Statistical Learning for Reliability Analysis
Prof. Monalisa Sarma
Subir Chowdhury School of Quality and Reliability
Indian Institute of Technology - Kharagpur

Lecture - 12
Continuous Probability Distributions (Part 1)

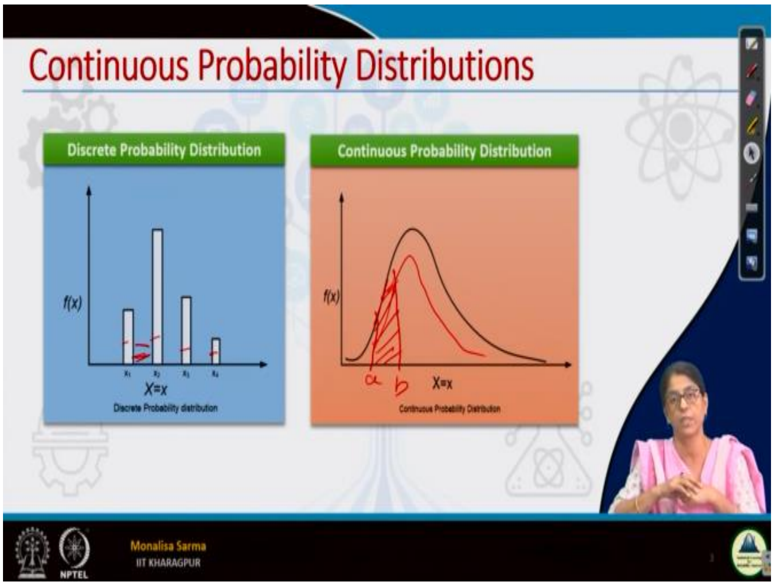
Hello everyone, so in continuation of our discussion of probability distribution, today, we will be discussing continuous probability distribution. Last few classes, we have discussed discrete probability distribution, as well as we have done some tutorials also on the discrete probability distribution. Now, we will be discussing continuous probability distribution, this continuous probability distribution, I will be taking 2 lectures for it.

In the first lecture, we will be covering around 3 continuous probability distribution, and other 3 I will be covering in the next lecture, followed by a tutorial class.

(Refer Slide Time: 01:01)

So in this lecture, I will be covering uniform distribution, normal distribution, and standard normal distribution.

(Refer Slide Time: 01:07)



Now, before covering going to uniform distribution, let us first take a quick recap of what we have learned of discrete probability distribution, the left side figure so the discrete probability distribution remember how like, we can express a discrete probability distribution in the form of a tabular form, in the form of a graph, as well as in the form of expression, the figure here shows that we have expressed the discrete probably distribution in the form of a figure.

So now, can you remember what figure is this? It is a bar chart so where, why how I can distinguish that it is a bar chart, not an histogram, because in a histogram, there will not be any gap between the this rectangles, let me take the pen, and then I will be showing you. So can you see this gap? This gap in a histogram, this gap will not be there basically, they will not be this gaps, it will be just close together.

Because the width of this rectangle in a histogram, it refers to the in the range of the values. So we cannot put any gap in between 2 rectangles, so this directly shows that is a bar chart. Anyway, that is not a main concern, main concern is here and trying just trying to say you how we basically can have a figure of a discrete probability distribution. Similarly, for continuous probability distribution, the figure is in a continuous curve.

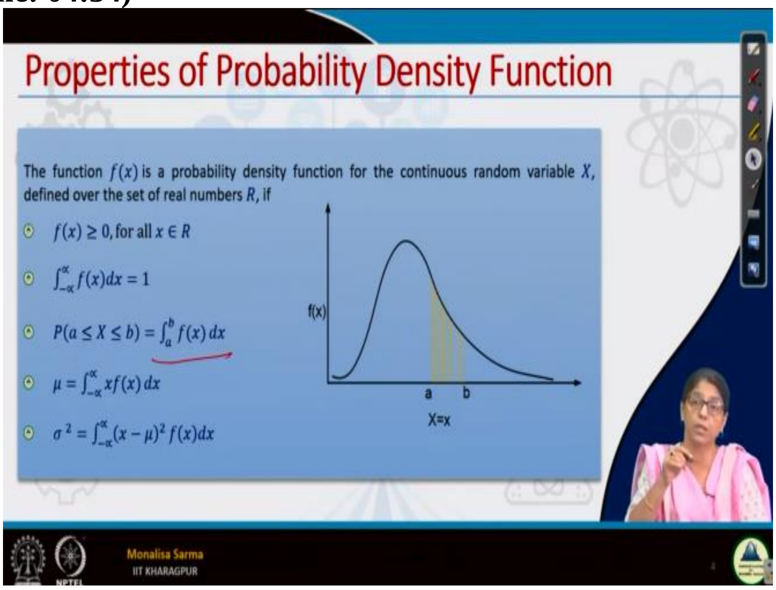
So here we see this is a continuous curve. So in this continuous probability distribution, the figure is basically a continuous curve. So now, the thing is that first, again, I will just, we will just go back a bit and we will see what was f(x) in case of a discrete probability distribution, what was f(x)? f(x) is the probability of the random variable X, the probability that X takes that a random variable X takes that is f(x), so value of f(x) cannot be greater than 1, it cannot be less than 0. So f(x) has to be between 0 and 1.

That is the case in case of discrete probability distribution. But in case of continuous probability distribution, I have already mentioned if you can remember, of course, f(x) cannot be less than 0, it has to be greater or equal to 0, but f(x) can be greater than 1 also, because here f(x) does not denote the probability it is here f(x) is a density function how µch the probability concentration is there in it within a certain interval.

When you want to find out the probability in for continuous, we will have to find out the area within that interval. Suppose I am considering this area suppose this area, so this area, I will have to find out the area of this area, I have to find out the area of this region. Suppose this is a this is b, I will have to find out the area of this region to find out the probability at this point. Sorry, it cannot be found when speaking of continuous probability distribution, it will never be at a single point.

I already mentioned that probability or a single point is equals to 0, it is almost equals to 0. So when we talk about probability and continuous domain, it has to be within an interval, maybe a very small interval, maybe a bigger interval. And, of course, for a random variable for the function to satisfy a continuous probability distribution rather, I should say function to satisfy a probability density function, we can say it is a probability density function, when it satisfies 2 of the properties if you remember.

(Refer Slide Time: 04:34)

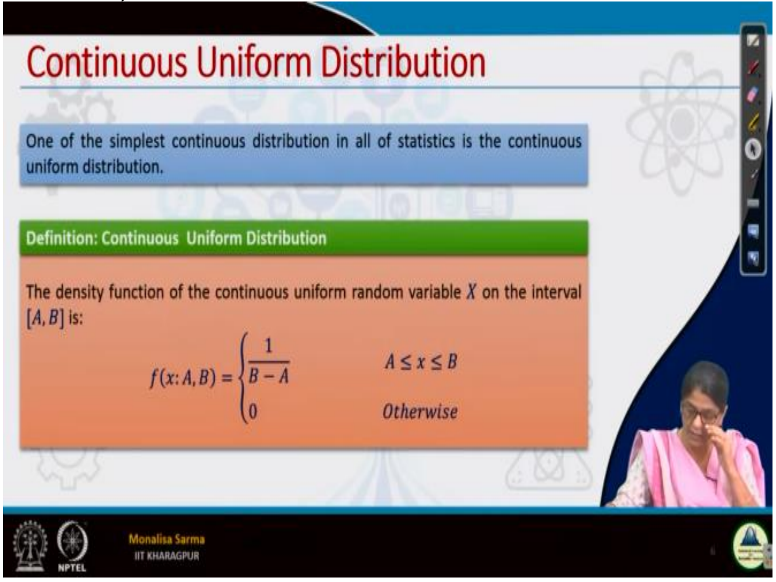


I am just putting it here again, it will it is called the probability density function. The first condition is that it f(x) will be greater than equal to 0 and integration of f(x) from all the range from - ∞ to + ∞ it should be equal to 1 these are the 2 conditions. So, now if I want to find out the probability between 2 values that a and b in the figure it is given this a and b.

So how we can find out it is just the integration of f(x) within the value a to b that is how we do in continuous random variables when the variable is continuous. So, for finding out the μ and that is the mean and the variance mean or I should say the expected value on the variance, that is same as what we have done for discrete probability distribution, but instead of summation here they are, we have used summation here we will be using ∫μf(x) ∫ x f(x) and for σ this $\int (x - \mu)^2 f(x)$.

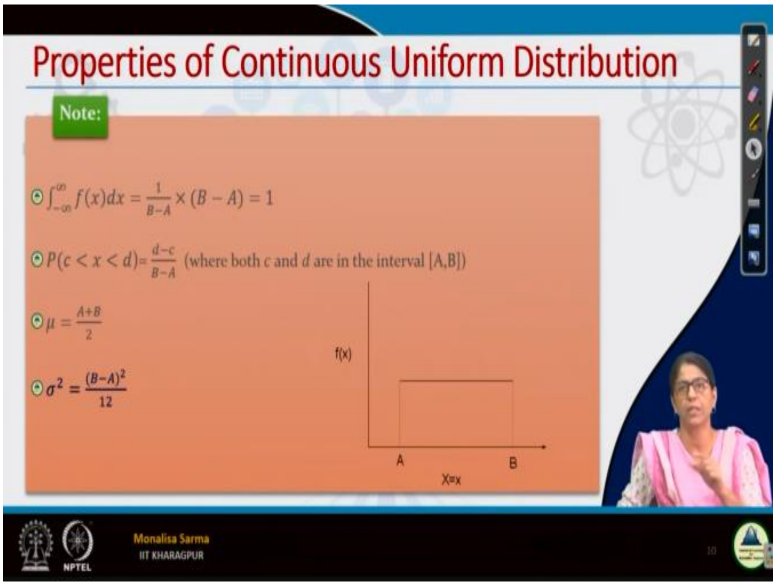
It is the same just summation get replaced by integration because here we are concerned of the forμla, the area within that range. So, we will be basically integrating it.

(Refer Slide Time: 05:39)



So, one of the simplest continuous probability distribution is the uniform distribution. So, we talk about uniform distribution, where the probability at each point is same.

(Refer Slide Time: 05:55)



So, basically, if I can tell you about the graph of a probability continuous uniform distribution that will be this sort of a graph here. So, we are trying to find out within the range the A to B, so, this is the graph here. So, it is equal probable in all the areas and of course, since it is a probability distribution, so, some ∫f(x) would be equals to 1. So, if ∫ f(x) = 1 so, what will be the probability of this probability of any value means probability in this range what is the probability it will be 1 / B - A is not it?

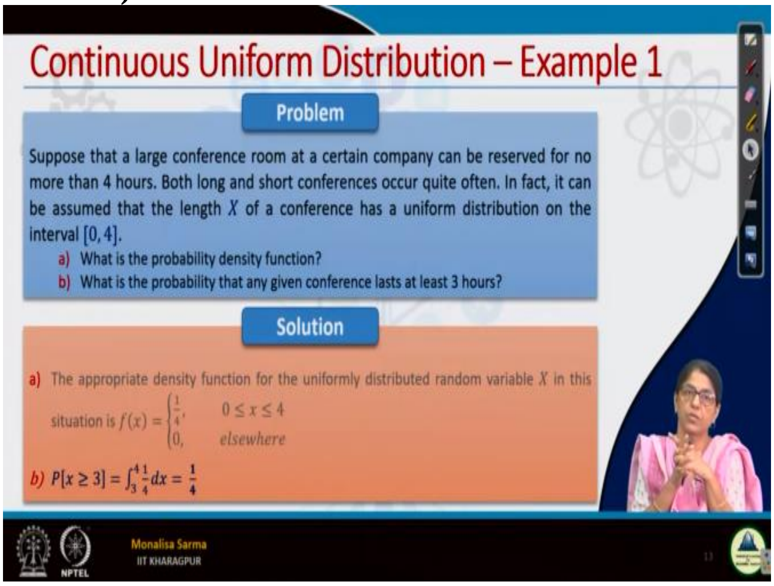
So, probability will so my f(x) will be as what it is given here f(x) will be 1 / B - A directly we can get ε since the area if we find out the area of this whole rectangle area of the rectangle is what? Length x breadth if now my this is my I have to find out my probability in this area this whole area has to be 1. So, what it will be to when will I get 1 because my total length is B - A that means, to get it one my I have to µltiply it 1 by B - A that directly gives me the probability is not it?

That I should not say probability that directly give me the density function that is f(x) it is not probability mind it that is the f(x) so, my f(x) is 1 / B - A, 1 / B - A from the range X ranging from A to B and in other areas it is 0. So, now, if we are interested in finding out any value within this range suppose we are interested in finding out what is the probability in this range c to d. So, what will be the probability in this range c to d what is my length of this area?

Length of these areas this d - c. So, d - c into what is my probability density function density function is 1 / B - A. So my probability become d - c / B - A. So, that is the probability within the range c to d probability always in continuous only distribution and I am repeating it again when we talk about continuous probability distribution, the probability is always within an interval we do not talk of a single point.

So, here by using the forμla same forμla integration of x f(x) for expected value, we will be getting A + B / 2 for σ^2 we will be getting B - A / 2 where B and A are the range of the random variable where X is a random variable X random variable can take value within the range of A to B.

(Refer Slide Time: 08:36)



Now, one example to illustrate our continuous uniform distribution remember the uniform distribution we had it for discrete also similar, we have in uniform distribution both for discrete and continuous. Now, coming to this example, see this example if you go to the example you will find it is very easily what it is given. Suppose that a large conference room at a certain company can be reserved for no more than 4 hours.

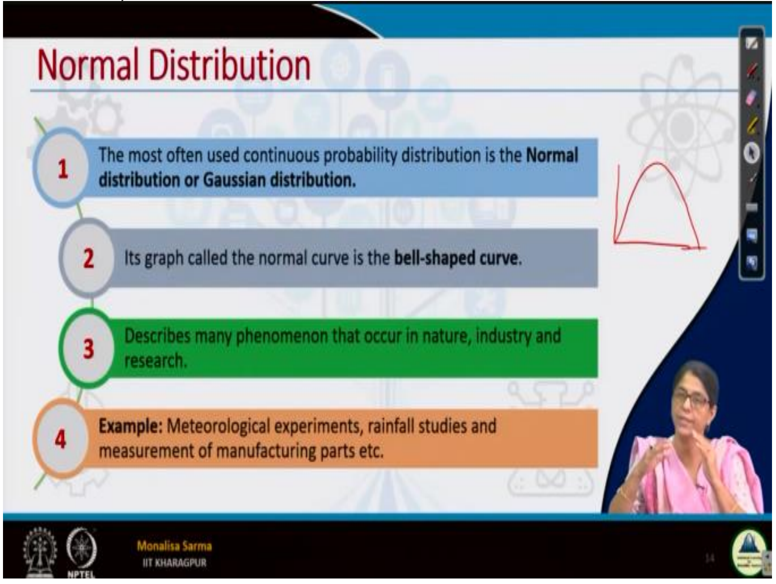
The large conference room we cannot book it for more than 4 hours. That means my value of my random variable, it ranges from maxiμm 0 to 4 booking cannot be negative, though, when it our continuous probability distribution. So, it can be - ∞ to + ∞ but here booking cannot be negative so, it will be from 0 to 4 rest every all other places have a random variable the probability of the density function will take the value 0.

So, then both long and short conference can occur quite often. In fact, it can be assumed that the length X of a conference has any form distribution on the interval 0 to 4. It is specified that the length X where X is a random variable, X is a uniform distribution in the range of 0 to 4. If the X has a uniform distribution in the range 0 to 4. Then what is the probability density function that means what is my f(x)? My f(x) = 1 / 4 is not it?

1 by 4 - 0, I have to find it from this range, this is 0, this is 4, this is my range. So my probability density function will be 1 / 4 - 0 that is 1 / 4. So that is my f(x), so first question what is the probability density function that is all it is 1 by 4 from which range from 0 to 4 and rest are all other area, what is the range? It is 0, what is the probability density function other than this range, that is from 0 to 4, it is 0.

That is how we define the density function. See here, the density function it is 1 by 4 from 0 is elsewhere from x ranging from 0 to 4 and 0, it is elsewhere. Now, what is the next question? What is the probability that any given conference last at least 3 hours? So it is given that any given conference last at least 3 hours so that means my random variable will go from 3 and above? So I will integrate it from 3 to 4 so that is a p(x) >= 3 at simple integration from 3 to 4 1 before dx, that is all that is very easy problem.

(Refer Slide Time: 11:25)



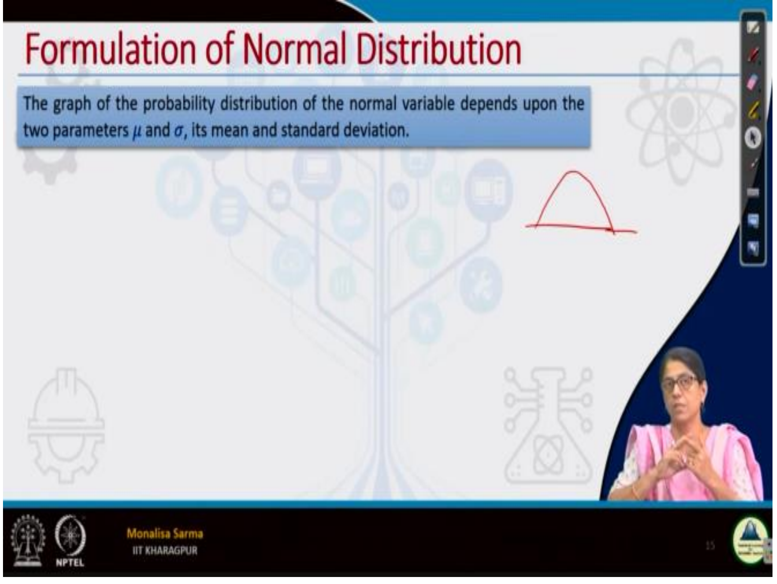
Now, we will discuss the next important distribution that is normal distribution is also Gaussian distribution this is a very important distribution in the sense that this distribution is it has wide application you can see this application of this in nature you can see application of design industry, research, business many places, you can see this application of this normal distribution this distribution was discovered by one professor,

Professor Carl Gauss, that is why by his name it is this distribution is also called Gaussian distribution and this distribution it has a basically a bell shaped. So, this is called a bell shaped curve, and describes any phenomenon on that occurs in nature industry and research. Some of the example is metrological example rainfall studies, measurements of manufacturing parts, etc.

Let me tell you, what made this professor Gauss to discover this distribution actually, he was trying to measure the movement of celestial body more he was trying to measure a moment of celestial body again and again of one celestial body, he was trying to measure and its moment on each measurement, he was getting different, different value. So, when he was trying to see there is an error in each measurement.

He saw that it fits into a distribution that distribution that he has come up with this distribution based on his error and measurement that is this bell shaped distribution that is a Gaussian distribution.

(Refer Slide Time: 13:08)



So, as I when I was discussing discrete probability distribution, I have already specified that when we talk about distribution, there are a few parameters which specify a distribution likewise, uniform distribution also we saw the parameters that describe the uniform distributions are μ and variants μ and σ. So, σ if I tell σ is a standard deviation like.

Similarly, for normal distributions, the parameters that are μ and σ here also when μ is the mean, and σ is the standard deviation, let me tell it in a different way actually, when we talk about distribution and distribution are characterized by a few parameters, which are called maybe scale parameter, shape parameter, location parameter. So, instead of telling as μ or standard deviation, that a parameter is defined by scale, what is the scale of the distribution?

So, then parameter is defined by the location where it is centered? Basically the expected value the μ, where it is centered, that is the location then we have called something called a shape parameter, what is the shape of the distribution? Shape of the distribution like as I talked about discrete probability distribution, I told you we are not very μch concerned about the state of the distribution while talking about discrete probability distribution.

Actually, we do not bother that μch of what to say. But in case of continuous probability distribution, when we are talking about the distribution of a continuous random variable, shape plays an important role. So, there is a shape parameter as well. Now, this parameters, these are also called moment there, I have not kept it in a slide because this is you do not need to consider as a part of this course.

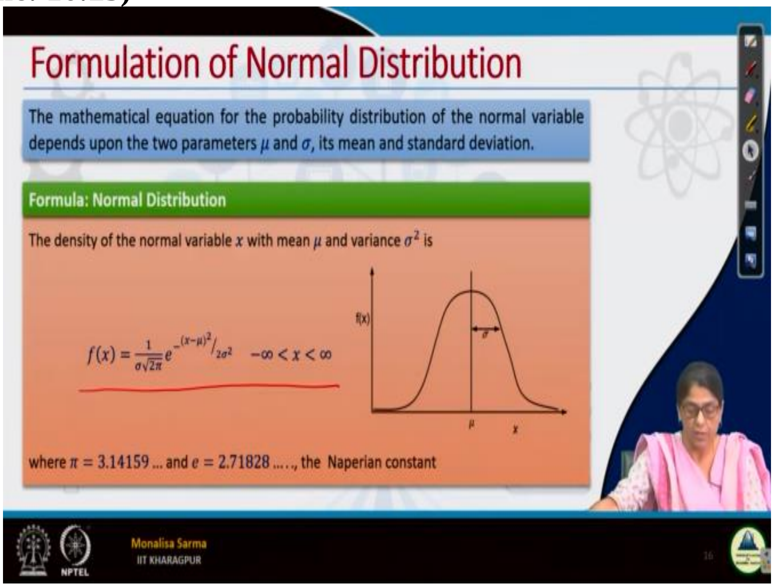
But for your knowledge for your information, I am just telling this we can also call the moments there are different moments which characterize the distribution, the first moment is the expected value that is the mean that is the location. Then the second moment is the scale parameter that is the standard deviation that is the spread how spread out my distribution is what the variants gives how spread out is.

So, that is the second moment then third moment is symmetric, how symmetrical is my distribution, whether it is skewed towards left it is skewed towards right or it is centered it is symmetrical to the mean. So, that is the third, what to say third moment, then fourth moment is the kurtosis, kurtosis gave the peakedness of the distribution and that is the height of the distribution.

So, these are the there are many other moments over these 4 moments are very necessary to describe characterize the distribution, but we will not go on discuss the moments here, but we will just stick to the parameters which are necessary to describe a distribution. So, talking of normal distributions, the parameters that describe a normal distribution basically, it is the mean and a σ mean is the location parameter and σ is the scale parameter here we do not have shape parameter why?

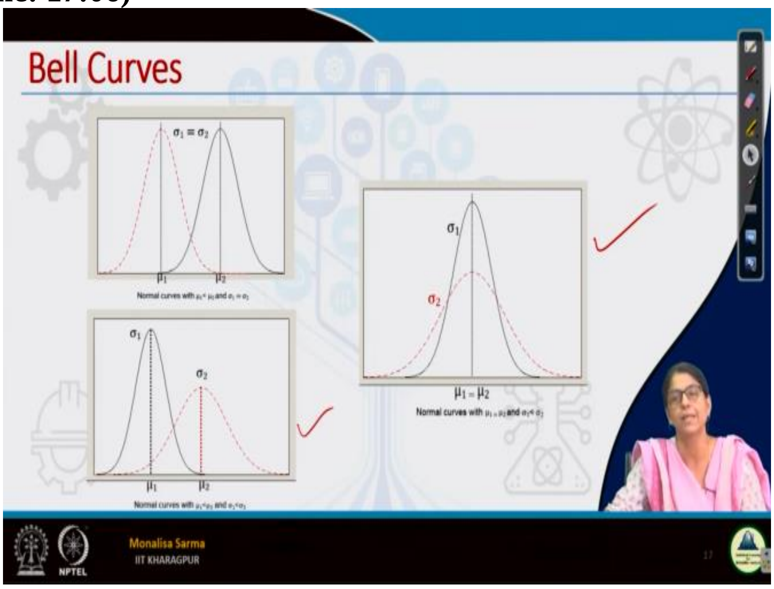
We do not have shape parameter because when we talk about normal distribution, normal distribution just has 1 shape that is the bell shaped just this one same. So, we do not have a special parameter to differentiate that different shapes geometric since normal distribution has this one set no point. So, normal distribution has this 2 parameter that is μ and σ it is also called the location parameter as well as the scale parameter.

(Refer Slide Time: 16:23)



Now, this is the probability density function for the normal random variable if x is a normal random variable. So, this is the density function like for uniform distribution, what is the density function? Density function is the f (x) = 1 / B - A where x range from B to A here x ranging from minus infinity to plus infinity my f (x) is this particular expression. Now, I will come to this expression again.

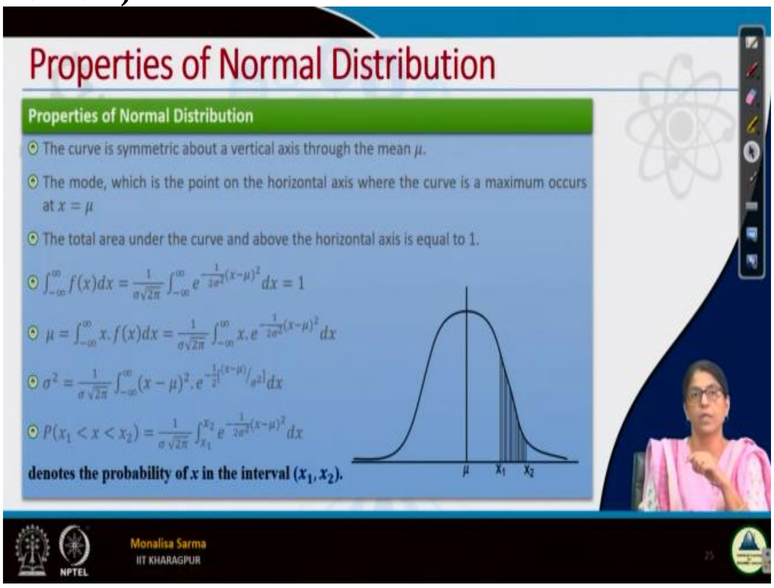
(Refer Slide Time: 17:06)



But before that, let us see some shapes of the normal distribution like here you can see different bell curve first one you see where it my σ 1 = σ 2 both you have seen here we have seen 2 bell curve where σ is same for both that case, but that means there my spread is same, but the location is different one location is μ 1 the another location is μ 2 you see how it is, then you see this one here the location is same that is μ1 = μ 2.

But both the scale is different for where here what is σ 1 is smaller than σ 2, σ 2 is more that is it is more spread out. Similarly, here we see the third one where μ 1 is less than μ 2 and σ 1 is less than σ 2 this this type of figure we basically get for different values of μ and σ.

(Refer Slide Time: 18:00)



So, now, for the normal distribution what the properties are the curve is symmetric about the vertical axis through the mean μ the curve is symmetric to the mean μ whatever may be the mean if this is my mean suppose, this is my y axis suppose this is my mean my curve is symmetric towards the mean and we also know what is mode see the mode is the point which is the point in the horizontal axis where the curve is maxiμm in a normal distribution curve as maxiμm mean at the expected value at μ.

So, mode is also μ here and at normal distribution my median is also μ here means what is median? At median 50% of the world's population falls in the left of the median with under 50% falls to the right of the median. So, here a normal distribution μ is also the median also in a normal distribution mean is equals to mode is equals to mean this is a very important characteristics please remember.

The total area under the curve of course, it has to be 1 in any continuous probability distribution we have seen f(x) gives the shape of the curve and the total area under the curve is equals to 1 that is why it we integrate f(x) from integration from minus infinity to plus infinity we get 1. So, this is the integration we get this is this whole is equals to 1. So, now what is μ? μ is integration x f(x) and σ^2 is integration $(x - \mu)^2$ and f(x).

Just put it into value and then we have to simplify. So, if we are interested in finding out what is the probability in this range x 1 to x 2 it is given here. If you want to find out what is the probability within this interval x 1 to x 2, just we have to integrate the f(x) value, whatever f(x) expression we have, we have to integrate f(x) from x 1 to x 2. Now see the problem here, when we are interested in finding the probability of this range x 1 to x 2.

Maybe any other range whatever range maybe we find you are interested in finding out the probability, then see, we will have to evaluate this expression, for this particular range evaluating this expression is not a very easy task we will have to do this is a complex integration solving this complex integration again and again for when trying to find out a probability of any interval it is a tedious job.

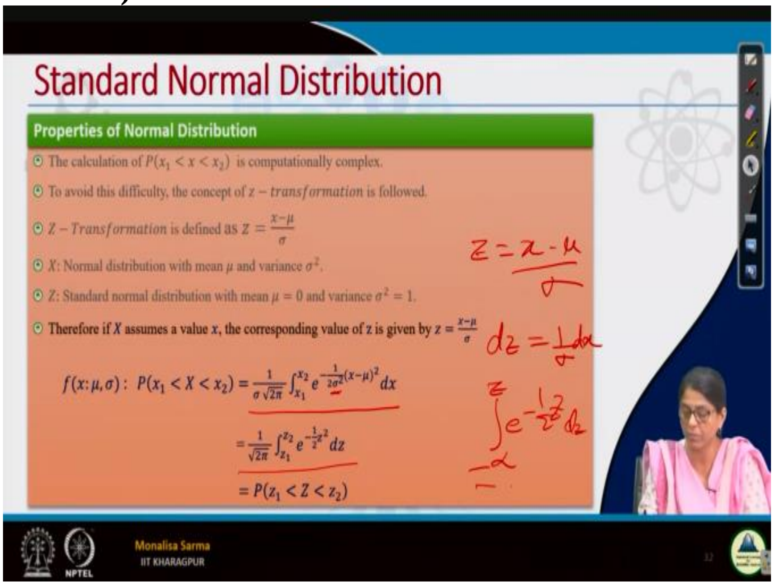
So, this job is made easy it can be I should say it can be made easy if we have a standard lookup table like what we see in the binomial distribution, I assume I told you that we have a table I also showed you remember similarly, if we can have a lookup table, then instead of doing this computing disintegration again and again we can refer to the table and we can come to the value that will be very easy.

So, now, the problem is that how many tables will have because this value this expression will have different values for different μ and different σ square my mean and σ can be anything, it can be any value mean can be starting from 1, 2, 3, 4 y 1 it can be 0, so, it can be any value, even σ also it can be any value, so, we will have to have different tables for different values, is not it? So, that is not possible.

So, it is not really not impossible to have a normal table for such a case, but there is a rescue when we can the rescue is that we can transform this normal random variable, we can transform this normal random variable to is z distribution it is called z distribution and under corresponding distribution is called standard normal distribution, how when we can transform this x to z distribution what happens then our mean is always 0 and a variance is 1.

So, that means, if we have a table for mean 0 and variance 1 for different probability values, then it is very µch visible we can have that we can look up the table. So, there are standard lookup table where we can find the value of the normal distribution, but this random variable, normal random variable first we have to convert it to z value. So, now once we can convert it to z value there are its corresponding distribution of z value has a mean of 0 and variance of 1.

(Refer Slide Time: 22:36)



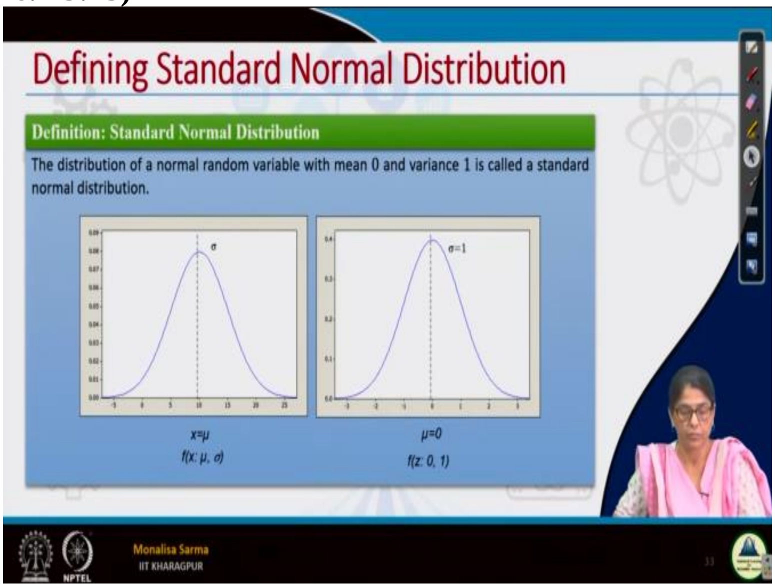
So, this is just how we can convert this z to x normal random variable x to z this is the forµla z = x - µ by σ. So, we need to convert x our random variable is x is not it? We need to convert x to z how we will convert, so, z = x - µ by σ. So, what we got is a z distribution or it is also called standard normal distribution, it has a mean of 0, always, its mean will be 0, this is the y axis, it will always mean will be 0 and standard deviation is 1.

So, we need just a couple of lookup tables with mean 0 and standard deviation 1, but having different because the probability might be different, is not it? So, for the different probability of occurrence, we can have different tables that is all. So, we will see some examples. So, that is what, see here remember this was the expression for f(x) if we are interested in finding out the values between in the interval x 1 to x 2 so, this was our expression.

Now, what we have taken we have taken z = x - μ by σ is not it? Now what will be dz? dz is equals to what? 1 by σ dx sorry 1 by σ dx, this is dx 1 by σ dx that is dz. So, here in this if I can substitute x - μ by σ square x - μ whole square by σ square this σ square is here if I can substitute by z and then substitute in dx by σ this dz so, I will be getting this and this value I can get it from the table but again one more thing there is one more point to that the table has the cuμlative value.

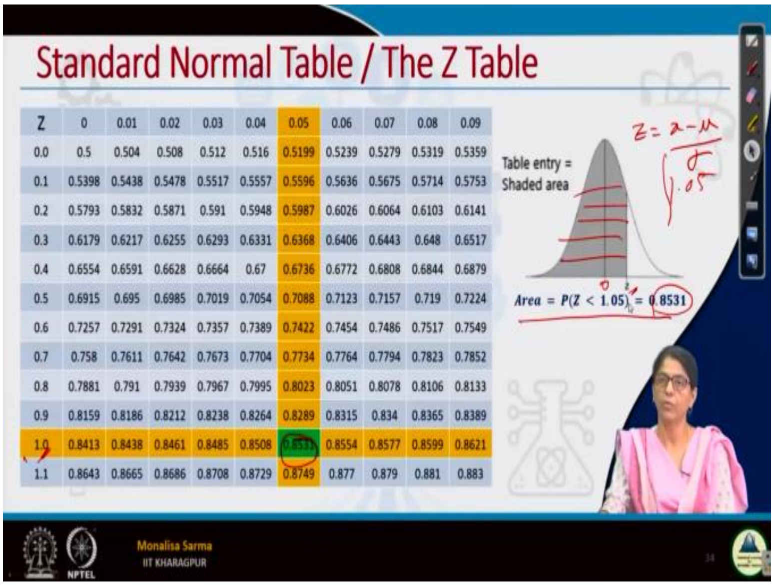
The table what we will see like in binomial distribution, binomial distribution table had the cuμlative value. Cuμlative probability distribution function values Similarly, here also normal table it has the cuμlative distribution value. So, cuμlative means I will have the value say from minus infinity to a particular value set to a particular value. So, z whatever is z and this expression $e^{-1/2 z 2}$ dz this value will get it from the table. So, once we can get the value from the table our job is done.

(Refer Slide Time: 25:25)



That is what the figure shows the difference between the standard normal distribution and normal distribution what I have already discussed.

(Refer Slide Time: 25:33)



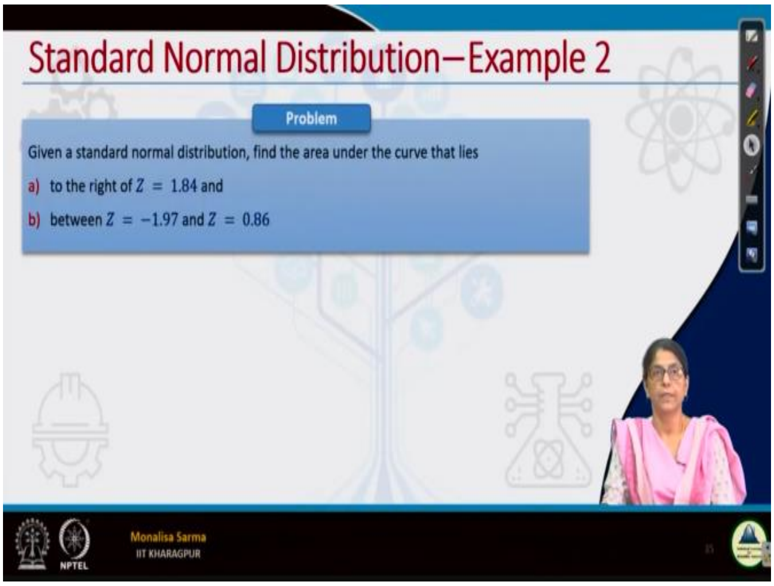
Z	0	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.5	0.504	0.508	0.512	0.516	0.5199	0.5239	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0.2	0.5793	0.5832	0.5871	0.591	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.648	0.6517
0.4	0.6554	0.6591	0.6628	0.6664	0.67	0.6736	0.6772	0.6808	0.6844	0.6879
0.5	0.6915	0.695	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.719	0.7224
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
0.7	0.758	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764	0.7794	0.7823	0.7852
0.8	0.7881	0.791	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.834	0.8365	0.8389
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.877	0.879	0.881	0.883

Now, see how we can look up the table see a table, this is the table suppose in the figure see in this figure there was some x I have found out from what is this z = x - μ / σ by substituting μ and σ value suppose I got my value of z as 1.05. Now, I will have to find out the probability corresponding to that means probability corresponding to 1.5 means integration of this value from - ∞ to a particular value.

So, whatever the to this value of z that is 1.05. As I told you it always gives the cuμlative value. So, how do I look the table see here is so, in the vertical you will see this is z value vertical and horizontal both are z value how we find in a vertical you see this yellow colour this is 1 I need 1.05. So, this is 1 this is 0.05 so, that is total together is 1.05 and where it meets this is my probability 0.8531.

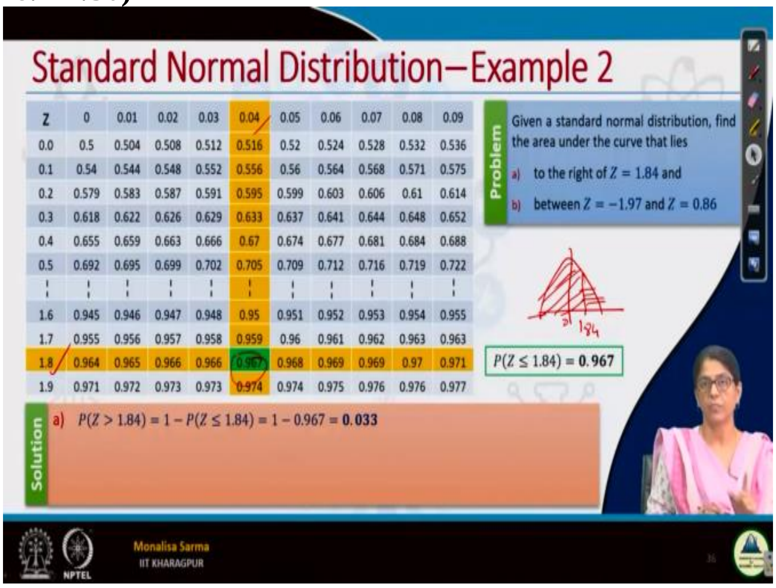
So, if as I told you standard normal distribution median is 0. So, this is 0 if this is the curve this is 0, so, 1.05 that will be 1.05 it will be to the towards the right of 0. So, this is my z value 1.05 and what is this area? This area is basically the probability that my value will lie in this range this area is nothing but the probability so, this area is 0.8531. This is how we look under the table. In the tutorial class we will be doing here also, of course, we will be doing some problem material first we will be doing more problems and things will be more clear.

(Refer Slide Time: 27:21)



Then say given a standard normal distribution find the area under curve that lies to the right of z = 1.84.

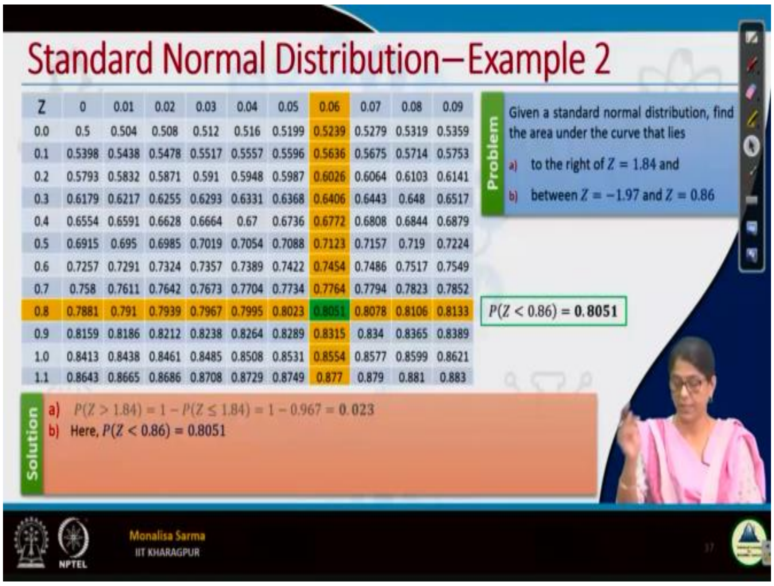
(Refer Slide Time: 27:30)



To the right of z = 1.84 so, right of that is was to 1.84 means, if this is my curve z = 1.84 because this is 0 my z value will be somewhere here 1.84 right means I need to find out this area. So, what but what my graph will give my table will give? My table will give this area left of this table is the cuμlative it will start on cuμlative means from minus infinity to 1.84.

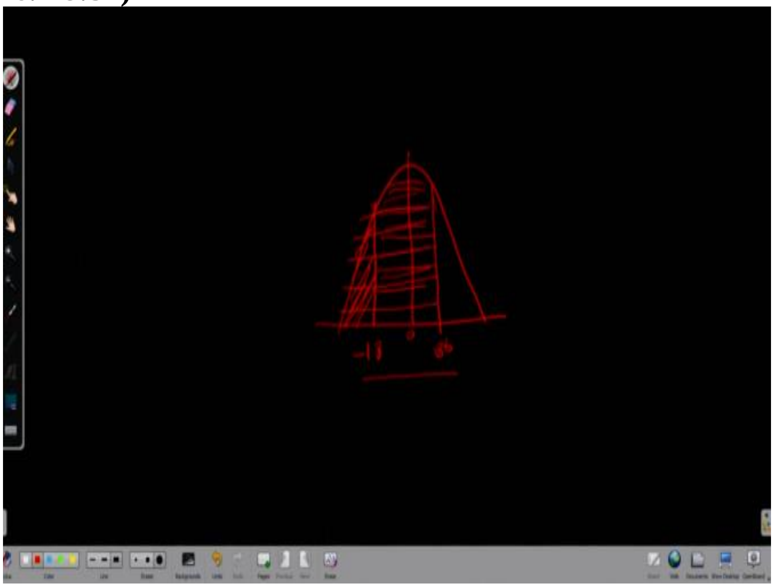
I will get this value, but I need this value that means one minus of that I will be getting this value right. So, here you see, so, it is 1.84 this is 1.8 and 0.04 will make it 1.84. So, I got this is the value then one minus of that is whatever value I needed.

(Refer Slide Time: 28:26)



Similarly between z = - 1.97 and z = 0.86. So, it is asking between z = 1.97 and 0.86 how we will do here suppose let me take the table then it will be better.

(Refer Slide Time: 28:51)



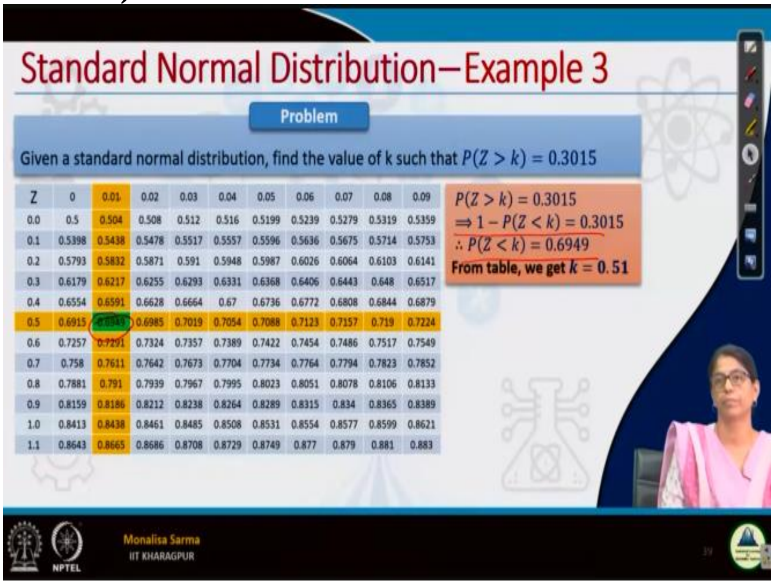
So, it is asking some minus one point I do not remember what was it say minus 1.5 suppose whatever it was in the question I do not remember and it was asking from minus 1 sorry it is not 1.5 will not be this, I will erase it here so, minus one point will be something here say minus 1.8 whatever it is, because this is 0 minus will be this side and between 0.86 so my 0.86 may be say this value, 0.86 so, I need this area 0, I need to find out this area.

That means basically I need to find out this area means what it is asking? I need to find out what is the probability that my random variable will live in this range 1.86 to 0.86. So when I am interested in finding this area, my cuμlative table will have all the values to the left of it.

So if I see 0.86, I will be getting this value from minus infinity to this value. And then when I am looking for 1.8 from here this portion.

But my intention is finding out this portion with between this. So, how will I find between this whatever value I got from 0.86 from this value if I subtract this value I will get this value. So, that is how we look it from the table okay fine.

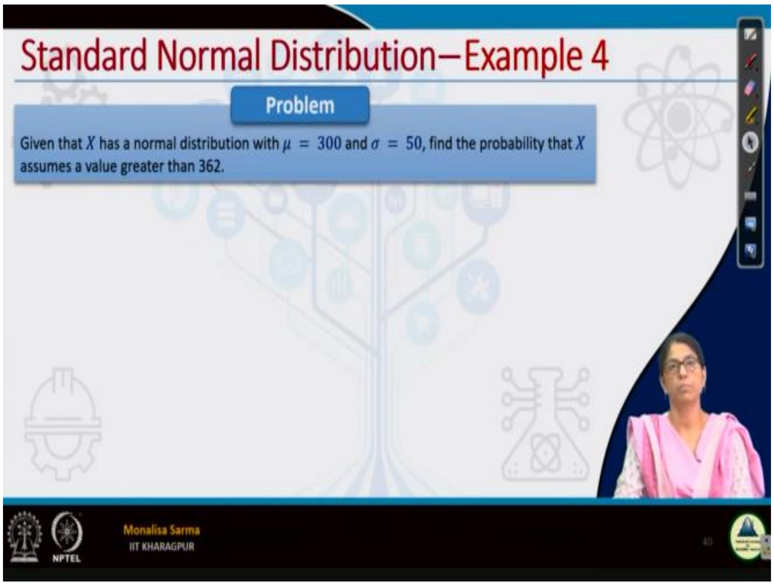
(Refer Slide Time: 30:33)



Similarly, given a standard normal distribution find a value of k such that probability of z greater than k is 0.315, now, we have to find out what is k here. So, it is given greater than k in the cuμlative table it is always less from - ∞ to a particular value. So, first of all when it is that greater than k what we have done first we have tried to find out what is less than k so, z less than k is this value 0.6949.

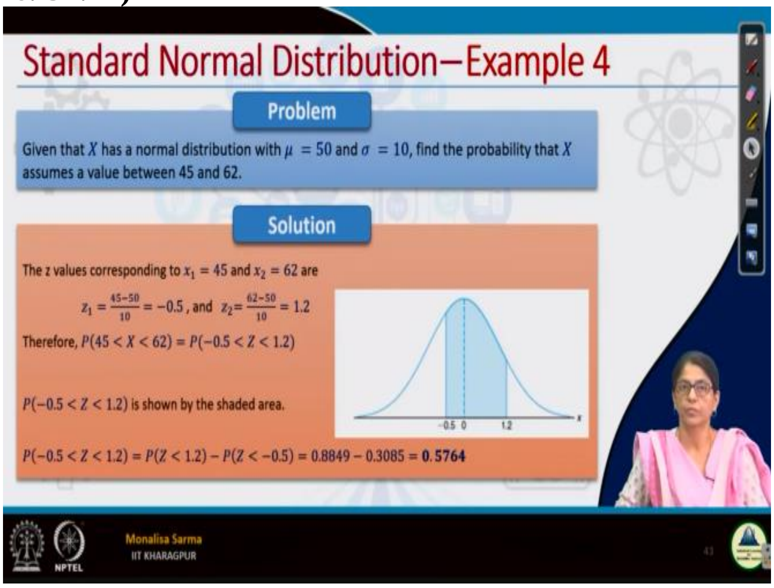
Now, from what is the value of k corresponding to this area this is the all this here in a table all the entries are nothing but the area within this curve so, here it is 0.6949. So, I found 0.6949 is this value. So, what is corresponding to 0.6949 it is 0.5 here it is 0.1 that means 0.51, my k value is 0.51.

(Refer Slide Time: 31:30)



So, one problem given that X is a normal distribution with mean μ= 300 and σ = 50 find a probability that X assumes a value greater than 262.

(Refer Slide Time: 31:44)

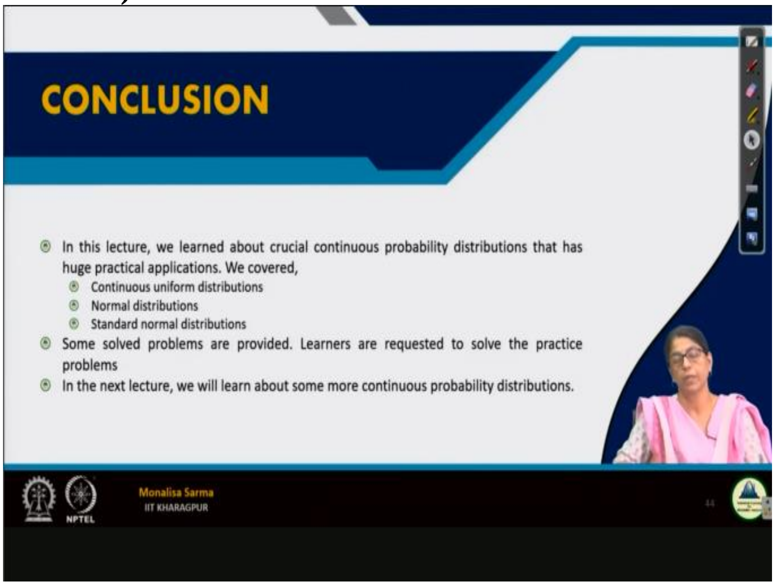


No, sorry, this is there is a slight mistake. That is the question given that X is a normal distribution with μ = 50. And σ is we will put them find a probability that X assumes a value between 45 and 62. So first, between 45 and a random value of the random variable is 45. Value of the random variable is 62 this X, I will have to first convert it to z. So, what is corresponding to x 45 what is the z value I found that value is - 0.5.

Corresponding to x value 62 my z value is 1.2. So, if we can see the figure that means I am interested in this blue shaded area, I want to find out what is the probability that my random variable will take the value from minus 0.5 to 1.2. But in question it was between 45 and 62, 45 and 62 get transformed to minus 0.5 to 1.2. So, from the table I can get this value, what is the probability of getting 1.2 - ∞1.2.

Then again, I can get what is the probability of getting the value minus 0.5. I will get this value. So, from this whole value, if I minus this value, I will get the blue portion that is the area of this blue portion that is the probability. So, that is how we solve this sort of problem.

(Refer Slide Time: 33:10)



So, we will be doing some real application of normal distribution in our tutorial classes. So now in this lecture, basically, we learned continuous uniform distribution, we learned what is normal distribution, what is standard normal distribution. And we have solved some problems though more problems we will be solving in our tutorial classes. And in the next lecture, we will discuss some more continuous probability distribution with that.

(Refer Slide Time: 33:35)

These are the references. Thank you guys.