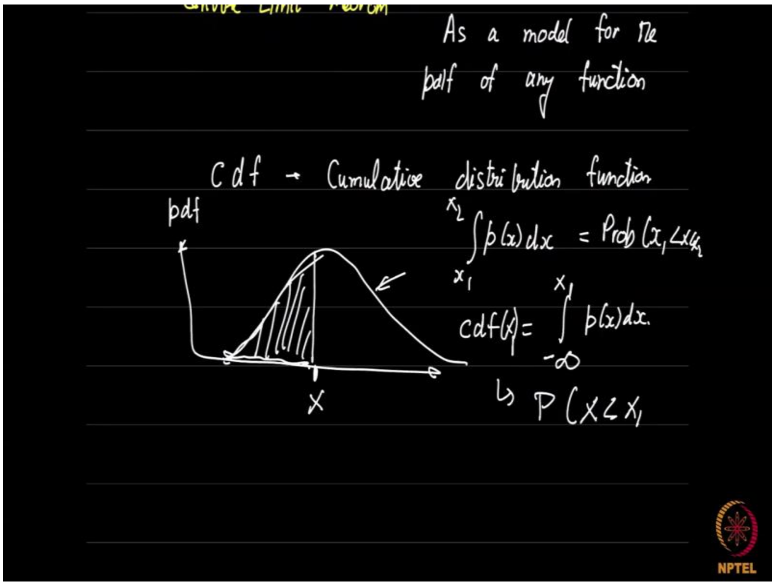
Inverse Methods in Heat Transfer
Prof. Balaji Srinivasan
Department of Mechanical Engineering
Indian Institute of Technology – Madras

Lecture – 35
Gaussian Distribution and the Standard Normal Table

(Refer Slide Time: 00:21)



Welcome back, in this video we will look at a very common probability distribution function. This is called the Gaussian distribution, many of you would have seen this before. This is a very commonly used function within a science and engineering and even in mathematics for several unique properties that it has. One of the most useful properties is something called the central limit theorem which I will talk about very shortly.

It turns out that the average of almost any independent distribution over multiple realizations if it is added together, if it is average together, it tends to words the Gaussian distribution and this is what makes it occur again and again and again. Now again unfortunately we will not have the time to get go through the nitty-gritties of the Gaussian distribution. We are going to more or less use it as a tool but I will just briefly introduce it.

Apart from the Gaussian distribution, there are several other practical distributions that exist within probability. There is for example the exponential distribution there are several other distributions that can be used have very usefully used. But what you should see the Gaussian distribution is as a model for the PDF of any function. As we will see shortly well, we will not really see it, I will just say it makes sense to use especially the cumulative distribution function what is called the CDF of the Gaussian to make predictions about the likelihoods of events.

So, CDF stands for Cumulative Distribution Function and in many ways the cumulative distribution function is more useful, more practically useful than the probability distribution function. Essentially the cumulative distribution function is simply. So, if you have a probability distribution function like this, then the cumulative distribution function. So, let us say,

$$\int_{x_1}^{x_2} P(x)dx = Prob(x_1 < x < x_2)$$

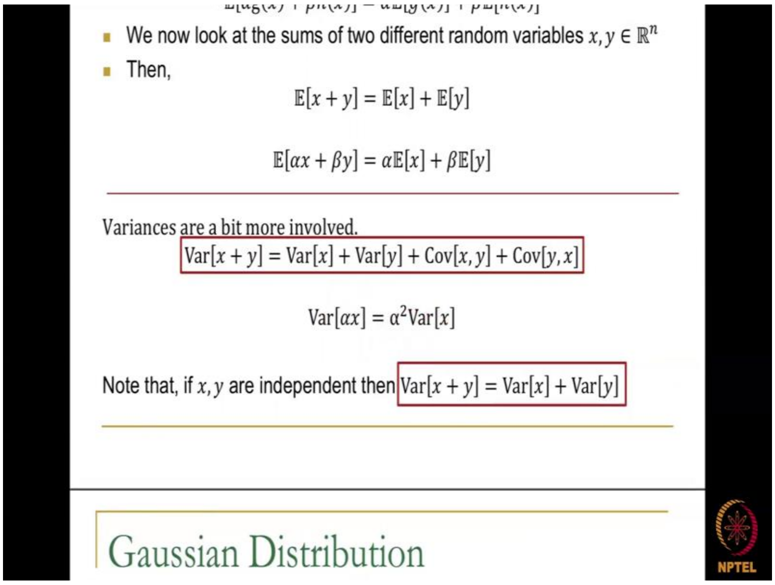
If that is the case then the cumulative distribution function is simply,

$$CDF(x_1) = \int_{-\infty}^{x_1} P(x)dx$$

So, the cumulative distribution function at x gives us the probability that or let us call it x_1, It is probability that $x < x_1$. So, it accumulates probabilities from zero to one of course the cumulative distribution function at Infinity is going to be 1.

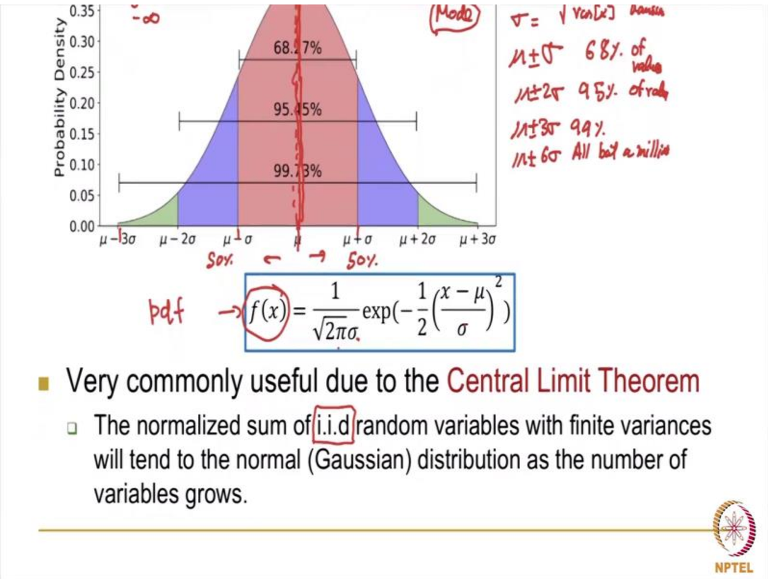
So, we will use the cumulative distribution function of the Gaussian, it has several very nice statistical properties we will not discuss all those here. So, let us move ahead to the Gaussian.

(Refer Slide Time: 04:03)



Before we move to the Gaussian, I will just want to recollect what we looked at the sums of random variables in the previous video. remember that expectations of two different random variables simply sums up as $E(x) + E(y)$. $Var(\alpha x)$ is $\alpha^2 Var(x)$ and variance in case to events are independent also adds up $Var[x + y] = Var[x] + Var[y]$. So, we will use this in what follows. So, I just wanted to recollect that.

(Refer Slide Time: 04:40)



So, let us come to the Gaussian distribution. This distribution is a very standard probability density function and it has the following expression you can see this here, $f(x)$ which is the Gaussian PDF is given by $\frac{1}{\sqrt{2\pi}\sigma}$, what is σ? σ is the standard deviation of the Gaussian distribution or you can see simply μ at σ as two parameters of the Gaussian of that PDF of the Gaussian.

Just Like a and b were two parameters of the arbitrary function that I wrote out in the previous video. Similarly, you can see μ and σ but why did we choose μ and σ and not something else, that is because there is this nice shape. the Gaussian Peaks at whatever value of mu it is and it turns out that the standard deviation of the Gaussian is exactly Sigma. So, mu for the Gaussian happens to be the mean, it also happens to be the median, it also happens to be the mode of the Gaussian.

So, Gaussian has this unique property that mean, equal to median, equal to mode. So, to recollect mode is the value which is the most frequent of course it is not really fair to speak about this for continuous distributions, but you can see that if I drew a histogram, then the

Gaussian would Peak right around new you can also see that it is the median in that 50 percent of the values are below the mean and 50 percent of the values are above the mean.

μ also happens to be the expectation that is if I take this function,

$$\mu = \int_{-\infty}^{\infty} xf(x)dx$$

It turns out with this expression all the sigma's will cancel out and it will come out to exactly μ. So, that is why this is defined that way. Now what about Sigma it turns out that the,

$$Var[x] = \sigma^2$$

So, σ is the, $\sigma = \sqrt{Var[x]}$, for the Gaussian there are other further properties which would have also seen earlier if you have not let us take a brief look at it within this video which is that within plus or minus Sigma 68 percent of all values are covered. So, for example let us look at the heights of adult males. So, if I look at let us say it approximately looks like a Gaussian, then if I look at the mean height plus minus one standard deviation usually 68 percent of men slides will come within that.

Now if I take mean plus or minus two standard deviations and this is really what enables a lot of polling, we will say with 99 confidence we think is that x y z will get So much percentage of the words. the reason they say it is with this. So, you take the mean number of words look at the standard deviation of those votes and let us say they lie on a Gaussian then μ plus one minus 2 σ covers 95 percent of the value.

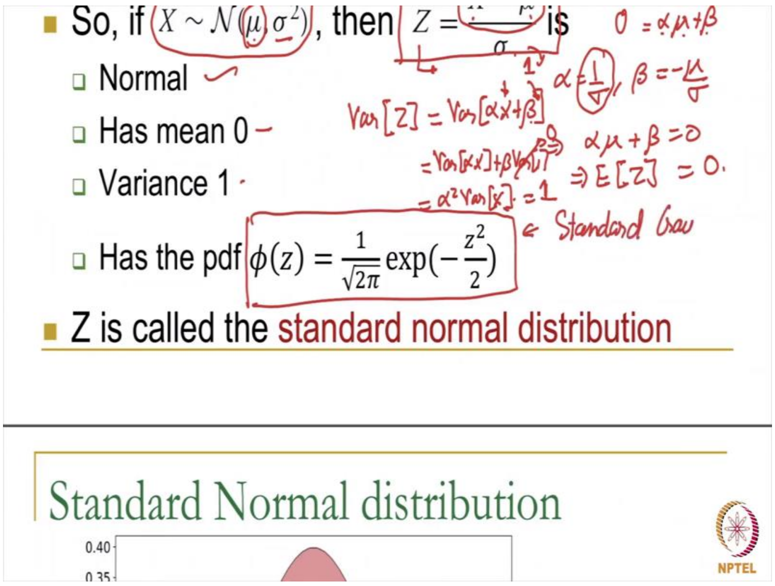
So, this is a very useful number 95 is always plus or minus 2 standard deviations. So, you can take a reasonably small sample and actually guess about 95 of the variances of what is happening provided your entire distribution is a Gaussian of course mu plus or minus 3 Sigma, gives you 99 of the value and you might have heard the word Six Sigma. So, mu plus minus 6 Sigma covers all bit a millionth.

So, so when they talk about $\mu \pm 6\sigma$ what they mean is they want accuracy such that only one in a million cases will fail. So, that is 6σ accuracy. So, this is extremely useful for all sorts of random experiments that is because even if the individual distribution is not Gaussian the normalized sum of identically distributed remember the term IID again independent identically distributed random variables with finite variances will tend to a Gaussian distribution as the number of variables grows.

So, I do not want to discuss this too much this, requires a full course of Statistics or at least it is usually a part of a full course and statistics and we are relegating almost all of probability and statistics to just this one week. So, obviously I cannot discuss this in detail. But this is the reason why you will see this kind of Gaussian coming again and again starting from quantum mechanics to solid mechanics to fluid mechanics the shape haunts us again and again and we will use it again next week also.

Of course, Gauss who was the person incidentally who came up with the least squares who also came up with this distribution. Now if you look at this, this has two parameters mu and sigma. Now it turns out that just like similar triangles you can make similar Gaussians and we standardized as a Gaussian to what is called a standard normal table.

(Refer Slide Time: 10:54)



And that standard normal table or the standard Gaussian looks like this the mean is at zero and one standard deviation is one. So, 68 percent the values lie within one. So, this is just a scaled version of the Gaussian why is this useful because it turns out every Gaussian can be rescaled to a standard Gaussian and we can just make the tables once and for all and as you will see later on in this video, that helps us solve a lot of practical problems.

So, we use in order to scale we reuse this original idea here the fact that $E[x + y]$ is $E[x] + E[y]$ and similarly this idea of a variance also. So, let us use that here. So, let us say,

$X \sim \aleph(\mu, \sigma^2)$ like we had in the original Gaussian table. Now suppose I want to change this to variable z such that.

So, usually we denote the original one as this the Gaussian as normal this N stands for normal or the Gaussian distribution with mean mu and standard deviation Sigma Square then we will say that I want to rescale this. So, that z is a normal distribution with mean 0 and standard deviation 1 for variance one. Now what scaling of the variable will enable that. So, suppose,

$$Y = \alpha X + \beta * 1$$

then expectation of y is going to be, by linearity Alpha times expectation of x plus beta times there is nothing, there is no random variable here that expectation is basically zero.

So, we can actually. So, this is scaled by Beta times 1 expectation of 1 is simply one. So,

$$E[Y] = \alpha E[X] + \beta$$

So, let us say we want expectation of y to be 0 and Alpha times expectation of this is Mu. So, this gives us, $\alpha\mu + \beta = 0$, in order for our rescaling. So, a simple idea is we take a rescaled variable and we use the meaning that I told you the meaning of mu is where it Peaks.

So, I know that my original Peak was at μ and I want the new Peak at 0. So, I subtract mu and I know that this was $\mu + \sigma$ and I want the sigma down to 1. So, I will scale the entire variable by a factor of σ. Now what happens here if I do this then this means, $\alpha = \frac{1}{\sigma}$ and $\beta = \frac{-\mu}{\sigma}$. So, you can immediately see that $\alpha\mu + \beta = 0$ which means, $E[z] = 0$.

Now what happens to the variance? The variance it turns out is,

$$Var[z] = Var[\alpha x + \beta]$$

Now we saw that the sum of the two variances can be the sum of the variances provided the two variances the two variables are independent. Now notice x is a random variable, here there is no random variable there is only one here. So, these two are independent. So, this is variance of alpha x plus beta times variance of a constant 1 which is just 0.

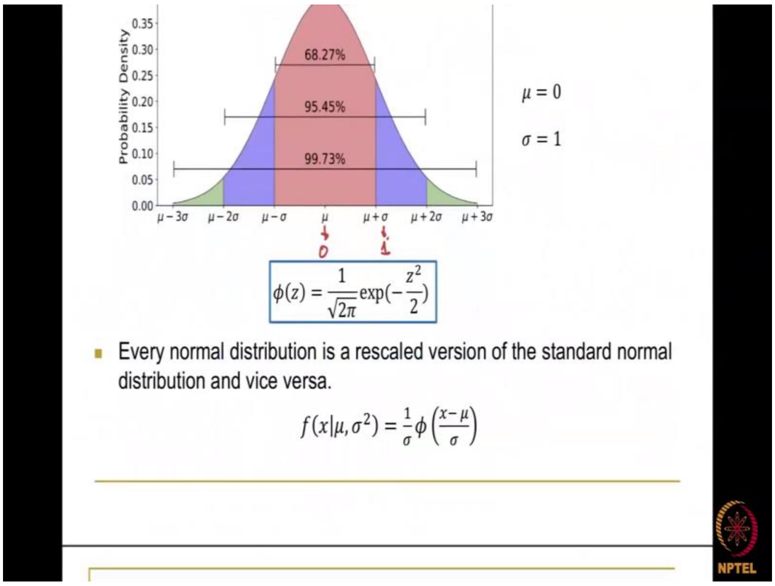
So, this is Alpha Square variance of x, which we had written before. Now Alpha square is 1 over a sigma Square Times the previous variance. So, you basically get just one over Sigma Square multiplied by Sigma Square. So, this is 1 so, all put together basically what you get is this new variable the scaled variable $\frac{x-\mu}{\sigma}$ is normal of course it has mean 0 and it has variance 1.

Now if I take a look at the previous one and rewrite this so, this becomes F of z is you should not call it FB in speaking I will call it,

$$\phi(z) \; = \; \frac{1}{\sqrt{2\pi\sigma}} e^{\frac{-z^2}{2}}$$

So, that is what I have written here. So, this is called the standard Gaussian or the standard normal distribution as I have written.

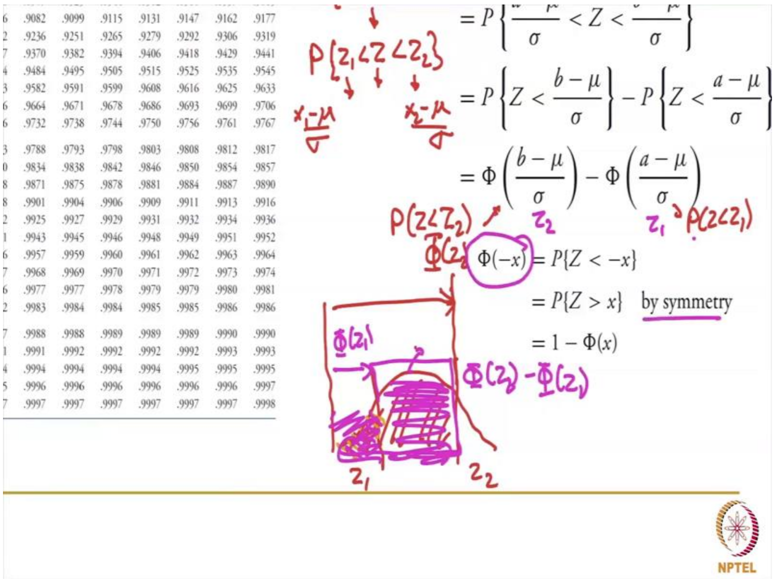
(Refer Slide Time: 16:37)



So, what is the standard normal distribution look like. it looks like the same thing except of course mu is 0, Sigma is one. Otherwise, there is no look it is just like zooming in or zooming out and just moving this around to the left or moving it to the right. So, we have moved it to the left by a value of mu and we have zoomed out. So, that the net variance is simply one. So, you just scale it so, that it is one.

So, the point here is every normal distribution is a rescaled version of the standard normal distribution. So, as I have written here mu equal to 0 and sigma equal to 1, and that makes it very powerful. Now why does this make it powerful. So, let us look at that in the next slide.

(Refer Slide Time: 17:30)



So, here we see the standard normal distribution. I will explain how we are showing it here shortly. This is from a book by Sheldon Ross I have just taken a screenshot from that book, this is a book by Sheldon Ross and he's written a series of books on probability highly recommended in case you want elementary probability Theory. Now notice what is written here is integral from minus infinity to x and this is capital Phi of x.

So, what we have here is actually the cumulative distribution function, because that happens to be very useful as I will show you very shortly. it is extremely useful, but the function we are using is the Gaussian $\frac{1}{\sqrt{2\pi\sigma}} e^{\frac{-z^2}{2}}$ or $\frac{y^2}{2}$ whichever way you want to do it. So, $\Phi(z)$,

$$\Phi(z) = \int_0^z \phi(x)dx$$

This is PDF this is CDF cumulative.

So, $\int_0^z \phi(x)dx$ so, till wherever you want if this is the PDF you sum it up and that is what gives you the cumulative distribution function. So, this gives you Capital Phi of z if this value is z. Now why is this useful, notice that at x equal to 0.5 the x equal to 0 CDF is 0.5 which makes sense. Why does this make sense? Because you have this if I integrate up until this point, the integrated value phi at zero is equal to 0.5 because exactly half of it is below zero and if I go to Phi of infinity that should be one.

So, you will see that here as we keep on increasing in value, we will see that about 3.5 as I told you 99.9 percent of the values are filled there you will get 0.9998. So, you get a high value when you come somewhere here. So, usually tables are written only from 0 to some positive value or from negative value and there are different ways of writing this table the thing the exercise I have asked you to look at a standard Gaussian table.

You do not need to take a look at only this, you just put standard Gaussian table you will find a million of those which is searched within Google. So, coming back here, let us say I have a variable this is the unscaled variable. This becomes clearer when we look at an example but let us say this is an unscaled variable, the original variable, let us say I have drawn Gaussians of heights. So, I want to find out what is the probability that any person is lower than let us say four feet high an adult plane is lower than four feet.

So, this is x this is the value this is the variable x and this is the value x_1 which we are trying to make this lower than. Now instead of writing a Gaussian in terms of x with the specific mu and sigma that we have we scale it we say that instead of doing $x < x_1$, I will instead transform this to $z < z_1$. So, where z is, $\frac{x-\mu}{\sigma}$ so, this is z. And similarly, z_1 is $x_1 - \mu$.

So,

$$z_1 = \frac{x_1 - \mu}{\sigma}$$

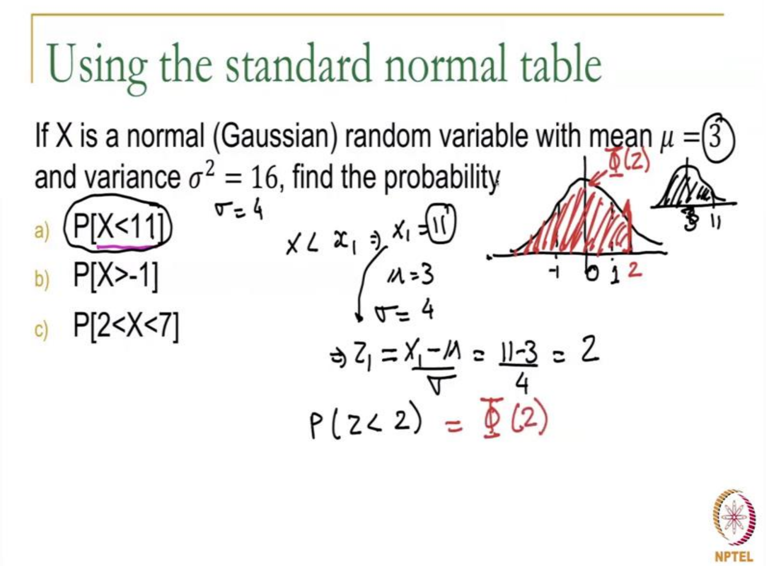
again, like I said easiest to see this in an example. Now when I want to find out when is x less than b, all I need to do is to find out where this value. So, this cumulative distribution function if I can find out this is capital Phi of z1. So, I locate the point this stays at and then simply look it up in this table and this is a standardized table. This is a standardized table you will look it up there.

Similarly, if I want the probability that $x_1 < x < x_2$, all I do is I transform this into another equation $z_1 < z < z_2$, where $z_1 = \frac{x_1-\mu}{\sigma}$ and $z_2 = \frac{x_2-\mu}{\sigma}$ and z is of course the standard normal table. Now how do we find out P, this is probability that $z < z_2$ this is probability that $z < z_1$. Now why does this work out.

Suppose I want the probability that z lies between these two variables z_1 and z_2 probability that z lies is less than z2 is given by $\Phi(z_2)$., that is the area under this curve. similarly, the area under this curve shown here let me color it in a different color, that is a pink, this area is $\Phi(z_1)$., what we actually want is all this area in the middle which is $\Phi(z_2)$, which is this whole area minus $\Phi(z_1)$. which is this restricted.

So, the area in the middle is given by $\Phi(z_2) - \Phi(z_1)$. Now if we want for a negative value. So, all we need to do is to flip the table and we can find that out by symmetry. So, let me show you a couple of examples of this and then this entire thing will become a little bit clearer. So, let us see an example now.

(Refer Slide Time: 25:05)



So, let us say x is a normal Gaussian random variable with the mean equal to 3 and variance equal to 16. So, I am asking you to find out the probability that x is less than 11. So, if you were to do it without the scaling, what will look like is you'll first draw the Gaussian and the Gaussian will be centered around mean equal to 3 and its standard deviation will be in this case Sigma is 4 variances of 16.

So, plus 7, minus 1 and then you would have to integrate this area of x less than 11. Now you will have to do this integral. Now how do you integrate this you have to integrate that PDF again the complicated looking PDF with $e^{(\frac{-(x-\mu)}{\sigma})^2}$ Now instead of doing all that we risk it we rescale the entire problem and pause it in terms of that.

So, these rescale the exam problem to say all problems will be turned to problems with z mean a zero standard deviation is one and I will map each one of these excess into this location. So, let us look at. So, this is x is less than let us say some x 1 and x 1 is 11, mu is 3, Sigma is 4 which means which means the corresponding z1 is, when we ask corresponding z1 really speaking what we are asking for is how many standard deviations is this value away from the mean.

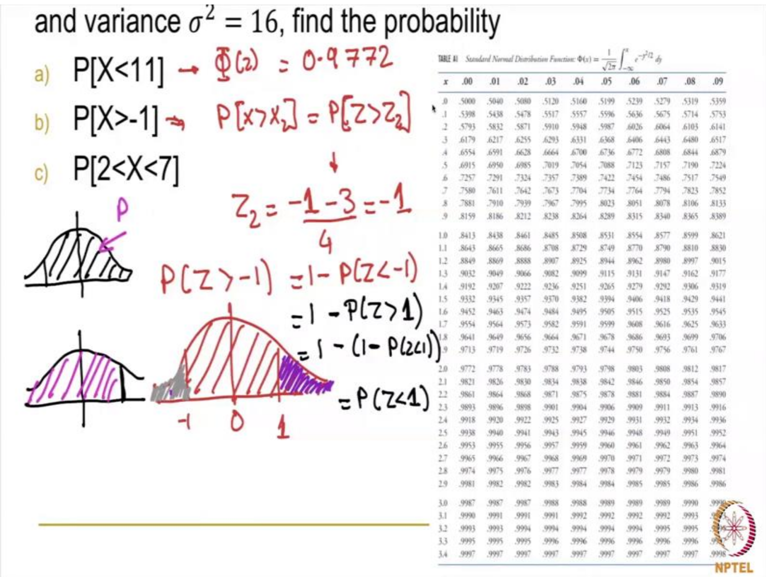
How many standard deviations up am I looking at really speaking that is what we are asking. So,

$$z_1 = \frac{x_1 - \mu}{\sigma} = \frac{11 - 3}{4} = 2$$

So, in some sense what we are asking is how many of these values are below 2 Sigma, starting from minus infinity. So, what we have to look at here is $P(z < 2)$. Now we have turned it to a problem which simply says the proportion of things which were less than 11 in the original graph, remember mean is 3 and standard deviation is 4.

There is the same as the proportion of things since we are looking at probability which are lower than a standard deviation of 2 or sorry a value of 2 within the z table. So, for this of course we need the integral and the integral is denoted by $\Phi(z)$, there is no analytical value there which is why we write it as a table $\Phi(z)$, z = 2. So, this is p of z less than 2 is Capital Phi of good. So, that is what we need to look up within the standard normal table in order to find this value on it. So, let us look at that value now.

(Refer Slide Time: 28:42)



So, here is our standard probability distribution table the way to read it is I want an x or a z of 2. So, I will come down here and look at 2. and it is 2.0 it is not 2.01 if it was 2.01, I would have come here, if it is 2.02 here but since it is just 2, I just look at it here and this is what I get

0.9772. So, 0.9772 basically means that $P(x < 11) = \Phi(2) = 0.977$. The second question is what is the probability is that $x > -1$.

So, once again let us look at the reference graph the reference graph is we had 3, 11, you do not need to draw this, I am just drawing this because if some of your unfamiliar with this. So, this is just one standard deviation below will actually give you minus one. So, in the original graph because remember the standard deviation was 4. So, 3, 7, minus one whereas in the z table so, let us say,

$$P[x > x_2] = P[z > z_2]$$

The same proportion of objects provided we map it correctly are going to be located there.

So, just as an example instead of calling my dice 1, 2, 3, 4, 5, 6 suppose I call it one by seven, two by seven, three by seven, four by seven, five by seven and six by seven that still does not change anything about how many times I am going to get one by seven it is still going to be the same value as before I just rescaled the basic what I call the basic values and that is the cleverness in this trick here.

So, if we come down to this,

$$z_2 = \frac{x_2 - \mu}{\sigma}$$

So, this comes to minus 4 this minus 1. So, basically one standard deviation to the left so, we want, $P[z > z_2] = P[z > -1]$. So, in terms of the graph z graph this is minus 1, this is one, we are looking for this area we are looking for this area. Now there are multiple ways of calculating it, one way of calculating it is like this I am going to show you one way here you can choose to calculate it in many other ways.

If you wish I will calculate this area, calculate this as,

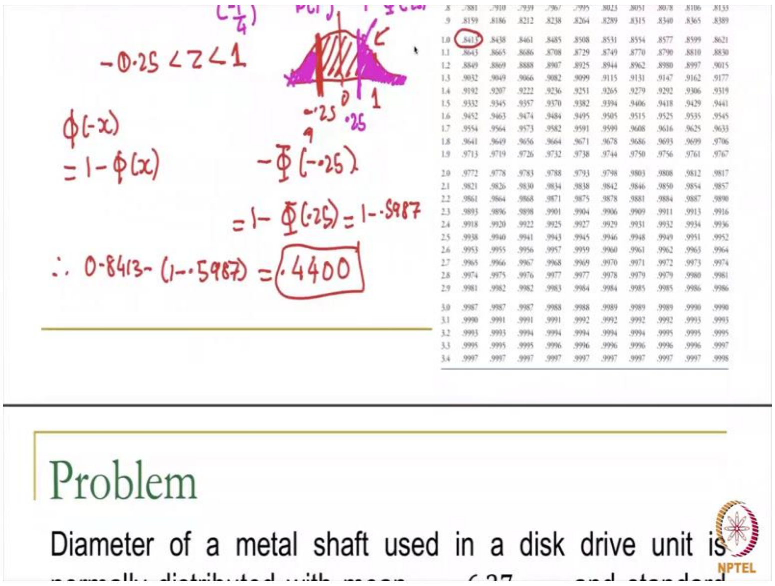
$$= 1 - P[z < -1]$$

But of course, I do not know that area also I am going to. Now calculate that area as same as. So, I'm doing this in a convoluted fashion just to show you the different ways in which this can work out. So, this is exactly the same as, $1 - P[z > 1]$ by symmetry but probability that z is greater than one is 1 minus probability that z is less than one does that make sense.

So, I I hope you can see that this is exactly the same as probability that z is less than one again you can do this directly through symmetry. So, this might look like an unnecessarily convoluted calculation I can show it to you in another way. So, I wanted the probability that z is greater than minus 1 which is this probability by symmetry you can see that this is exactly the same as this area.

So, this is probability that z greater than minus 1 this is probability that z is less than. So, the two are the same which basically is Phi of one. So, if you see Phi of one it is here 0.8413. So, you can say that this probability is 0.8413, what probability, the probability that x is greater than minus 1. So, finally we have the third part which is probability that x lies between 2 and 7. So, let us calculate that now.

(Refer Slide Time: 34:00)



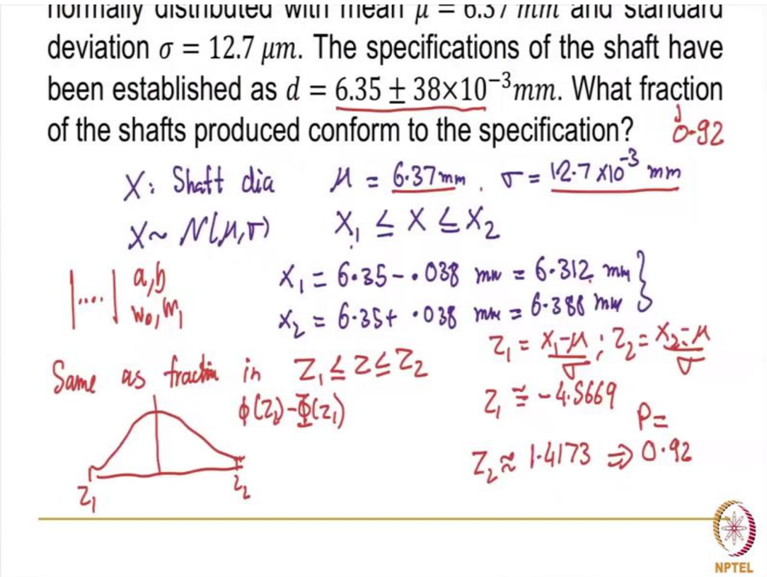
We will use the same trick as before we will say probability that x lies between x1 and x2. So, which is the same as the probability that $z_1 < z < z_2$; z1 is 2 minus 3 divided by 4 and z2 is 7 minus 3 divided by four. So, this basically gives us probability of minus 1 by 4 probability one way sorry not probability of one by four minus one by four and one. So, we basically want the location that probability or let us say write it more accurately as z lies between minus 0.25 and 1.

So, once again let us draw this for reference minus 0.25 here 0 is the mean, one is little bit later and I want this area. So, the way to find out this area is to say this is the same as $\Phi(1)$, minus $\Phi(-0.25)$, $\Phi(1)$ is straightforward we just calculated it $\Phi(1)$ is we can see that here 0.8413 but what about $\Phi(-0.25)$,. So, $\Phi(-0.25)$ calculates this area here.

Now how do we calculate this area. So, what we can do again is by symmetry we had to draw $\Phi(0.25)$, and say this area here is going to be the same as this area here. Now what is this area here, this area here is $1 - \Phi(0.25)$. So, we can use this relationship that Phi of minus x is equal to 1 minus Phi of x and that gives us, this is $1 - \Phi(0.25)$. what is $\Phi(0.25)$ we can come back here0.2 and 0.05, 0.5987, this is point one minus point magnitude.

So, overall, we have 0.8413 minus 1 minus 0.5887. So, if we calculate this this constitutes 0.4400. So, the point of this exercise was to show that you can take variables with some mean and some variance and rescale them so as to be able to calculate it using the standard Gaussian table. Now I want to use this example for a slightly more practical problem which is the next problem that we will see now.

(Refer Slide Time: 38:14)



So, let us look at this problem let us say we have a disk drive unit and we have a metal shaft and typically after manufacturing you are not going to get exactly the same size for everything. So, you can see that after manufacturing the mean disk drive unit diameter the shaft diameter is 6.37 mm, whereas the standard deviation, there is some error in the machine. So, you have this in micrometer notice this if it was in millimeter obviously you cannot do anything with this machine at all.

In this case you have a machine it is after machining you have a small amount of variation within micrometers. So, what we want to find out is there are specifications over which this this drive unit will work we want to check whether how many of the things that we manufacture are going to lie within those specifications. we are told that the specifications are such that we want to be within this range, we want to be 6.35 plus minus 0.038 millimeter.

So, we want to check how many of the shafts that we produce actually conform to this specification. So, let us write this out formally once more. So, let us say x is the random variable which is the shaft diameter. what do we know we know mu of x is 6.37 mm we know Sigma is 12.7 into 10 powers minus 3, I am going to write everything in mm. Now we also know that x is normal with of course mu n Sigma as our mean and standard deviation. Now what we want is x should lie between some limits.

So, x1 less than equal to x less than equal to x2 this is what we want. x1 of course is 6.35 minus 0.038 mm. So, x1 is 6.312 mm, x2 is 6.35 plus 0.038. So, that is 6.388 and. So, what we want to know is after doing our computation how many of the values will lie here and the way we do it is of course we are going to keep the proportion the same, we are going to say same as fraction in z line between z1 and z2.

Now what is z1 the z1 is x1 minus mu by Sigma where mu is given by 6.37 and sigma is 12.7 into 10 power minus 3 and z2 is x2 minus mu by signal. So, now if you calculate these values it comes out to be. So, you can see x1 is 6.312 minus 6.37 divided by this 1.27 or 12.7 to 10 power minus 3. If you calculate this, this comes to approximately minus 4.56 somewhere around there.

Similarly, z2 comes out to approximately 1.4173. So, once again I always find it convenient to draw these figures z1 is somewhere here z2 somewhere here what we want is $\Phi(z_2) - \Phi(z_1)$. Now if we go back to our normal distribution table you can see that this is 3.5 there itself it is approximately 0.99. So, if I look at z1 that is far more below than that. So, this is minus 4.4 something. So, you might as well treat $\Phi(z_1)$ you can read it as $1 - \Phi(-z_1)$. 1 minus Phi of minus z1 which is here which is entire table.

So, this is approximately zero. So, this is approximately one or you can call this minus z1 and plus z1 whichever way you want it. So, after four standard deviations nothing really survives. So, nothing is going to be lower than the lower end of the diameter whereas Phi of z2 uh z2 was 1.4173. So, we can look that up here. So, 1.41 or I can look at 1.42 approximately is 0.92.

So, we can say that about 92 percent of these are going to be within the specification. Now this is very useful because you can. Now control the quality of manufacturing and you can actually make some predictions about what will happen when you do manufacturing with some amount of tolerances. You can also probably guess that this kind of thing is extremely used for heat to be useful for heat transfer.

We can say in how many cases when you run a full app in how many cases you are you expected to actually blow up or expect to reach temperatures which are too high. Similarly in the cases that we were looking at with this slab also not only do we give a and b or W0 and W1 you can actually start talking about the uncertainty in the weight estimation that you have done.

Instead of simply giving that these are the optimal weights or this is the flux you can actually start saying well 95 percent of the times the flux value will lie between this and this value. So, that is far more useful. So, if you give mu and sigma at least it is far more useful than just giving mu which is what happens to be the case when we were doing our typical function fit. So, what we saw in the current video was the Gaussian distribution.

And how you can use it to make a range of predictions, in the range of probabilistic estimations we will use all our ideas. So, far put together we will use that in the next week to give probabilistic versions of the algorithms that we have looked so far. So, I hope this week was useful I will see you again in the next week, thank you.