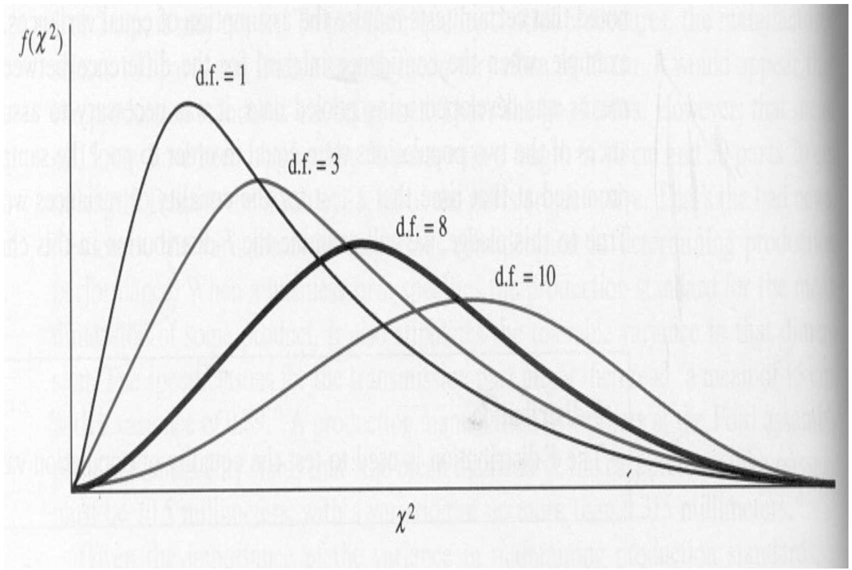

Fig 11.3: Chi-squared Probability Curves

We should note the following features of the χ^2 distribution.

1) As Fig. 11.3 shows, the χ^2 is a positively skewed distribution. Its degree of skewness depends on its degrees of freedom. For lower degrees of freedom, the distribution is highly skewed. As the number of degrees of freedom increases, the distribution becomes increasingly symmetric. In fact, for degrees of freedom more than 100, the variable $\sqrt{2\chi^2} - \sqrt{(2k-1)}$ can be treated as a standard normal variate, where *k* is the degrees of freedom.

2) The mean of the chi-squared distribution is *k*, and its variance is *2k*, where *k* is the degrees of freedom.

3) If Z_1 and Z_2 are two independent chi-squared variables with k_1 and k_2 degrees of freedom respectively, then $Z_1 + Z_2$ is also a chi-squared variable with degrees of freedom equal to $k_1 + k_2$.

As in the case of the normal distribution, a similar table has been prepared for the χ^2 distribution also (see Appendix Table A2 at the end of this book). We have to just consult this table to obtain the required probability of the χ^2 variable for different degrees of freedom.

In Table A2 *df* represent degrees of freedom. While the columns $\chi^2_{0.05}$ and $\chi^2_{0.01}$ denotes χ^2 values for 5% and 1% *level of significance* respectively. We will explain the concept of level of significance in Unit 14.

The following example illustrates the use of the chi-squared table.

Example 11.4

What is the probability of obtaining a χ^2 value of 34 or greater, given the degrees of freedom 25?

Solution: We can see from Appendix Table A2 that if we move down the degrees of freedom column to reach the figure of 25, the nearest figure to 34 that we find across the corresponding row is 34.08747. The probability for 34.08704, as we can see from the table, is 0.10. Hence the required probability is 0.10.

11.3.3 The Student's-*t* Distribution

W.S. Gosset presented the *t*-distribution. The interesting story is that Gosset was employed in a brewery in Ireland. The rules of the company did not permit any employee to publish any research finding independently. So, Gosset adopted the pen-name 'student' and published his findings about this distribution anonymously. Since then, the distribution has come to be known as the student's-*t* distribution or simply, the *t* distribution.

If z_1 is a standard normal variate, i.e., $z_1 \sim N(0,1)$ and z_2 is another independent variable that follows the chi-square distribution with k degrees of freedom, i.e.,

$z_2 \sim \chi_k^2$, then the variable $t = \frac{z_1}{\sqrt{(z_2/k)}} = \frac{z_1\sqrt{k}}{\sqrt{z_2}}$

is said to follow student's-*t* distribution with *k* degrees of freedom.

The probability curves for the student's-*t* distribution for different degrees of freedom are presented in Fig. 11.4

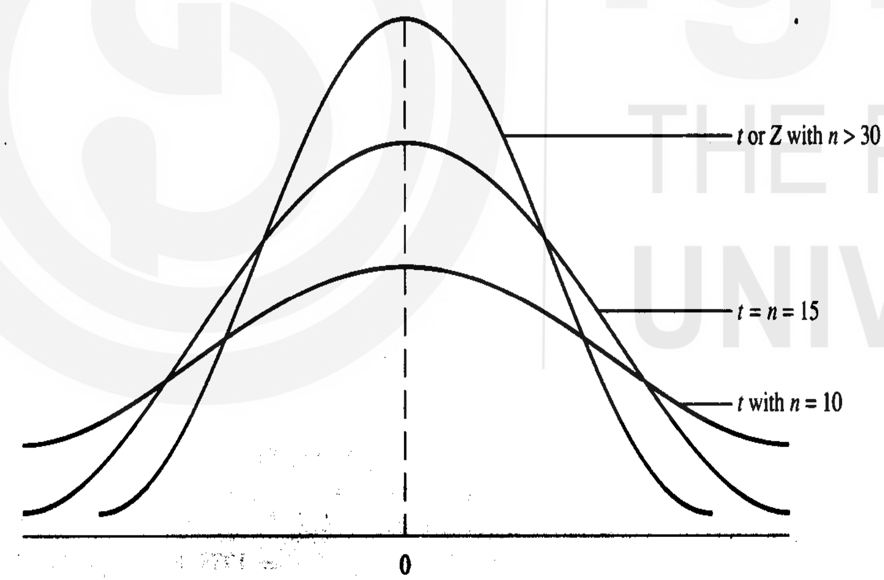

Fig 11.4: Student's-t Probability Curves

We may note the important characteristics of this distribution.

1) As we can see in Fig. 11.4, like the normal distribution, the student's-*t* distribution is also symmetric and its range of variation is also from – ∞ to + ∞; however, it is flatter than the normal distribution. We should also

note here that as the degrees of freedom increase, the student's-*t* distribution approaches the normal distribution.

2) The mean of the student's-*t* distribution is zero and its variance is $\frac{k}{(k-2)}$, where, k is the degrees of freedom.

Like the normal distribution, the student's-*t* distribution is often used in statistical inferences and tests of hypotheses to be discussed in Block-7. The task involves the integration of its density function; which may prove to be tedious. As a result, in this case also, like the normal distribution, a table has been constructed for ready-reference purposes (see Appendix Table A3).

We shall consider an example to see the use of Table A3.

Example 11.5

Given the degrees of freedom equal to 10, what is the probability of obtaining a *t* value of (i) 2.7638 or greater, (ii) –2.7638 or lower?

Solution:

(i) In Table A3 for student's-*t* distribution, first, we move down the degrees of freedom column and reach the figure of 10 and then look across the corresponding row and locate the figure of 2.7638. The corresponding lower probability figure of 0.01 is the required probability.

(ii) Since the student's-*t* distribution is symmetric, the probability of obtaining a *t* value of –2.7638 or lower is also 0.01.

11.3.4 The *F* Distribution

Another continuous probability distribution that we discuss now is the *F* distribution.

If z_1 and z_2 are two chi-squared variables that are independently distributed with $k_1$and k_2 degrees of freedom respectively, the variable

$$F = \frac{z_1/k_1}{z_2/k_2}$$

follows s *F* distribution with k_1 and k_2 degrees of freedom respectively. The variable is denoted by F_{k_1,k_2} where, the subscripts k_1 and k_2 are the degrees of freedom associated with the chi-squared variables. We may note here that k_1 is called the numerator degrees of the freedom and in the same way, k_2 is called the denominator degrees of freedom.

Some important properties of the F distribution are mentioned below.

1) The *F* distribution, like the chi-squared distribution, is also skewed to the right. But, as k_1 and k_2 increase, the *F* distribution approaches the normal distribution.

2) The mean of the F distribution is $k_1/(k_2 - 2)$, which is defined for $k_2 > 2$, and its variance is $\frac{2k_2^2(k_1 + k_2 - 2)}{k_1(k_2 - 2)^2(k_2 - 4)}$ which is defined for $k_2 > 4$.

3) An F distribution with 1 and k as the numerator and denominator degrees of freedom respectively is the square of a student's-t distribution with k degrees of freedom. Symbolically,

$$F_{1,k} = t_k^2$$

4) For fairly large denominator degrees of freedom k_2, the product of the numerator degrees of freedom k_1 and the F value is approximately equal to the chi-squared value with degrees freedom k_1, i.e., $k_1 F = \chi^2_{k_1}$.

As we have mentioned earlier with reference to other continuous probability distributions, F distribution is also extensively used in statistical inference and testing of hypotheses. Again, such uses also require obtaining areas under the F probability curve and consequently integrating the F density function. However, in this case also our task is facilitated by the provision of the F Table.

11.3.5 Distributions Related to the Normal Distribution

We have already seen from the features of the chi-squared, student's-t and the F distributions that for large degrees of freedom, these distributions approach the normal distribution. Consequently, these distributions are also known as the distributions related to the normal distribution. This relationship between the chi-squared distribution, the student's-t distribution and the F distribution on one hand and the normal distribution on the other has tremendous practical implications. When the degrees of freedom happen to be fairly large, instead of using the chi-squared distribution or the student's-t distribution or the F distribution separately as the situation may demand; we can uniformly apply the normal distribution. As a result, our task gets considerably simplified.

Check Your Progress 2

1) What is the probability of obtaining a χ^2 value of 8 or greater, given the degrees of freedom 20?

……………………………………………………………………………......…

……………………………………………………………………………......…

……………………………………………………………………………......…

2) Given the degrees of freedom equal to 25, what is the probability of obtaining a t value of 1.708 or greater?

…………………………………………………………………………......

…………………………………………………………………………...

...…………………………………………………………………………

3) Given k_1 = 10 and k_2 = 8, what is the probability of obtaining an F value of 5.8 or greater?

……………………………………………………………………………………

……………………………………………………………………………………

……………………………………………………………………………………

……………………………………………………………………………………

11.4 LET US SUM UP

In this unit, we studied some continuous probability distributions. Among these distributions, the normal distribution is considered to be the most important one. We have learnt its characteristics and seen its practical applications. We have considered the important concept of a standard normal variate. We have also learnt the technique of using the table for the areas under the standard normal curve for solving problems relating to the normal distribution. Besides the normal distribution, we have considered three other continuous probability distributions. These are: the chi-squared distribution, the student's-*t* distribution and the *F* distribution. These three distributions use the notion of the degrees of freedom. So, we have tried to explain the concept of the degrees of freedom. We have learnt the characteristics of these distributions and seen how these distributions can be applied to various situations by using tables relating to these distributions. Finally, we have seen that for fairly large degrees of freedom, these three distributions approach the normal distribution.

11.5 ANSWERS OR HINTS TO CHECK YOUR PROGRESS EXERCISES

Check Your Progress 1

1) a) 0.0552
   b) 0.0668
   c) 0.0062
2) The proportion of persons having heights between 67 inches and 69 inches = 0.1586

   Therefore, the number of persons having heights between 67 inches and 69 inches = 1000 × 0.1586 = 159 (approximately).

3) $$z = \frac{24-34}{5} = -2$$

   From the standard normal table, the area between z = 0 and z = 2 is 0.4772. Therefore, the area between z = 0 and z = –2 is 0.4772 (because the standard normal distribution is symmetric). For finding the number of batteries that will have to be replaced, we have to consider the area to the left of z = –2, which is equal to the area to the right of z = 2. Now, the area to the right of z = -2 is 0.5 – 0.4772 = 0.0228. Therefore, the probability that a battery is defective is 0.0228. Thus out of 5000 batteries, the number of batteries that will have to be replaced $= 0.0228 \times 5000 = 114$.

Check Your Progress 2

1) 0.99
2) 0.05
3) 0.01