PROBABILITY THEORY FOR DATA SCIENCE

Prof. Ishapathik Das

Department of Mathematics and Statistics

Indian Institute of Technology Tirupati

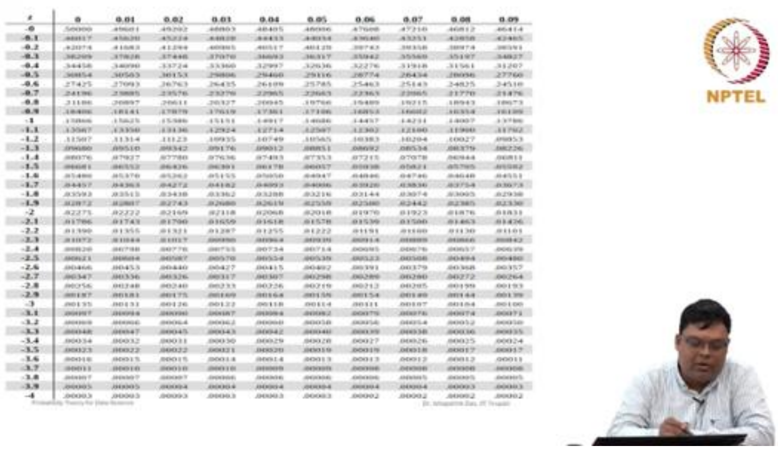
Week - 06

Lecture - 28

Numerical Examples of Normal Distributions

So, the standard normal variate has a probability curve, and the probabilities are given here. This is called a probability table for the standard normal distribution function. Now, to use the standard normal distribution function, we can find probabilities like this: suppose Z follows a normal distribution with a mean of 0 and a standard deviation of 1, which means it's standard normal. So, what would the density function be? $f(z) = 1 / \sqrt{(2\pi)} * e^{(-z^2 / 2)}$. This is the density function. Now, what is the cumulative distribution function? For the standard normal variate, the cumulative distribution function, or CDF of Z, is denoted as $\varphi(z)$. So, $\varphi(z) = \int$ from $-\infty$ to z of f(t) dt.

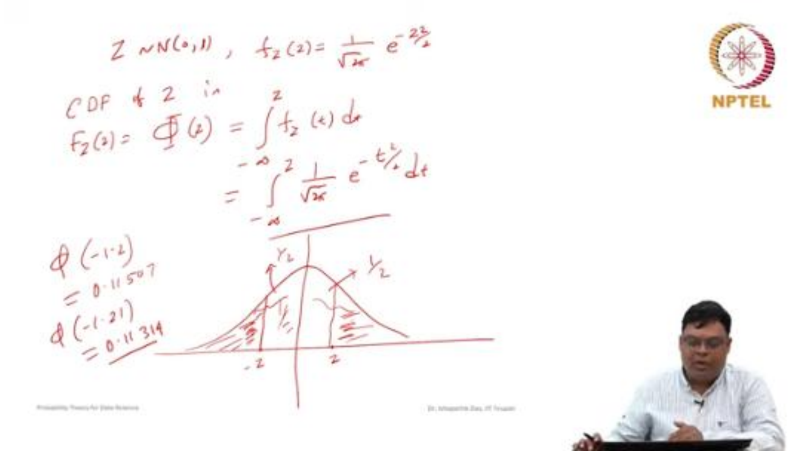
Now, this is nothing but the integral from $-\infty$ to z of $(1 / \sqrt{(2\pi)}) * e^{(-t^2 / 2)}$ dt. But this is intractable. So, this is just how the distribution density function is represented. So, for any z here, suppose z is this value, you want to find the area under the curve. This area, for z anywhere here, cannot be found directly; it requires some computation. But if you use the table, you can easily find this probability. For example, suppose you want to find the value for z = -1.2. So, you can see that this area is given. Suppose the value here is -1.2. You can see that this value is $1 - \varphi(-1.2)$. So, this probability value will be 0.11507.



We say that if we take $\varphi(-1.2)$, where $z = -1.2$, it can be found here as 0.11507. So, in this way, for all these negative values, it's given because of symmetry. Suppose z is here, by symmetry... This probability will actually be the same as for z. If z is positive and we want to find the probability for -z, then by symmetry, we can find this probability. So, you can find this probability as well. This probability will be because the total is 1. So, this is 1/2. The total is this 1/2, and this is also 1/2. You can find this probability by subtracting from the 1/2 minus this area.

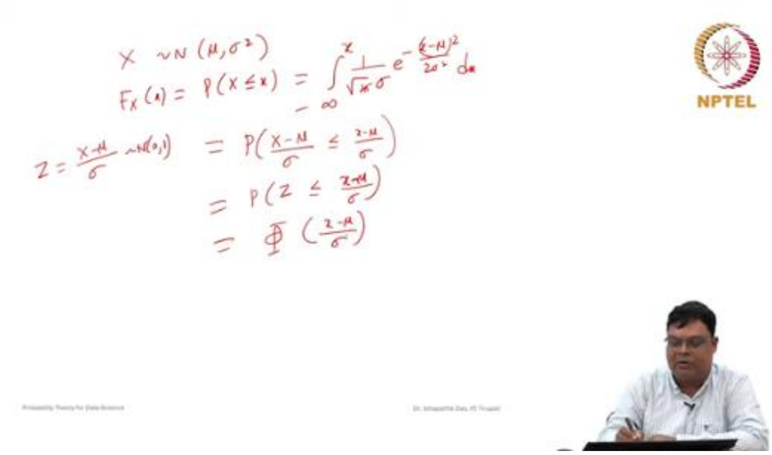
In this way, we can find any probability, because all these negative values and probabilities are given. These probabilities are given up to two decimal places, like -1.21. So, this is -1.21. You can see this in the column with -1.21, then 0.11314. So, that is why this is $\varphi(-1.21)$. This is nothing but 0.11314 in the second decimal place.

So, this is 0.11314. In this way, we can find any probability. Whenever you know the probability of the standard normal variate, suppose you want to find the value. Suppose X is any normal distribution with mean μ and variance σ^2, and now you want to find this value, $f_X(x)$. This is nothing but $P(X \leq x)$. This is also not tractable because the integration will involve $-x * (1 / \sqrt{2\pi\sigma^2}) * e^{-x^2 / 2\sigma^2}$.



This is the density function of the normal distribution with mean μ and variance σ^2 dx. You cannot find this easily, so you have to perform numerical integration. However, we can use this transformation because $Z = (X - \mu) / \sigma$, which has a normal distribution with mean 0 and variance 1. So, what will we do? This probability is equal to the probability that $(X - \mu) / \sigma \leq x$, where X is a real number.

Now, $(X - \mu) / \sigma$ is nothing but Z. So, the probability that $Z \leq (X - \mu) / \sigma$ is known. This is the probability that $Z \leq$ some value. Actually, this is the probability that $Z \leq (x - \mu) / \sigma$. So, this value we can find; this is nothing but, by notation, $\varphi((x - \mu) / \sigma)$. So, $x - \mu$ and σ are known. Then, for any small x value, $(x - \mu) / \sigma$ can be found. From the table, we can compute this probability. If you have understood this, we can also look at some numerical examples. By using numerical examples, it will be clearer how to use the table and find the probability.



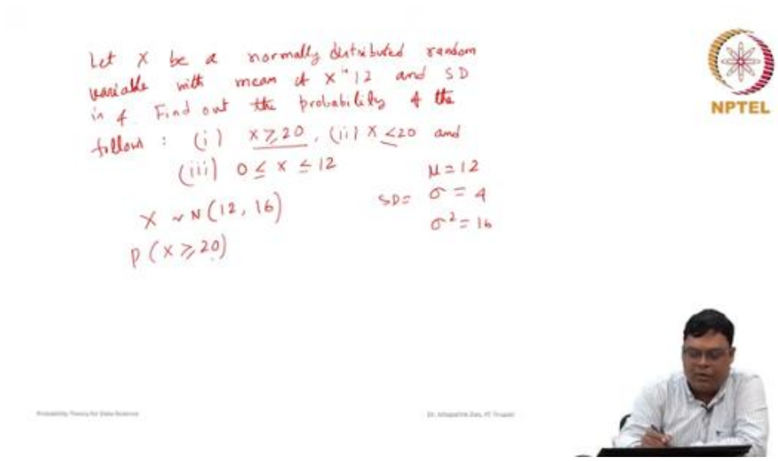
Let us discuss some numerical examples for the normal distribution. So, one problem. Let X be a normally distributed random variable. Let X be a normally distributed random variable with a normal distribution, where the mean of X is 12 and the standard deviation is given.

So, the standard deviation is nothing but the square root of the variance. The standard deviation is 4, and the mean is 12. Let X be a normally distributed random variable with a mean of 12 and a standard deviation of 4. Find out the probability of the following. So, some of the probabilities you have to calculate.

What are those? One, what is the probability that $X \geq 2$? And what is the probability that $X \leq 20$? And the third one is, what is the probability that $0 \leq X \leq 12$? So, let us find out these probabilities.

There are some other questions as well that we will do. So, whenever X is a normally distributed random variable, the mean is 12 and the standard deviation is σ. The mean, μ, is equal to 12, and σ is the square root of the variance, which is the standard deviation. So, the standard deviation is σ, and it is 4, which implies that $\sigma^2 = 16$. This means the variance, σ^2, is 16.

So, X has a normal distribution with mean $\mu = 12$ and variance $\sigma^2 = 16$. Now, suppose you want to find the probability that $X \geq 20$. So, what we will do to find that is, whenever we want to find a probability with respect to something, we will use a transformation. For any normal distribution X with mean μ and variance σ^2, if we want to find the probability that $X \leq x$, or any probability you want to find, we discussed this earlier as well. Whenever you want to find any probability, we will use a transformation such that it is with respect to the standard normal distribution, allowing us to use this table to find the probability.



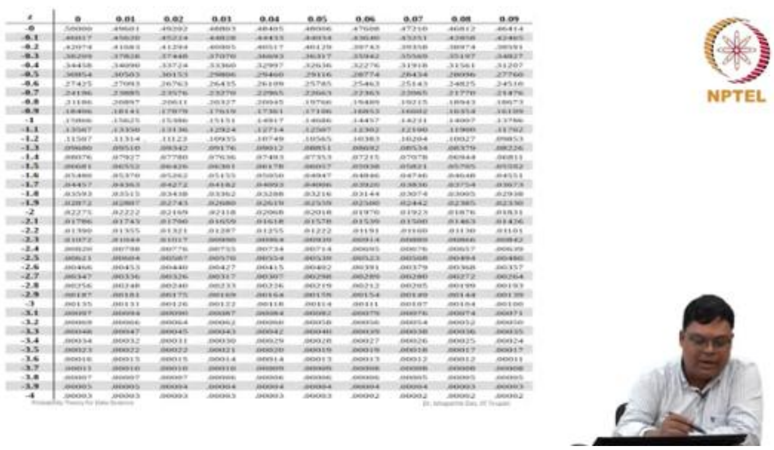
This is because we cannot easily perform the integration, so we need a numerical approach and computational techniques. Now, for $X \geq 20$, what will we do? We will subtract the mean, μ, which is 12, and divide by the standard deviation, σ, which is 4. So, for x, we will subtract 12 and divide by 4. This inequality will remain the same if we do the same thing on the right-hand side as well: (20 - 12) / 4.

So, this is equal to, now $Z = (X - \mu) / \sigma$, which in this case is $(X - 12) / 4$. So, this becomes $Z > (20 - 12)$, which is 8. Therefore, $8 / 4 = 2$. So, the probability is asking for $Z > 2$. Now, Z is a standard normal variate.

So, suppose this is the standard normal distribution, just approximately represented by this curve. This is 1, this is 2, and similarly, this is -1, and this is -2. Now, if you want to find $Z > 2$, the area, by symmetry, will be the same as $Z \leq 2$. Since it is a continuous distribution, this does not change. So, this is the same as writing ≥.

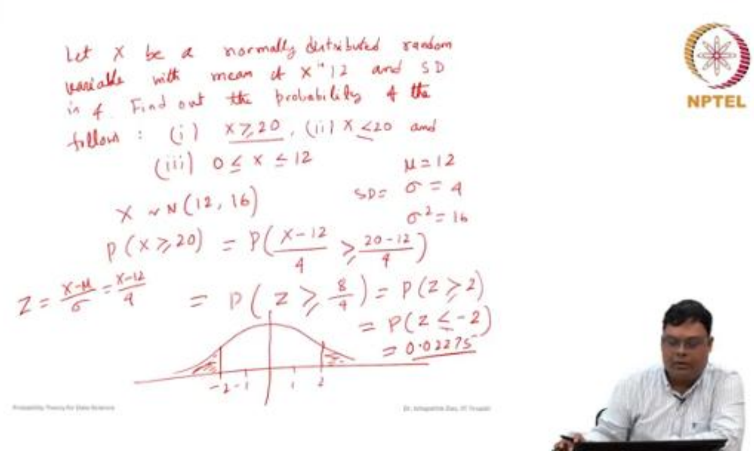
The inequality will be the same whether you write > or ≥. Now, sorry, this should be -2. So, this will be Z ≥ 2, and this area will be the same as Z ≤ -2. So, this is Z ≤ -2. We can find this probability using the table because Z ≥ -2.

So, this probability for negative values is given in the table. Where is this table? The table is here. We will find -2 here. This is -2. The probability is 0.02275. So, this probability is 0.02275.



Sorry, this is nothing but 0.02275. Let’s see this. For -2, the value is 0.02275. This is the probability. Now, the next question is: What is the probability that X ≤ 20?

So, we found that the probability X ≥ 20. The probability that X ≥ 20 is 0.02275. Now, for X ≤ 20, note that because it is a continuous random variable, this is the same as X > 20. If you write 'equal to' or '≥' at a specific point, the probability is 0. Now, the probability that X ≤ 20 will be 1 - P(X > 20), because it is the complement of this random variable.



So, this is 1 - 0.02275. So, whatever the value is. So, this is 0.97725. So, approximately, this is the value. You can check that this computation is correct, and then this will be the probability.

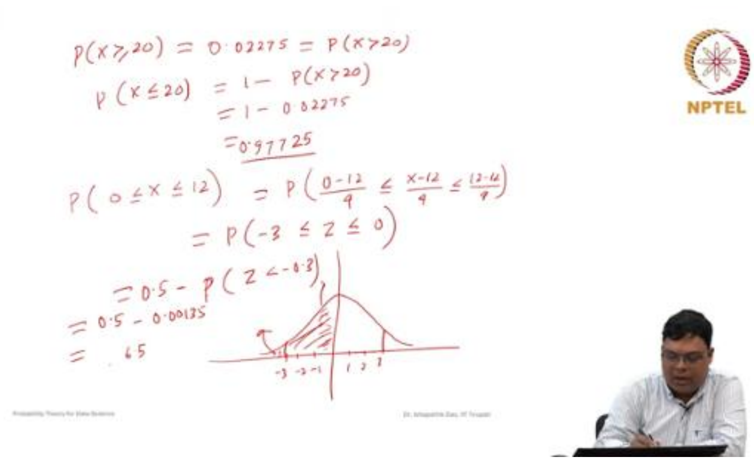
Next, it is asked: what is the probability that $X \geq 0$ and $X \leq 12$? The probability that $X \leq 12$ and $X \geq 0$. So, now again we have to apply this transformation: $Z = (X - 12) / 4$. So, we will see that this becomes $(0 - 12) / 4$. This is the transformation $(X - 12) / 4$, where $X \leq (12 - 12) / 4$.

So, basically, this probability is nothing but the probability that $(X - 12) / 4 \leq 0$. This is equal to $Z \leq 0$. Now, if you take the normal distribution curve—I'm just trying to draw it approximately for understanding—suppose this is 1, 2, 3 on the positive side and -1, -2, -3 on the negative side. So, then it's asking, sorry, not this part. It is asking for the probability that Z is between -3 and 0.

So, now we can find the probability for these values. This probability can be found from the table. So, this is nothing but 0.5, because the total probability here is 0.5. So, this is 0.5 - $P(Z \leq -0.3)$. The probability for $Z \leq -0.3$ is the same.

So, now this is the total 0.5 - this probability. Now, the probability, sorry, $Z \leq -0.3$ can be found from the table. For values up to -3, this probability is 0.00135. So, we subtract 0.00135 from 0.5. So, 0.5 - 0.00135 gives us the value.

So, this probability is 0.49865. Now, another problem is also asked: Find the value of x such that $P(X > x) = 0.24$. Find x' such that $P(X > x') = 0.24$. This probability is given, and you need to find the value of x'. So, how can we find that?



Since we don't know $P(X > x')$, we will apply the transformation again. So, this implies that $P((X - 12) / 4 > (x' - 12) / 4) = 0.24$. We can find this value, which is the same as $P(Z > z')$, and this is equal to 0.24. So, Z is the standard normal variate, $N(0, 1)$. $Z = (X - 12) / 4$, and $z' = (x' - 12) / 4$. So, if we can find z', then we can solve it and find the value of x'.

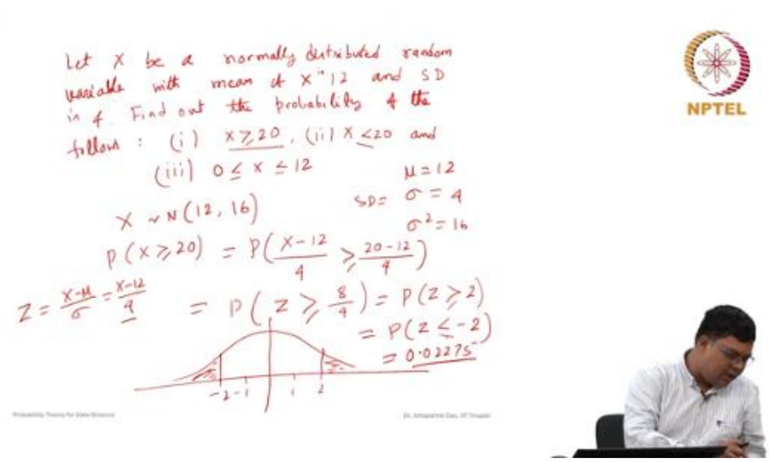
Now, how can we find z'? We want this probability to be 0.24. So, it's the normal distribution curve. Suppose this is the case. $Z > z'$, some value z'. This is greater than 0.24. By symmetry, the negative value of z' can be found. Here, this will also be 0.24. So, from the table, we have to find the probability. Actually, it's an inverse thing. You have to find the probability of 0.24 such that this z' corresponds to this value.

So, 0.24, where it is appearing, we have to find something like -something that corresponds to 0.24. So, 0.24 is approximately 0.23885, and then 0.23885 gives us 0.24. Since the table provides values up to a certain decimal place, we can consider this value of 0.23885 as close to 0.24. Alternatively, we can consider the value 0.24196, which is also close to 0.24. So, basically, it is approximately -0.71, which corresponds to the value.

So, this is close to 0.24, and the value here corresponds to z' = -0.71. Therefore, we get that z' = -0.71. So, similarly, this is nothing but equal to 0.71. So, we found that z' = 0.71, which implies (x' - 12) / 4 = 0.71, which implies x' = 12 + 4 * 0.71. So, we just need to compute it. So, it is 12 + 4 * 0.71.

So, 2.84, please check if this is correct. So, this is nothing but 14.84. So, this x' value is 14.84. So, this is the one question. We can do a similar type of question. Let's see if there's any other problem, similar to this one. Let X be a normal random variate with a mean of 30 and a standard deviation of 5. Let's do another problem. Let X be a normal random variate with a mean, $\mu = 30$ and a standard deviation, $\sigma = 5$. Find the probabilities that:

1. $P(26 \leq X \leq 40)$

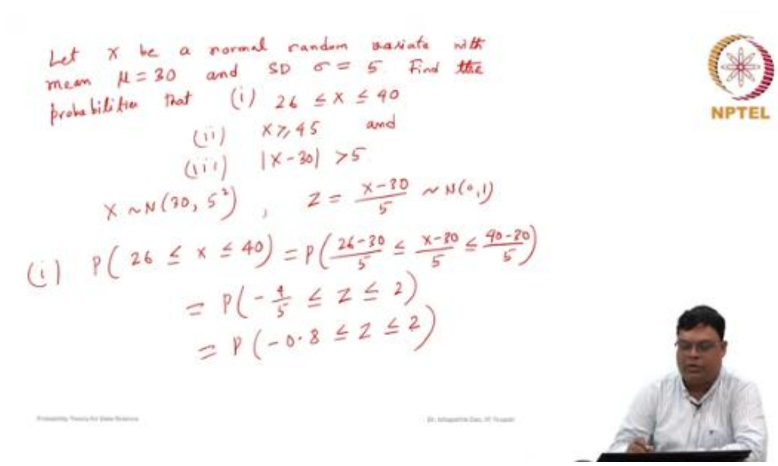


2) $X \geq 45$, and 3) $|X - 30| > 5$. So, this problem is similar to the previous one, and you can also solve it by trying it yourself. I am just writing down the solution here. In this case, we are given that $\mu = 30$ and $\sigma^2 = 5^2$.

We will take the transformation, where $Z = (X - 30) / 5$, which follows a standard normal distribution with $\mu = 0$ and $\sigma^2 = 1$. Using this transformation, and with the help of the table, we will calculate the probabilities. First, the probability we need to find is for question 1: $P(26 \leq X \leq 40)$.

This is equal to $P((26 - 30) / 5 \leq (X - 30) / 5 \leq (40 - 30) / 5)$. This value shows that $(X - 30) / 5$ is the random variable Z.

So, this is nothing but $P(-4 / 5 \leq Z \leq 10 / 5)$, which simplifies to $P(-0.8 \leq Z \leq 2)$. So, how will we find this value? So, finally, we have a normal distribution.



Suppose we are considering the standard normal variate. Now, we want to find $P(Z \geq -0.8$ and $Z \leq 2)$. So, suppose this is -0.8 and this is 1, 1, 2. The area they want is $P(Z \leq 2$ and $Z \geq -0.8)$. You need to find $P(-0.8 \leq Z \leq 2)$.

It is not exactly coming out. So, $P(Z \leq 2)$. How can we find that? First of all, the total area is 1. We can find the probability of this area.

Once we have this probability, we can subtract it from 0.5, and then we will get this area. Similarly, we can find this area by symmetry. If 2 is on one side, -2 will be on the other side. So, the area for -2 will be the same as the area for 2. So, then we will take half of the value and subtract it to get this area.

I think we need to find it two times, which is just this. First, we need to find $P(Z \leq -0.8)$

from the table. Since it's a negative number, we will refer to the table. For 0.8, the value from the table is 0.21186. So, $P(Z \leq -0.8) = 0.21186$.

So, we want the probability, not just that, but this area. This area is actually $P(Z \geq -0.8$ and $Z \leq 0)$. So, this area will be 0.5 - 0.21186. This area corresponds to that. So, for this area, we just need to subtract this. This is 4, then 1, then 8, 8 and finally 2. So, this is nothing but 0.28814.