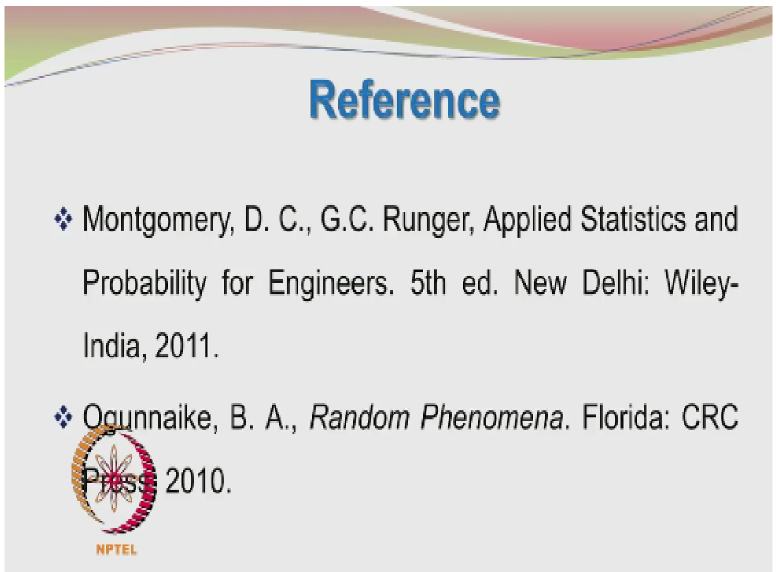
Statistics for Experimentalists
Prof. Kannan. A
Department of Chemical Engineering
Indian Institute of Technology – Madras

Lecture – 03
Discrete Probability Distributions

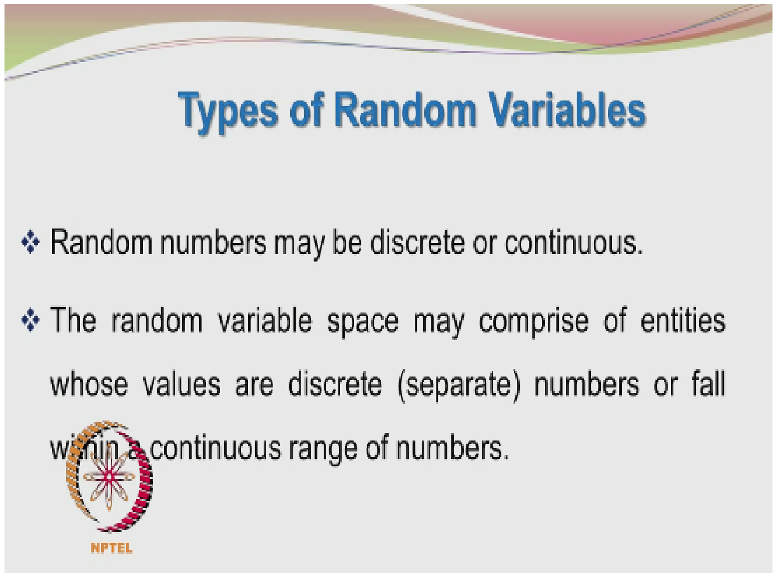
Hello again. We will be going to the next lecture today. We will be discussing on Discrete Probability Distributions.

(Refer Slide Time: 00:26)



The reference books for this topic are presented here; Montgomery and Runger's book, Applied Statistics and Probability for Engineers, 5th Edition, Wiley-India, 2011. Then, you have the book by Ogunnaike, Random Phenomena by CRC Press.

(Refer Slide Time: 00:53)



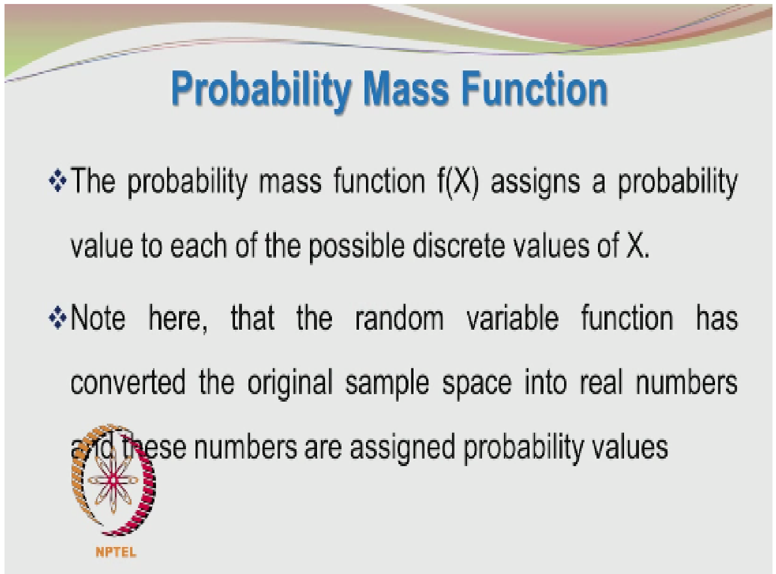
In the previous lecture, we saw about random variables, okay. When we conduct a random experiment, we are unsure about its outcome. So, the different possible values are called as random variables and they make take some specific real values, okay. The possible values of random numbers may be discrete entities or they may be continuous range of numbers.

What I mean by continuous range of numbers is, it may be falling between lower value and upper value, okay or a lower limit and an upper limit if you want it like that but in between these 2 limits, okay, it can take any possible value. As an example, if the random variable mole fraction is considered from an experiment, okay a distillation experiment for example. The mole fraction can take values only between 0 and 1, okay, both included.

But in between 0 and 1, it can take any value 0.12, 0.123, 0.1234 whatever. It depends upon the accuracy of your measurement device, okay. Of course, the calculator can give any value of the mole fraction running up to 7 or eight digits, okay. So, the mole fraction is a continuous random variable here. But when you throw a die, okay, you will find that the die a can take only discrete values.

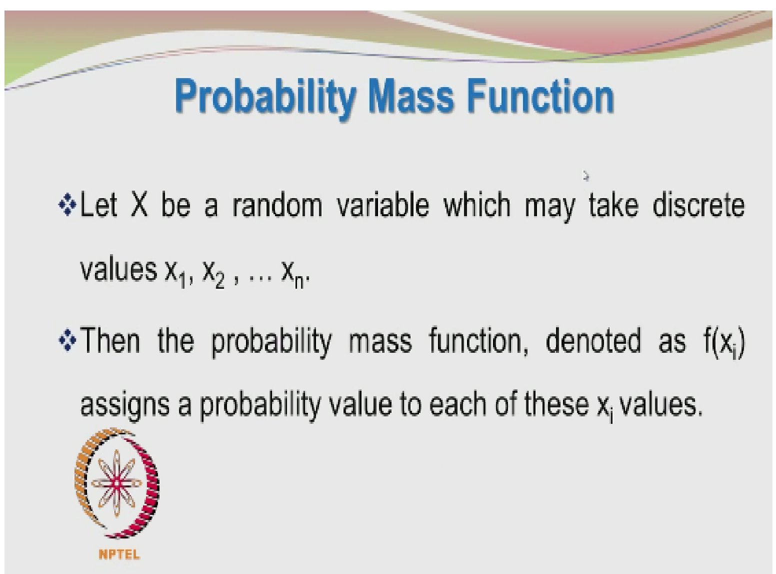
For example, one or 2 or 3 up to 6, okay. In this case, the random values or the random number values are discrete, okay.

(Refer Slide Time: 03:16)



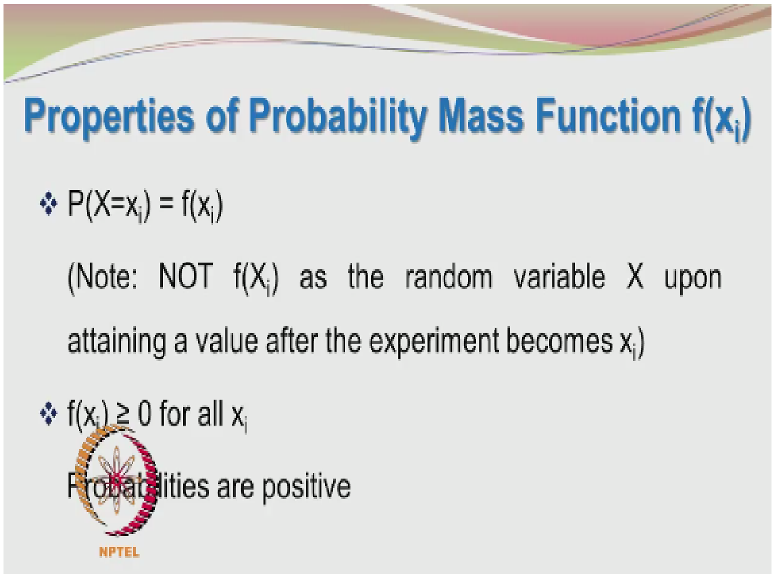
So, we are going to talk about a very interesting function, the probability mass function f of x assigns a probability value to each of the discrete values of X. If you recollect, the random variable function has converted the original sample space into real numbers and these numbers are assigned probability values. This we covered in one of the previous lectures, okay.

(Refer Slide Time: 04:14)



So, the random variable may take discrete values x1, x2, … so on to xn, okay. So, we now define a probability mass function and denote it as f of xi, okay and the role of the probability mass function is to assign a probability value to each of possible values xi, okay.

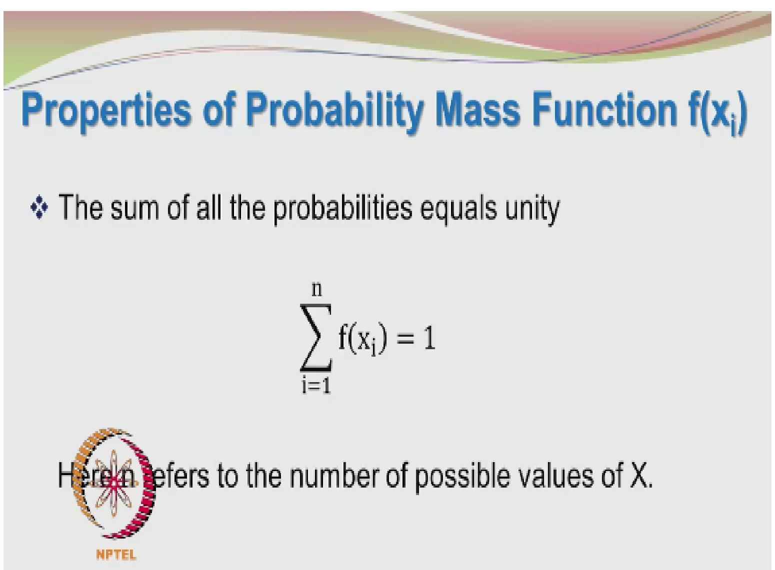
(Refer Slide Time: 04:51)



So, please note the definition again, probability of X=Xi, that means what is the probability that the random variable X takes the value small xi and that is given by the probability mass function f of xi, okay. When you write down the statements pertaining to statistical calculations, you have to make sure that the terminology use is correct, okay. The notation you use is correct. You cannot write f of xi, okay. You do not have Xi. You have only small xi.

You have X and small xi but you do not have XI, okay because once a random variable is assigned a value, it becomes small x of i and as we stated earlier f of xi should be >=0 for all the possible values of xi. It means that the probability values are all positive.

(Refer Slide Time: 06:23)

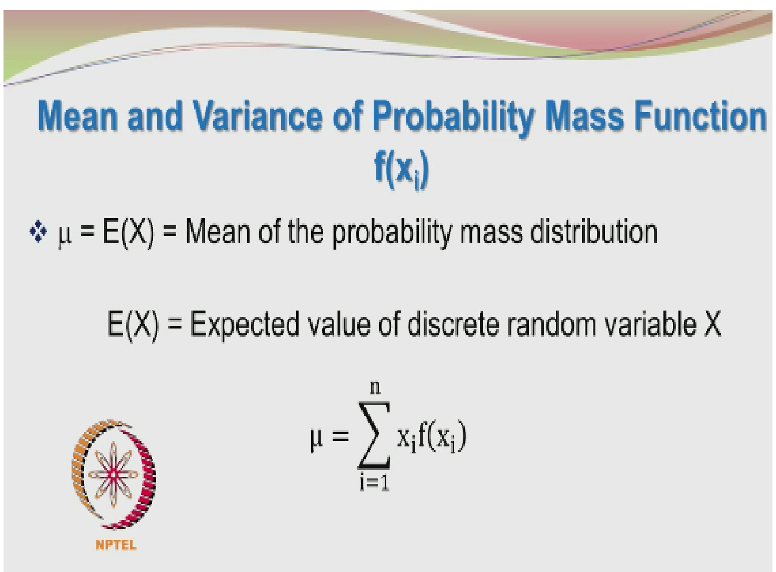


Now, the other important thing to observe about the probability mass function f of xi is the sigma I=1 to n, f of xi should be = 1. It means that the sum of all the probabilities is = 1. You have different your entities, each entity is having a probability value and the probability values of assigned by the probability mass function in such a way that when you add the total probabilities, they should be = 1, okay.

The probability value assigned to each of the xi may be same or they may be different, okay. For example, the probability of number coming up, let us say, 1. When you throw a fair die is 1/6. Similarly, the probability of number 2 coming is also 1/6. So, you have 6 numbers and all of them share the same probability of 1/6, okay. On the other hand, there may be other examples where you may have a discrete number of possible outcomes but all of them need not have the same probability value, okay.

They may be having different fractional probability values, but when you add them up they should become = 1, that is very important.

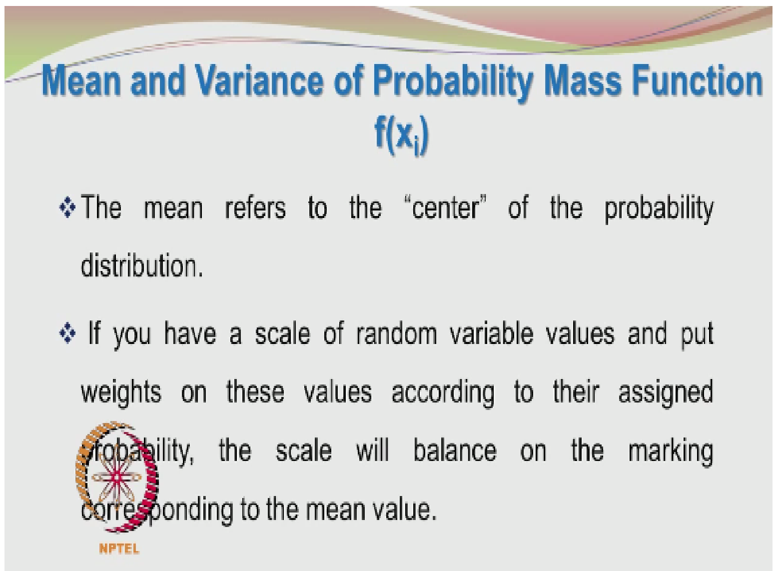
(Refer Slide Time: 08:07)



Mu=E of X, okay, that is the mathematical notation. Here, mu means the mean of the probability distribution, okay. E of X means the expected value of the random variable X, okay. The expected value of the random variable X is defined as mu=sigma i equals 1 to n xi f of xi, okay. The mean is also referred as the average okay.

When you take any cricketer okay and he has gone and performed in a test match series or even when he is coming out to bat after the previous wicket had fallen okay, the first thing which is shown on TV is the number of matches he has played, the number of runs he has scored and then the average, okay. The average is a measure of his overall performance, okay. If a batsman has scored an average of 50, then he is supposed to be very good but it does not mean that he will score 50 in the current innings. He may score 100 or he may score a duck, okay.

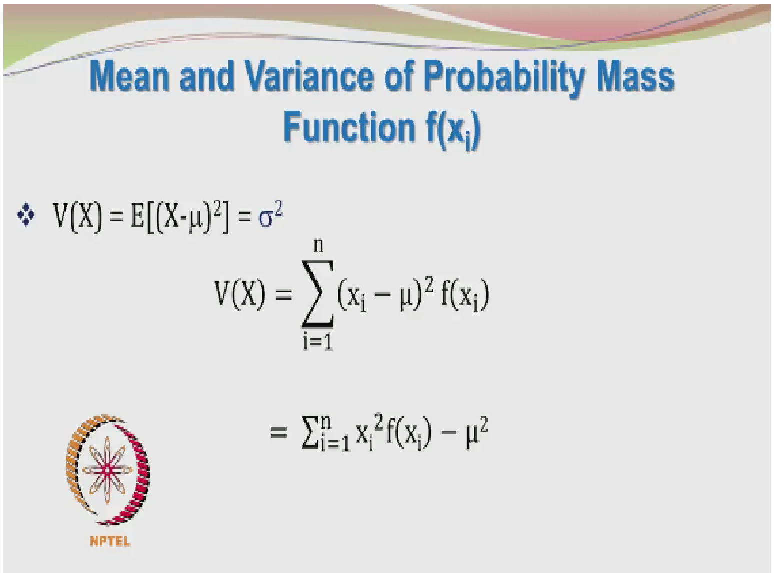
(Refer Slide Time: 09:45)



So, the mean refers to the centre of the probability distribution, okay. Imagine you have a long scale and on the scale you have markings corresponding to the possible values of the random variables, okay and each of these values will have a probability assigned to it, okay. Let us say that you have numbers 1, 2 and 3 and each of these numbers 1, 2 and 3 will have a probability value assigned to it, okay. Number 1 may have a probability of 0.2, number 2 may have a probability of 0.3 and number 3 may have a probability of 0.5, okay.

So, you put coins on the scale or you put weights on these numbers and you put the weights in such a way that it is proportional to the probability value, okay. So, if you put coins for number 2, you put more coins for number 3 and you put even more coins for number 5. The number of coins or weights you put for these values will be in proportion, okay. Then you put the scale on a knife edge and see where are the knife edge will balance, okay and the knife edge will balance at the mean value.

(Refer Slide Time: 11:28)



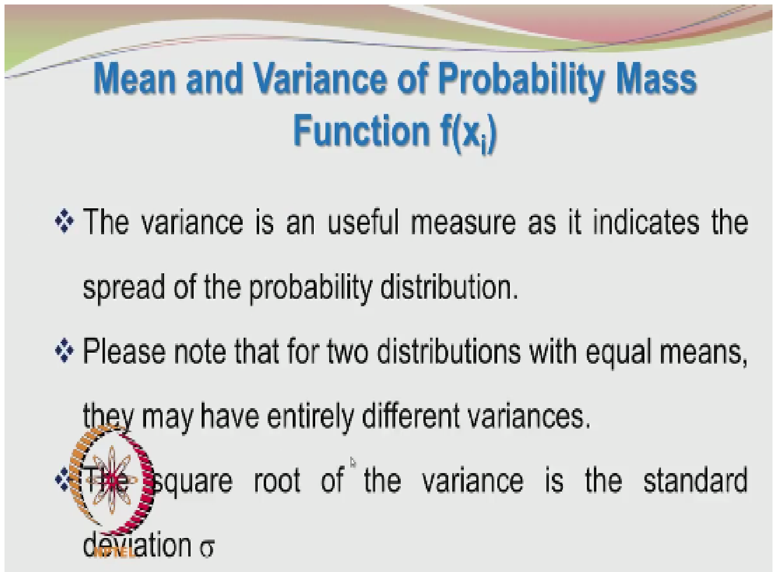
Now, let us come to the variance. We know from our earlier discussion because of the variability in the data you had to find the average. The batsman was not scoring 50 in all the innings, okay. He was scoring 0 in one, 100 in the next, 30 in the third, 70 in the 4th. So, when you take an average over a long period of time, you get the average of 50. So, the performance of the batsmen is variable, okay and that is why you had to find the mean or average value.

Now, the variability is quantified in terms of another parameter called as variance. We refer to it as variance of the random variable X and that is defined as the expected value of the deviation of the random variable from the mean square, okay. So, what you do is you first find the deviation of the random variables from the mean mu and then square it. This is called as deviation square and the expected value of this deviation square or the expected value of the square of the deviation = sigma square, okay.

So, now by using the definition for expectation, we write V of X=sigma i=1 to n xi-mu whole square f of xi, okay. Please do not forget to put f of xi, okay. You may recollect that the sum of all the deviations from the mean will be = 0. When you square the deviations from the mean you take the sum, they will not be = 0, okay and each of such square deviations, okay, should be multiplied by the corresponding f of xi.

When you expand this sigma=1, you get xi square f of xi-mu square. This derivation is quite simple and you may want to try it out yourself.

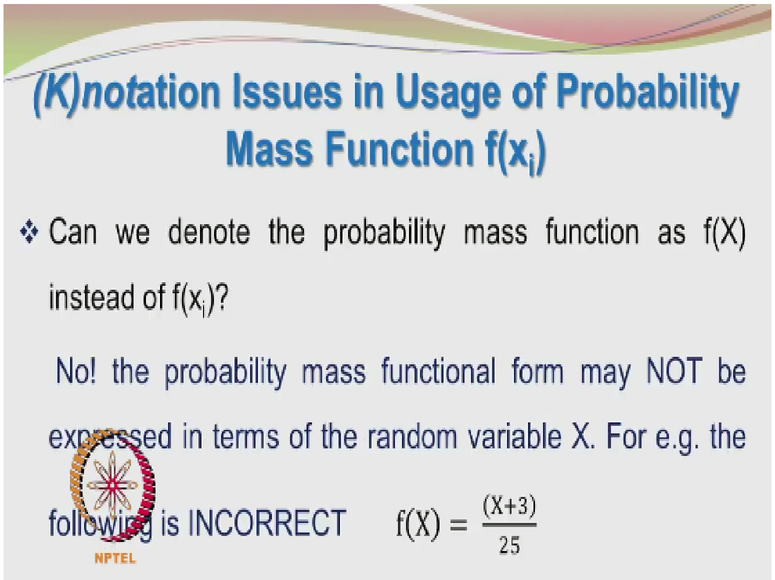
(Refer Slide Time: 13:50)



The variance is a useful measure as it indicates the spread of the probability distribution, okay. Even if there are 2 distributions having the same mean or average, it does not imply that they should have the same variances. You may have the same average but you may have different variances. For example, in one class the average may be 15 and in another class the average may be 50, okay. But in the first class the lowest mark may have been 20 and the highest mark may have been 100, okay.

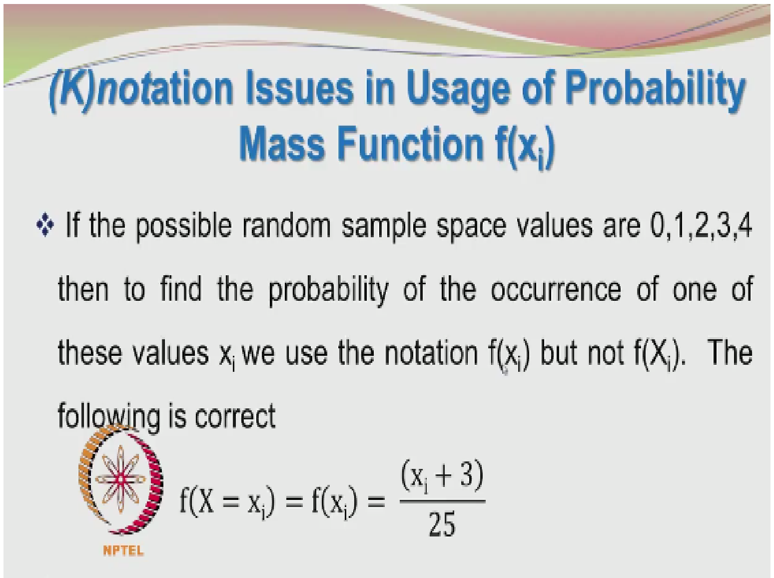
So, average is still 50 when you add up all the marks and divide by the total number. The second-class, the lowest mark may have been 40 and the highest mark may have been 60 and the average would have been 50, okay. There is a lot more variation in the first class in the marks when compared to the marks in the second class, even though both of them had the same average. By definition, the square root of the variance is the standard deviation. Variance is denoted by sigma square and so standard deviation is denoted by sigma.

(Refer Slide Time: 15:21)



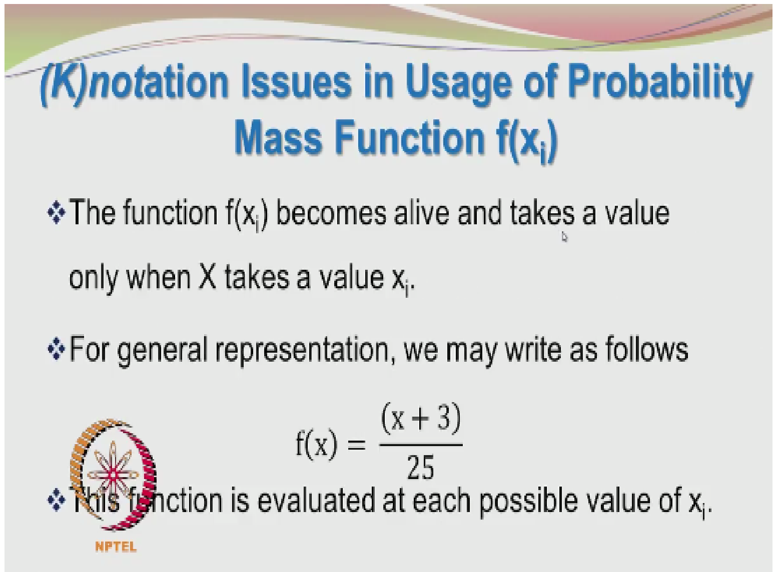
As I said earlier, we should look at the notation or terminology use whenever we use the language of statistics, okay. The grammar has to be correct. So, let us look at some of the notation issues in the usage of the probability mass function f of xi, okay. So, can we denote the probability mass function as f of X instead of f of xi, okay. You cannot put X in the argument of the probability distribution function or the probability mass function, okay. It should be f of xi. So, writing the probability mass distribution as f of X=X+3/25 makes no sense, okay.

(Refer Slide Time: 16:23)



If the possible random samples space values are 0, 1, 2, 3 and 4, then to find the probability of the occurrence of one of these values xi, we use the notation f of xi and not f of Xi, okay. The following form which I have shown here is correct. F of random variable x taking a value xi is given by f of xi and that is = xi+3/25, okay. This is correct notation.

(Refer Slide Time: 17:05)

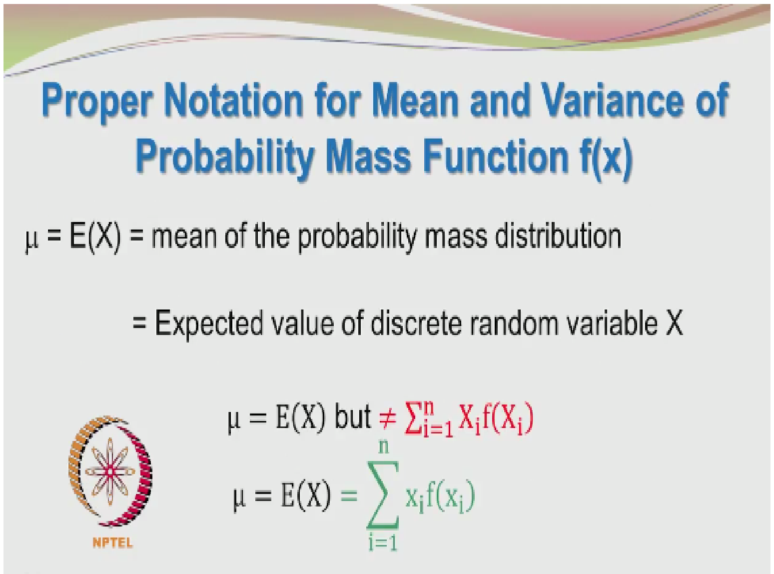


The function f of xi becomes alive and takes a value only when the random experiment has been performed and X random variable can take a particular value xi, okay. Before throwing the dice or before throwing a die, the random variable was x. Once you have thrown the die, perhaps when playing a game of Ludo or Snake and Ladders, you know what is the value shown by the die and then that value is xi, okay. Then, you can find what is the probability of this number xi equals one or 2 occurring and you can estimate that value.

Sometimes, you may not even know the probability beforehand, okay. In the case of a die, you know that all numbers are equally probable and so you can beforehand itself say that probably of number 1 occurring is = 1/6. But there are many random experiments whose outcome is not known for sure and you also cannot predict what that top probability of that outcome is going to be, okay. So, you have to conduct the random experiment, obtain the value of xi and then find out the value of f of xi, the probability value, okay.

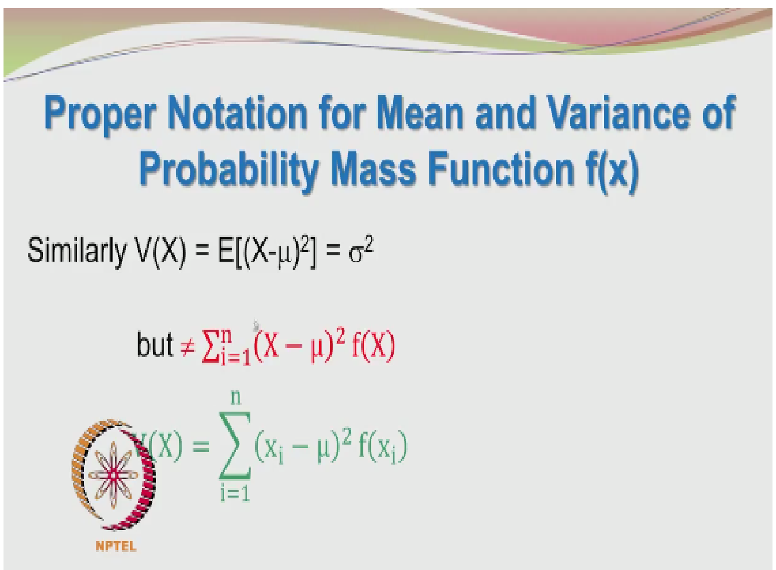
Generally, we can represent the probability density function. Sorry it is not probably density function. Probability density function is used for continuous distributions. Now, we are dealing one with discrete distribution. So, the probability function or the probability mass function f of x is given by x+3/25, small x is used throughout. This function is evaluated at each possible value of xi, okay.

(Refer Slide Time: 19:13)



So, we saw about the mean. When we write the definition for mean mu, we write it as expected value of X, okay. Expected value of the discrete random variable X. Mu=E of X but it is != Xi f of Xi where the Xi are capital Xi, this is wrong. When we actually go about calculating the expected value, then we use xi and f of xi and the argument in f of xi is also xi. So, do not use Xi in f of xi. Use xi*f of xi in finding the value of mean. So, this is important notation or terminology.

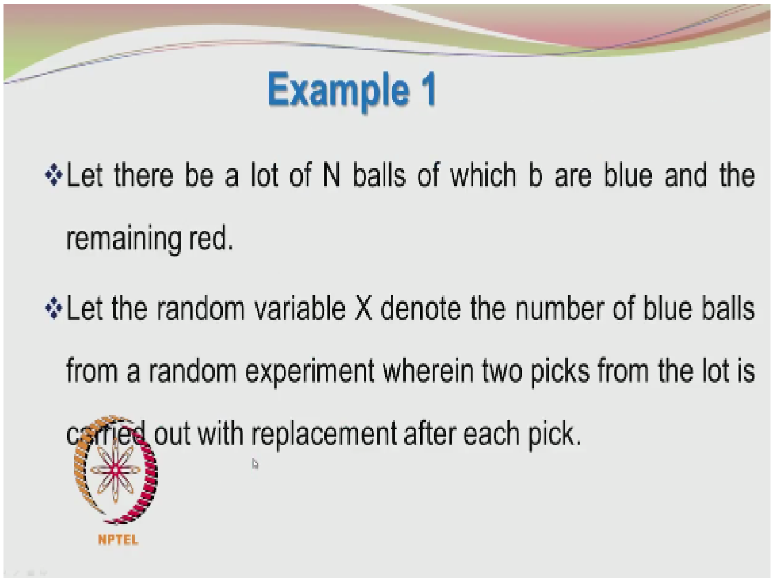
(Refer Slide Time: 20:12)



Similarly, when you are defining the variance of the random variable X, it has expected value of

E of X-mu whole square=sigma square but this should not be written as sigma=1 X-mu whole square f of x, okay. You should not use X here but you should rather use xi.

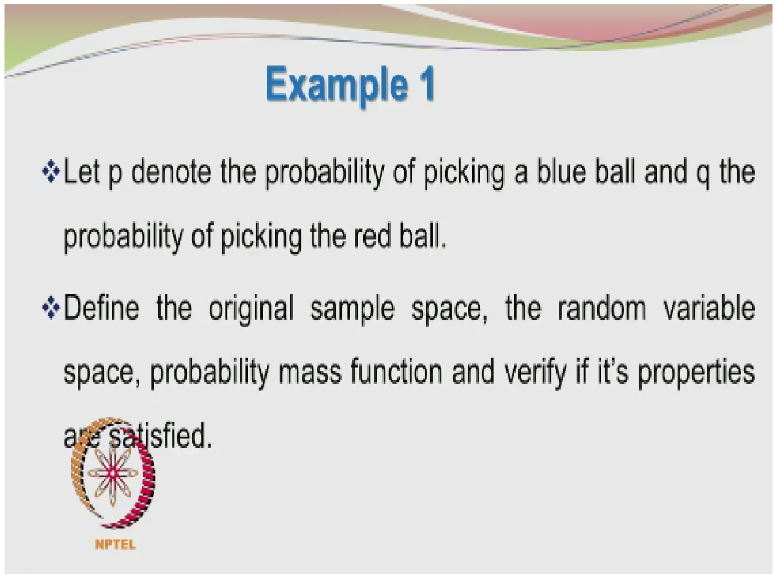
(Refer Slide Time: 20:36)



We will take a simple example, okay. So, let there be a lot of N balls of which b are blue in colour and the remaining are red, okay. I am sure that this kind of problems or variance of this problem you would have seen several times in the past. So, let us define the random variable as the number of blue balls that have been picked. Random experiment is performed by taking 2 picks from the lot, okay.

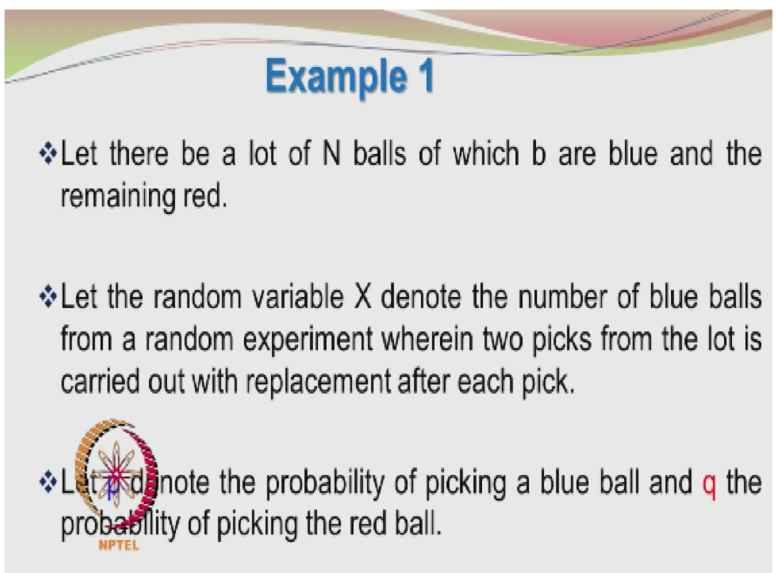
You take one ball, note the colour, put it back, take the next ball, note the colour and then put it back, okay. So, this is called as picking and replacing.

(Refer Slide Time: 21:28)



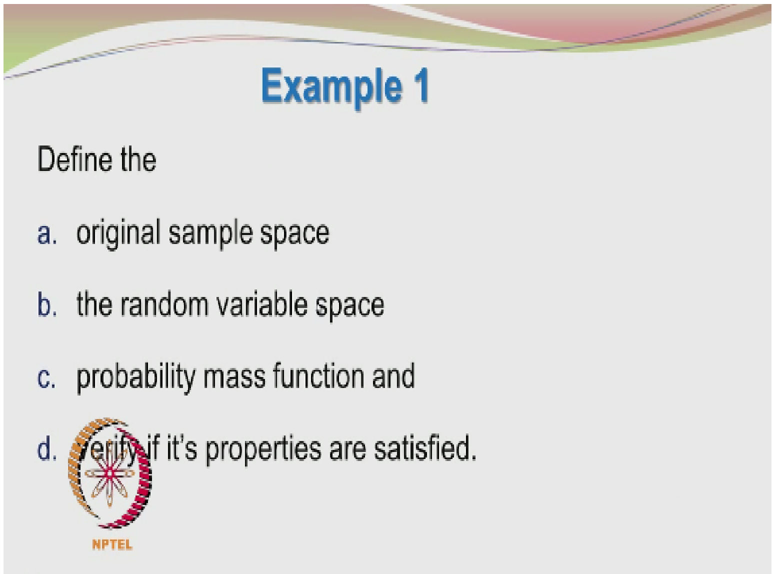
Let p denote the probability of picking up a blue ball and q the probability of picking up the red ball, okay. The questioner is you have to define the original sample space, the random variable space, the probability mass function and verify whether its properties are satisfied.

(Refer Slide Time: 21:53)



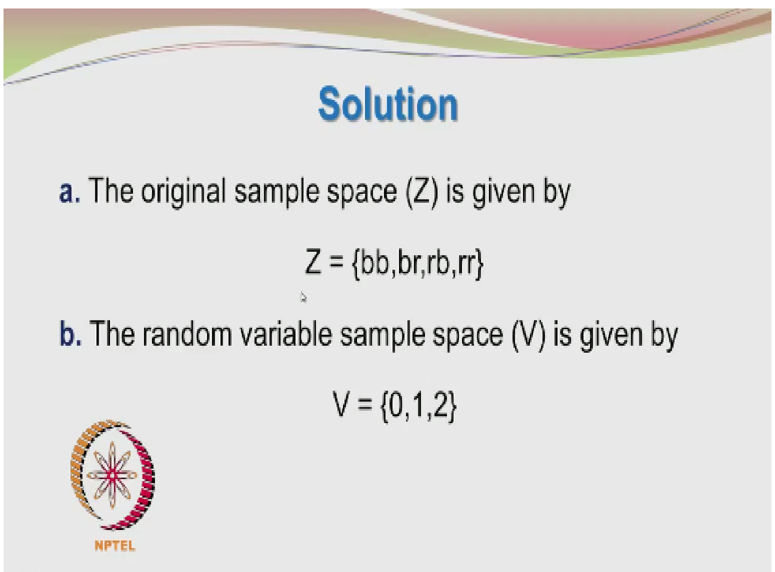
To summarize, p is the probability of picking up blue ball and q is the probability of picking the red ball.

(Refer Slide Time: 22:08)



So, we have to find the original sample space, random variable space, probability mass function and see whether the properties are satisfied.

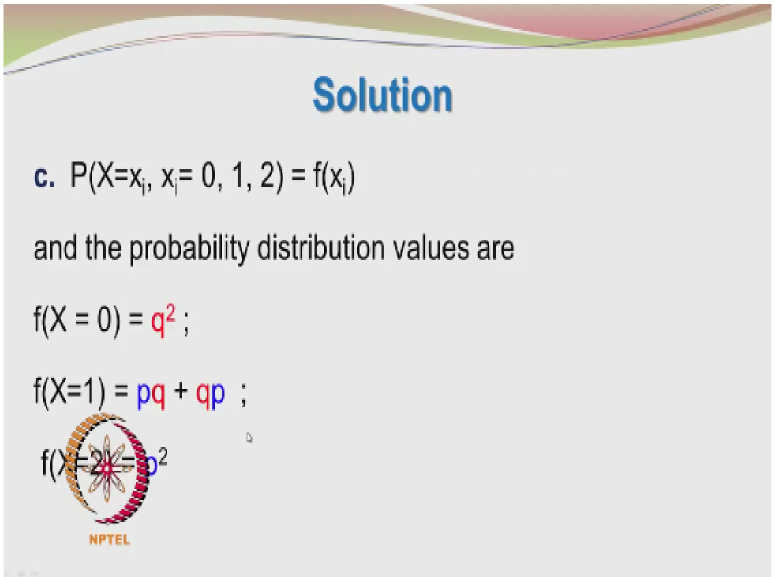
(Refer Slide Time: 22:16)



The original sample space is given by bb, br, rb, rr, okay. So, the first event is both of them are blue balls. The next event may be blue ball and red ball or red ball and blue ball. It is also possible that you pick up 2 red balls, okay. Since, the random variable X was defined as picking up the number of blue balls, okay. So, the random variable X is denoting the number of blue balls which have been picked from the random experiment, you have the possibility of 0 blue ball or one blue ball, okay, either of these 2 or 2 blue balls, okay.

So, these are the possible outcomes expressed in the form of random variables which take the values of 0, 1 and 2, okay. The possible outcomes are no blue ball, one blue ball or 2 blue balls. So, the original sample space which was having 4 entities was reduced to only 3 entities in the random variable space.

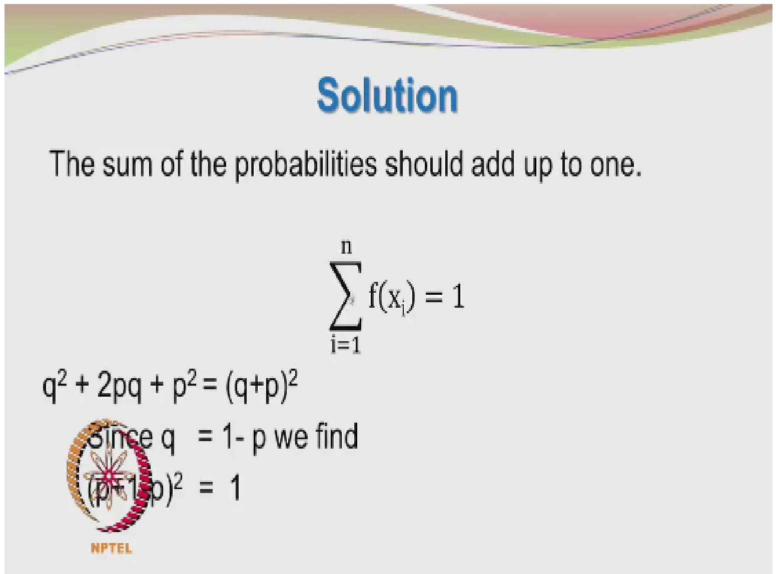
(Refer Slide Time: 23:45)



So, the probability of the random variable taking 0, okay, that means no blue balls have been picked. Then both the balls which have been picked are red in colour, okay and the probability of that happening is q square. Similarly, f of X=1, the probability is the first ball picked was blue, the second ball picked was red. So, the probability would be pq or the first ball picked was red and the second ball picked was blue, so the probability would be qp, pq+qp will be 2pq.

The probability of picking up 2 blue both balls = p square, f of x=p square, okay. Now, you can verify whether the sum of the probability is = 1. So, you add up q square+2pq+p square which is nothing but p+q whole square, okay.

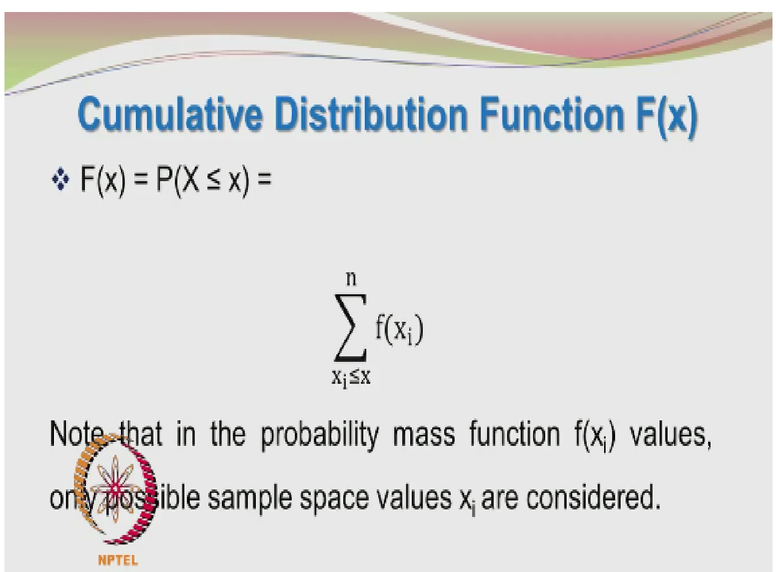
(Refer Slide Time: 25:00)



So, sigma f of xi should be = 1. So, you get when you add up all the probabilities q square+2pq+p square which is nothing but q+p whole square. Since, the probability of picking up a blue ball is 1-probability of picking up the red ball, okay or the probability of picking up a red ball is 1-probability of picking up a blue ball. We have q=1-p. You put q=1-p here, you get 1. So, you are having the sum of the probabilities totalling or adding up to 1, okay.

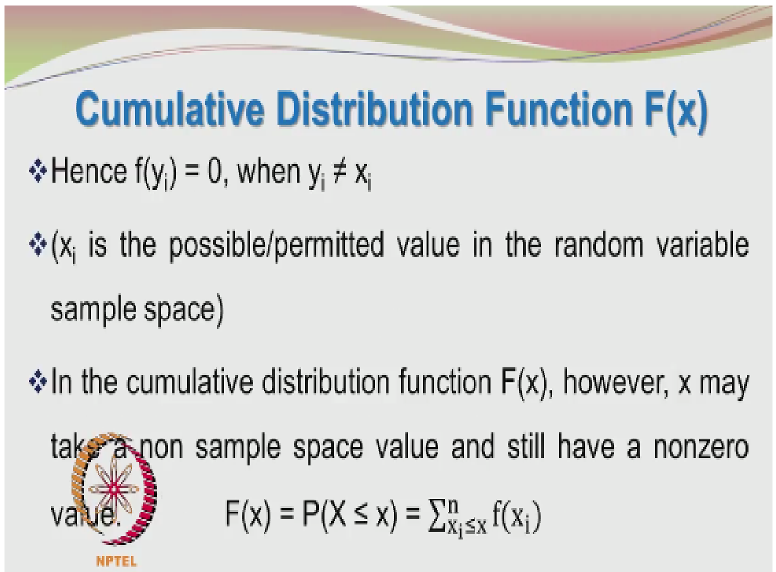
Obviously, the probability values are fractional. You have eight balls in a box, 3 of them are blue and 5 of them are red, then the probability of picking up a blue ball would be 3/8 and the probability of picking up a red ball would be 5/8, okay. So, you can see that q=1-p.

(Refer Slide Time: 26:11)



Next, we come to another interesting and important function, the cumulative distribution function F of x and that is given by F of x=probability of x<=x, okay. What is the probability that the random variable takes any value which is <= x, okay? So, that is the cumulative distribution function and that is written as sigma x1<=x up to n f of xi, okay. So, you are summing over all the values of xi until you either go below x or reach x.

(Refer Slide Time: 27:13)

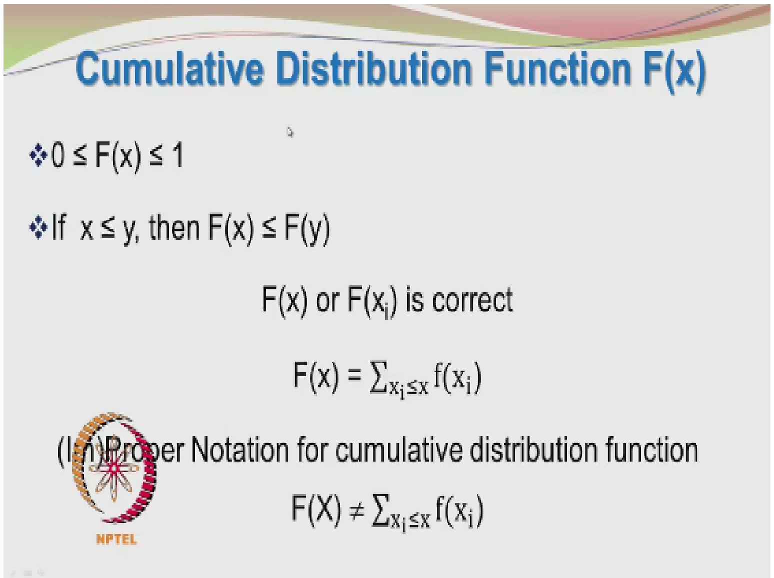


There is an interesting distinction between the probability mass function and the cumulative distribution function, okay. In the probability mass function, you can only apply the function to the possible values of the random variable x, okay. So you can only apply the probability mass function to the xi values. If for example you have another value yi which is not equal to any of the possible xi, f of yi=0, okay. But when you do the cumulative distribution function, this x value need not be xi all the time.

It may even be a value between 2 permitted xi values. For example, if the permitted xi values are values of 0, 1, 2, 3, 4 and 5. You can even find the cumulative distribution function probability of random variable x<=3.5, so that is permitted and that value need not be = 0, okay. So, you are adding up the probabilities as we saw in the mathematical equation. This is the cumulative distribution function. Cumulative means adding or totaling, okay.

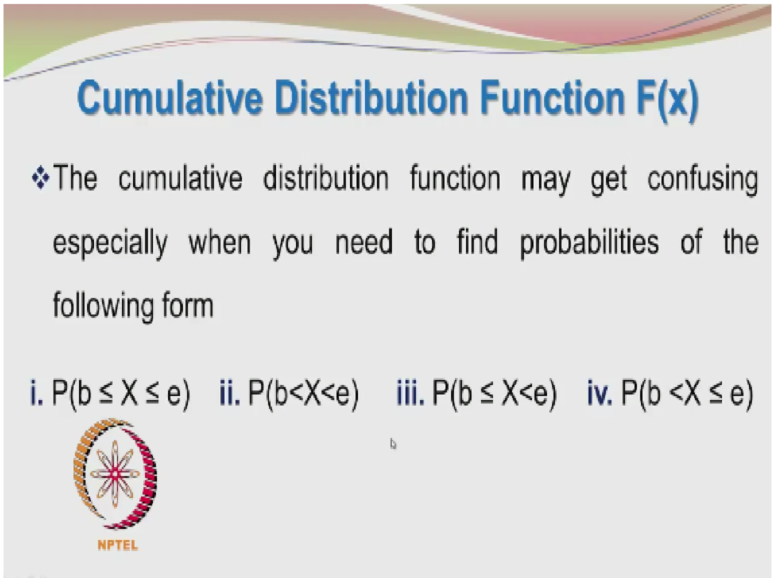
So, that is why you add up the probabilities up to the value of x, okay. That x may be equal to allowed value of xi or it may be != xi, okay.

(Refer Slide Time: 29:26)



The cumulative distribution function ranges between 0 to 1, okay. Even though you are adding, you are adding the probabilities and the maximum value would be = 1 because some of the probabilities is = 1 and if x<=y, F of x<=F of y. So, you can either write it as F of x or F of xi, okay but you cannot right F of X. Whenever you are doing the cumulative distribution function calculations, there is a possibility of committing mistake, okay.

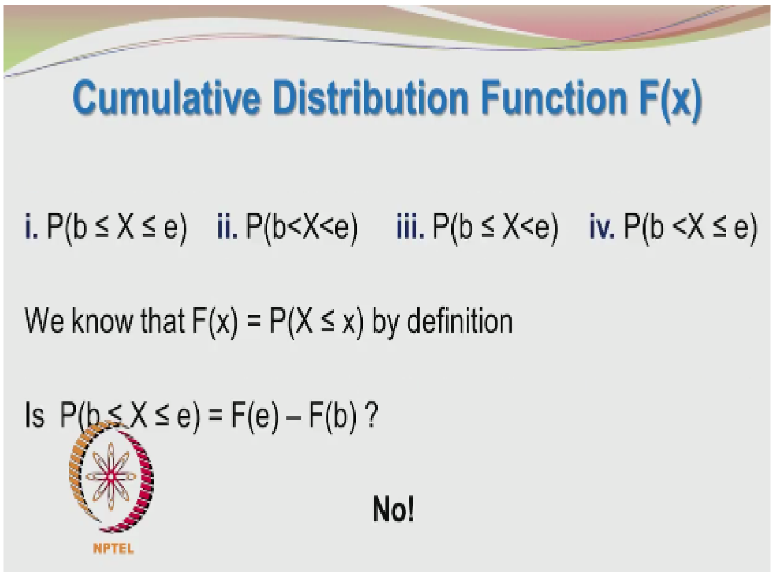
(Refer Slide Time: 30:10)



So, let us see. The probability mass function is given to you and you have to find the cumulative distribution function, okay. The statement may be similar but slightly different, small differences can be noticed. In one case, you have to find probability of b<=x<=e. In the second case, it is probability of b<x<e. In third case, probability of b<=x<e. In the 4th case, it is probability of b<x <= e, okay.

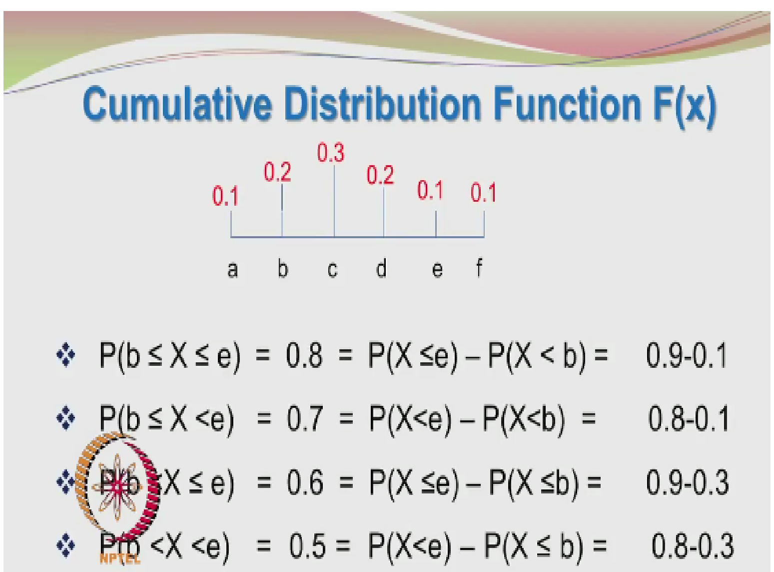
So, you may think that all of these would result in the same value, okay but that is not going to be the case as I will demonstrate with an example.

(Refer Slide Time: 31:11)



Can we write probability of b<=x<=e as F of e-F of b. No, it cannot be written in this fashion.

(Refer Slide Time: 31:30)



Let us take an example, okay. So, you are having a cumulative distribution function F of x and these are the possible outcomes a, b, c, d, e, f. These are the possible values taken by the random variable and these denote xi values. The probability of a occurring is 0.1, b occurring is 0.2, c occurring is 0.3, d is 0.2, e is 0.1 and f is 0.1, okay. These are the probability values, okay and the length of this line corresponds to the probability.

This line is twice as long as this line because the probability is 2 times higher, okay. Now, let us see the probability of the random variable x taking a value <=e and >=b. So, this is what you want to find out, okay. You want to find the probability that the random variable will take any value between b and e both b and e included. So, the probability of the random variable taking the value b is 0.2, c is 0.3, d is 0.2, e is 0.1, you add up 0.1+0.2=0.3, 0.3+0.3=0.6 and then 0.6+0.2=0.8, so that is the value you get, okay.

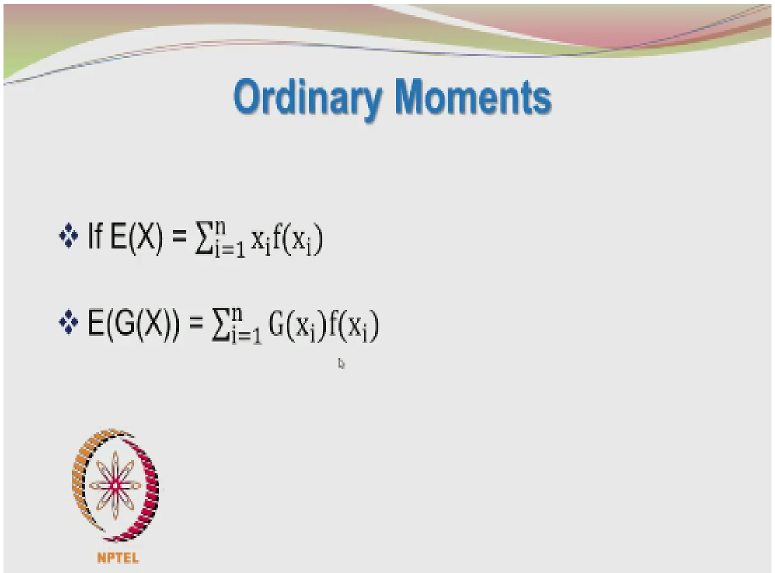
That may be written as probability of x<=e-probability of x<b, probability of x<=e is 0.9, 0.1+0.2=0.3, 0.3+0.3=0.6, 0.6+0.2=0.8, 0.8+0.1=0.9 that is why you get 0.9 here and then you get probability of x <b. Less than b is only a and the probability is only 0.1, it is point 0.9-0.1. The second case is probability of b<=x<e, okay and that is = 0.7.

Here b is included but e is not included, okay. So, it should be < e. So, we count 0.2 corresponding to probability of occurrence for b. So, 0.2+0.3=0.5, 0.5+0.2=0.7, okay. We are excluding e because it is probability of x< e-probability of x<b, okay. Probability of x<e is 0.8, 0.1+0.2=0.3, 0.3+0.3=0.6, 0.6+0.2=0.8 - probability of x<b, okay.

Probability of x<b as we saw in the previous case is 0.1. So, you get 0.8-0.1 which is 0.7. Similarly, you can easily show that probability of b<x<=e is 0.6 and probability of b<x<e=0.5, okay. So, coming back to our original problem statement, what is the probability of b<=x<=e that was != F of e - F of b, but it is probability of x<=e-probably of x<b, okay and that was 0.9-0.1 which was = 0.8, okay.

So, do not try out some formulae based on intuition, okay in these kind of situations, okay. Rather than using these formulae, I would rather advise you to construct the distribution of probabilities and then do the calculations. Anyway the mathematical formulae are here but from my point of view, these formulae need not be memorised but they can be easily implemented, okay, by just using the diagram.

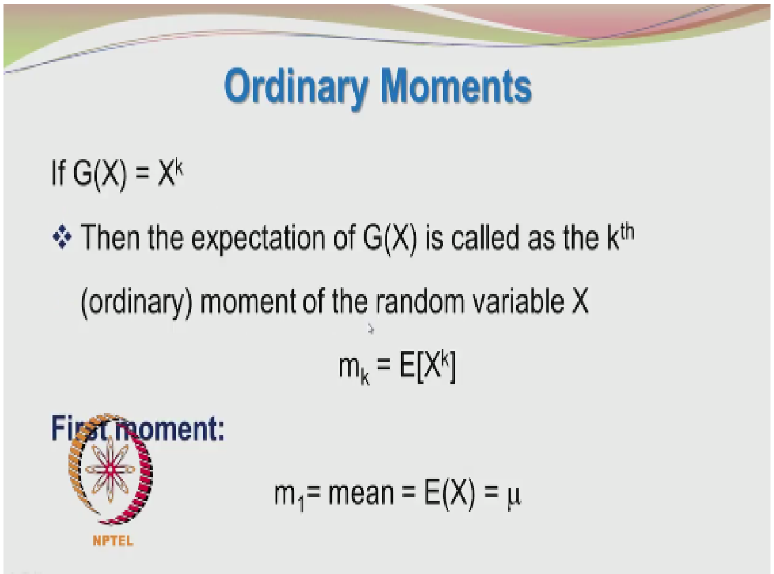
(Refer Slide Time: 36:34)



Now, we are going to look at something known as moment, okay. In physics, you might have seen the term moment come into use rather frequently, okay, force*distance, okay. So, here also we have ordinary moments and central moments. So, I will just give the definitions here. So, the expected value of E of X is given by i=1 to n xi f of xi.

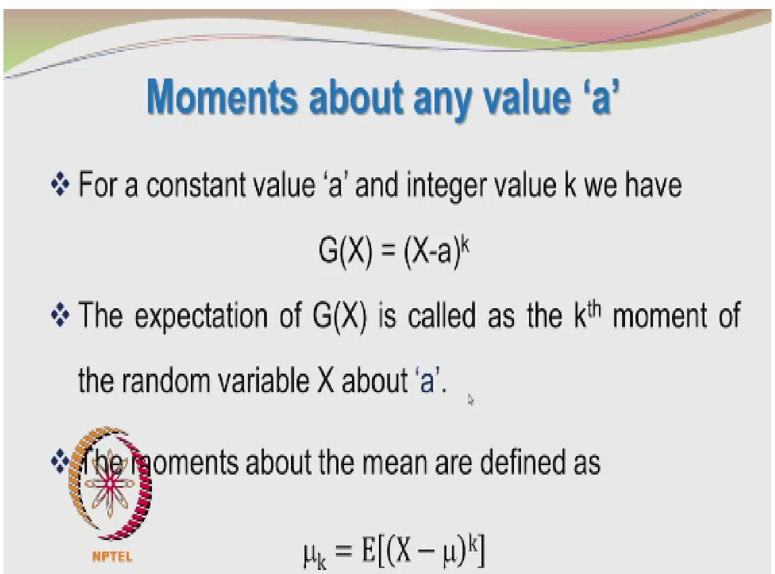
This is called as the first ordinary moment. In general, expected value of G of X=sigma i=1 to n G of xi*f of xi, okay. Here, you have random variable x, so you put xi here. Here you have G of X and so you put G of xi here. The function defined by the random variable is implemented on xi inside the submission, okay.

(Refer Slide Time: 37:53)



So, if we define G of X as X to the power of k, then the expectation of G of X is called as the kth ordinary moment of the random variable X. So, the kth ordinary moment is represented by m of k is equal to expected value of X of k. The first ordinary moment is represented by m1 which is equal to mean and that is equal to expected value of X because k is 1, so you have expected value of X and that is = mu.

(Refer Slide Time: 38:30)

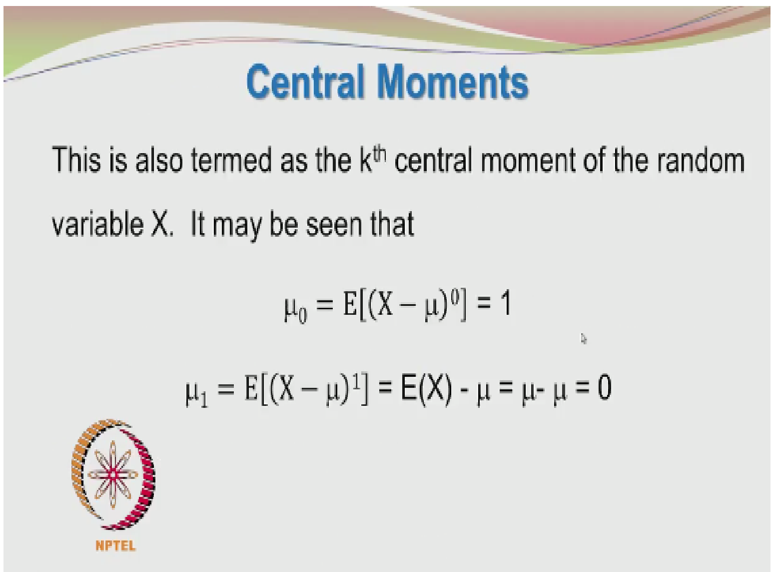


So far, we have been defining the function G of X as X power k, but we can also define the random variable X as a deviation from a where a is constant value, okay. Earlier, we had by default used a as 0 but it need not be the case always. So, we want to put a as constant value and we are going to find the kth moment, okay. Remember k should be an integer value. So, G of

X=X-a to the power of k. The expectation of G of X in such a case is called as the kth moment of the random variable X about a, okay.

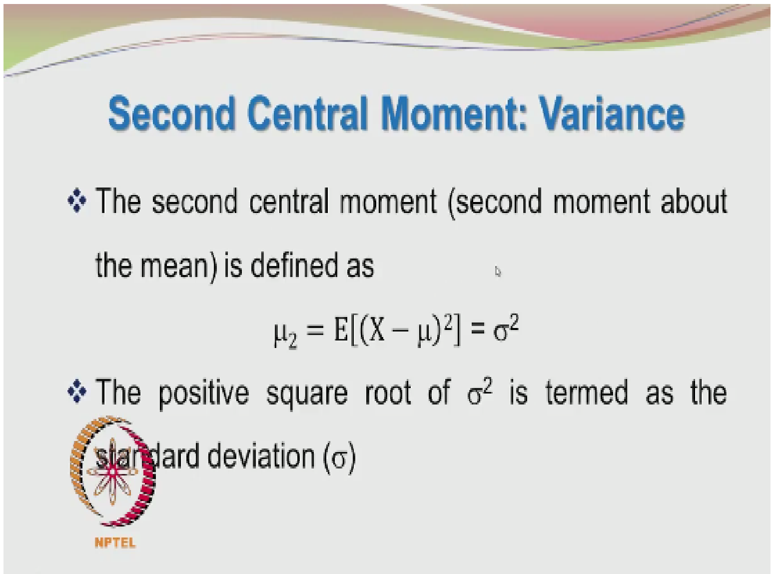
The kth moment of the random variable X about a. The moments about the mean are defined as mu of K = expected value of X - mu to the power of k. So, k can be 0, k can be 1, 2, and so on. Remember k should be an integer, okay. So, here we are finding the moment about the mean, okay. The mean is the centre point of the distribution. So, it is X-mu, expected value of X-mu to the power of k is called as mu k and these are referred to as the moments about the mean for different values of k.

(Refer Slide Time: 40:23)



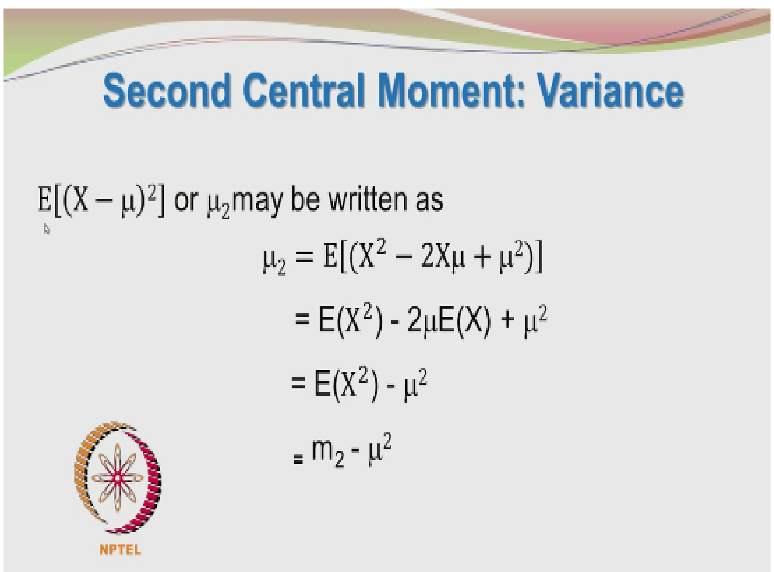
It is also termed as the kth central moment of the random variable X. Since, we are taking the moment about the mean, we refer to it as the kth central moment of the random variable X. So, when you put k=0, mu 0 is E of X-mu to the power of 0, expected value of 1 is 1. Mu 1 where we put k=1. So, expected value of E of X-mu to the power of 1 is expected value of X-mu=mu-mu=0, okay. So, this is the expected value of X-mu, okay that we call as mu1 which happens to be 0. So, mu0 is 1 and mu1 is 0, okay.

(Refer Slide Time: 41:24)



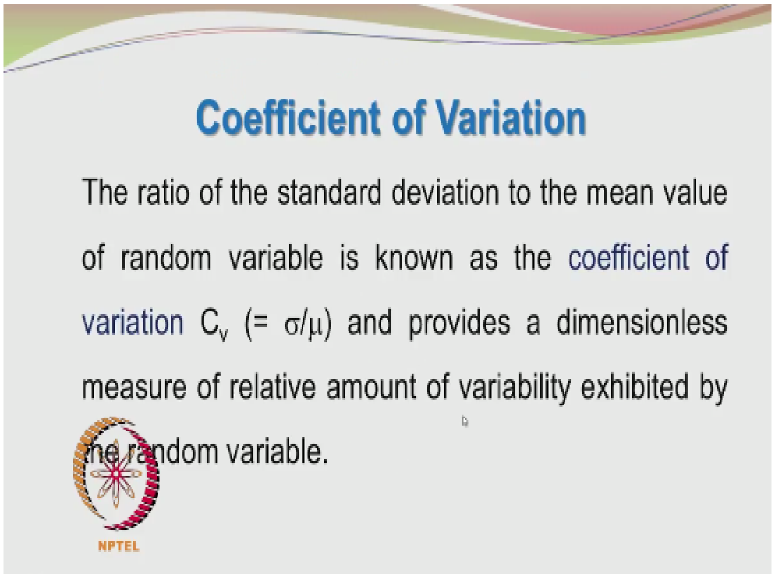
The second central moment is termed as the variance. The central moment is taken with respect to the mean, so k=2. The expected value of X-mu whole square=sigma square. This definition we had seen previously but now this is called as the second central moment or the second moment about the mean. The square root of thee variance is called as the standard deviation.

(Refer Slide Time: 42:00)



So, E of X-mu whole square or mu2 may be written as m2-mu square, second ordinary moment - mu square, okay.

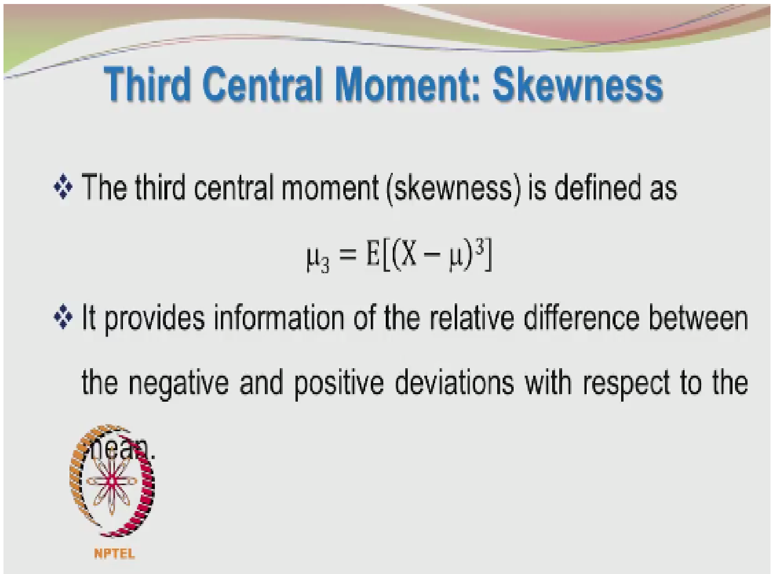
(Refer Slide Time: 42:22)



So, you have found the mean, variance and the standard deviation, so you have to scale it properly, okay. So, you want to see what percentage of the mean or what fraction of the mean is the standard deviation and so you write the coefficient of variation Cv as sigma/mu. It provides a dimensionless measure of the relative amount of variability exhibited by the random variable, okay.

If sigma is very high, sigma is 100, okay, you may think that there is a lot of variation, the standard deviation is pretty high but you have to find it with respect to the mean value. If the mean is in the order of 200 and the standard deviation is 100, 100/200 is pretty high, okay. But if the mean value mu is in the order of 100/10,000 okay which is 1/100 and that would be 0.01 which is pretty low, okay. So, you have to compare the standard deviation with respect to the mean value.

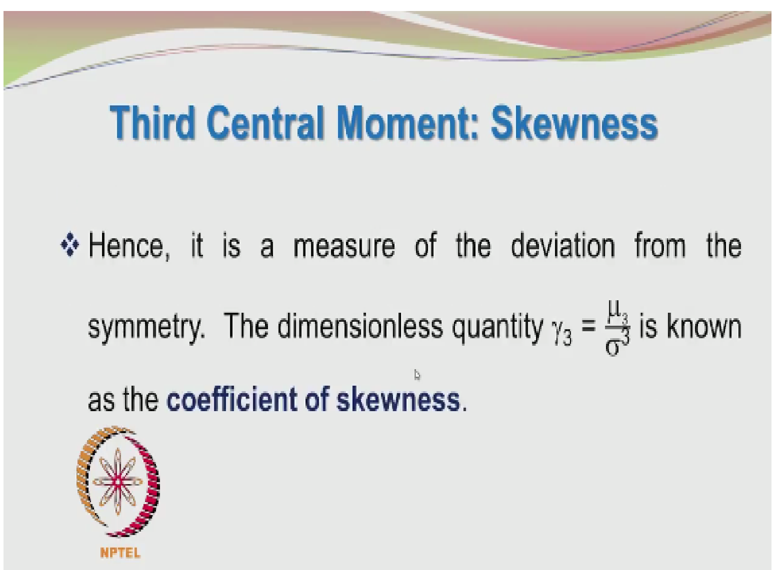
(Refer Slide Time: 43:52)



The third central moment or the moment about the mean mu is defined as skewness, okay and that is given by mu3=expected value of X-mu whole cube, okay. The significance of the skewness is it illustrates or indicates the asymmetry of the distribution, okay. You can understand that when there are several values of X, they may be distributed about a mean value, okay and they may not be distributed uniformly with respect to the mean value, so you may have negative deviations and the positive deviations.

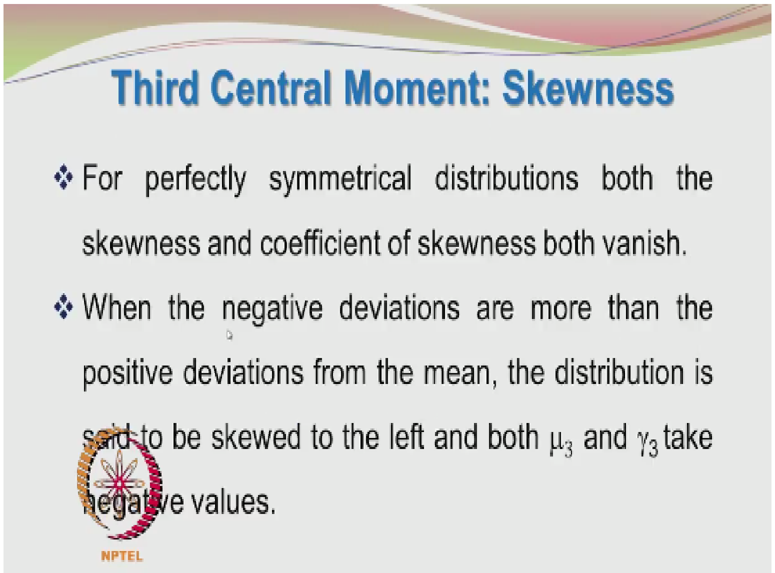
The relative difference between the negative and positive deviations with respect to the mean is given by the third central moment.

(Refer Slide Time: 45:01)



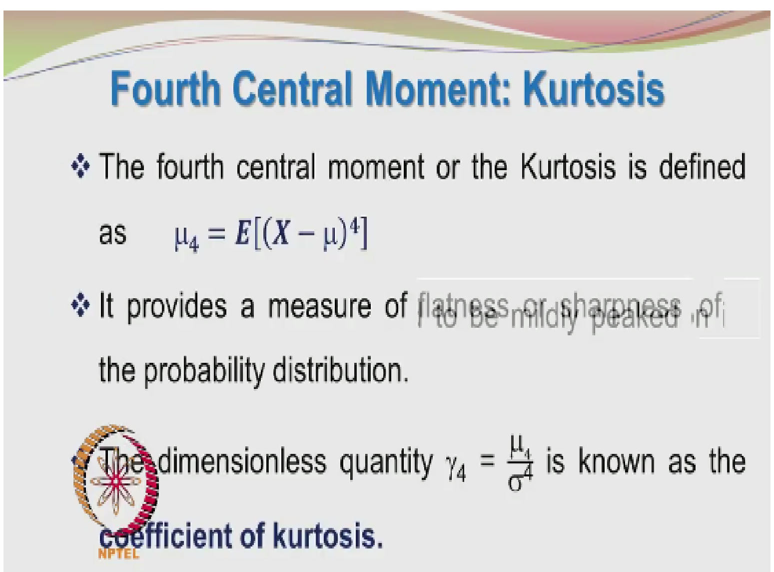
So, the third central moment skewness is a measure of the deviation from the symmetry, okay. The dimensionless quantity mu3/sigma cube is known as the coefficient of skewness.

(Refer Slide Time: 45:18)



If the distribution is perfectly symmetric, both the skewness and coefficient of skewness both vanish. When the distributions are such that the negative deviations are more than the positive deviations from the mean, the distribution is said to be skewed to the left, okay. There is dominance of negative deviations when compared to the positive deviations, so the distribution is said to be skewed to the left and both mu3 and gamma 3 take negative values.

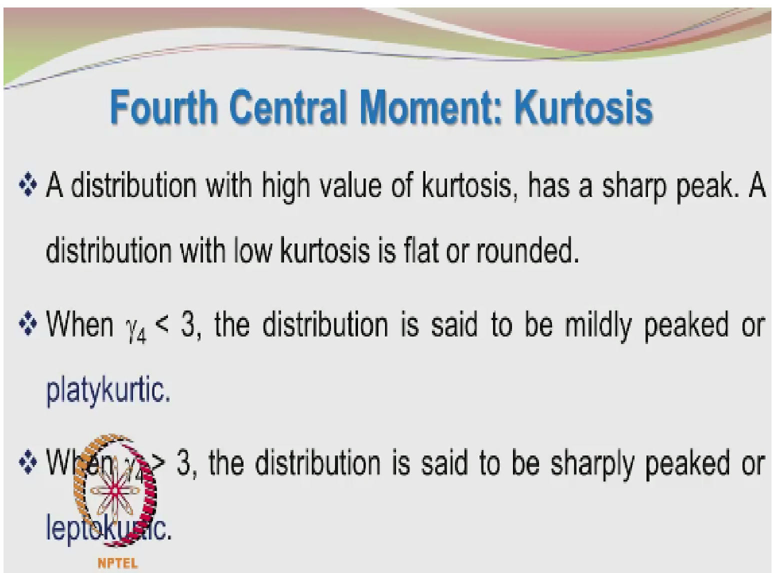
(Refer Slide Time: 45:57)



The 4th central moment is called as the Kurtosis and that is defined as mu4=expected value of X-mu to the power of 4, okay. It is the measure of the flatness or the sharpness of the probability distribution, okay. The ratio of the Kurtosis to the 4th power of standard deviation is termed as the coefficient of Kurtosis, okay. A distribution with high value of Kurtosis has a sharp peak. On the other hand, a distribution with low value of Kurtosis is flat or rounded.

When the gamma 4 value is < 3, the distribution is said to be mildly peaked or platykurtic, okay. When gamma 4>3, the distribution is said to be sharply peaked or leptokurtic. These terms you may encounter in standard texts or even in research papers showing distributions and explaining their characteristic features. You may often come across these terms platykurtic and leptokurtic. So, depending upon whether the peaks are sharp or the distribution is flat and rounded, the appropriate terminology is used.

(Refer Slide Time: 47:25)



The next topic in our discussion is median for discrete probability distributions. The median xm for a discrete probability distribution is the point within the range of the allowed values of the random variable X such that the cumulative distribution value at xm is exactly 0.5. In mathematical form, sigma i=1 to m f of xi=the cumulative distribution value at xm=0.5. F of xi is the probability mass function.

The xm value here is such that the probability of X < xm = probability of random variable X>xm are both = 0.5. This concludes our brief discussion on the discrete probability distribution functions. We will be now moving on to continuous probability distributions. They are also called as the probability density functions.