INDIAN INSTITUTE OF TECHNOLOGY MADRAS

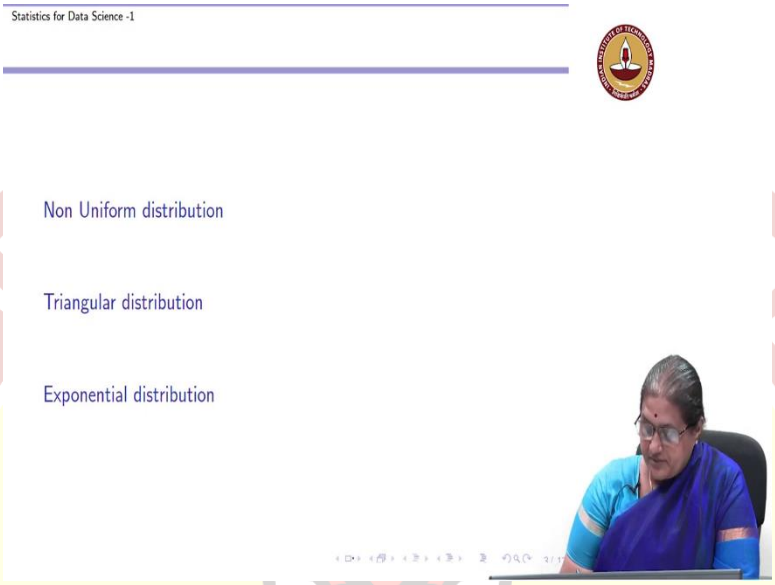
सिद्धिर्भवति कर्मजा

IIT Madras

ONLINE DEGREE

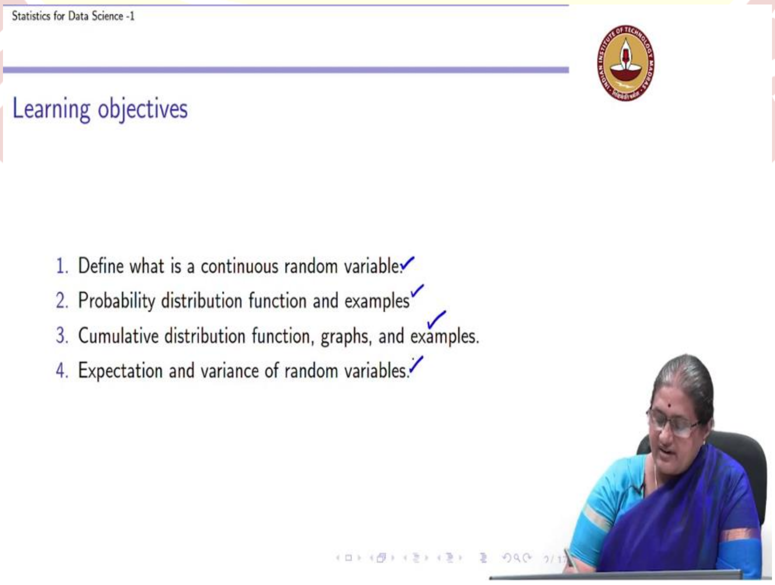
Statistics for Data Science- 1
Professor. Usha Mohan
Department of Management Studies
Indian Institute of Technology, Madras
Continuous random variable - Non uniform and triangular distribution

(Refer Slide Time: 0:18)



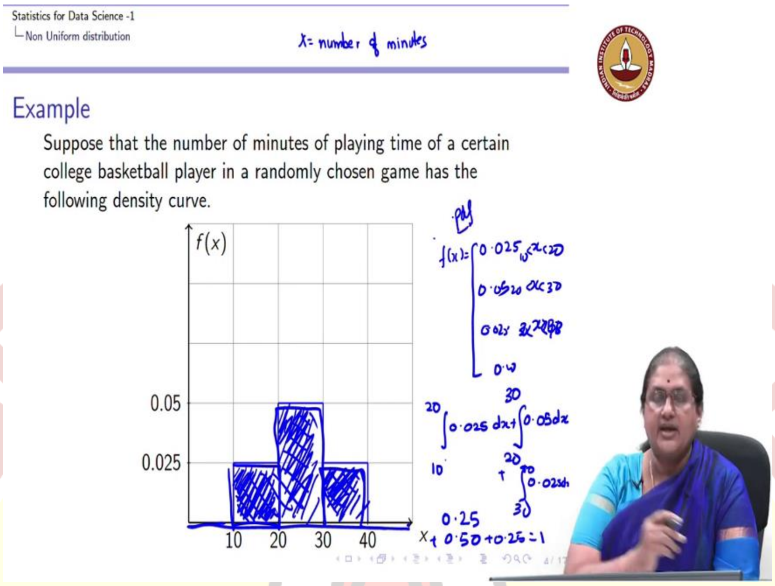
Now, we are going to look at another two or three naturally arising distributions. Namely, I will just, we will first look at something which is not uniform. We have looked at uniform distribution, so we will look at a non-uniform distribution, then we will just look at a triangular distribution, and then we will spend some time understanding what is a exponential distribution.

(Refer Slide Time: 0:48)



And as in every case we will find out what is the random variable the PDF, the CDF and we will get hold of the application.

(Refer Slide Time: 0:59)

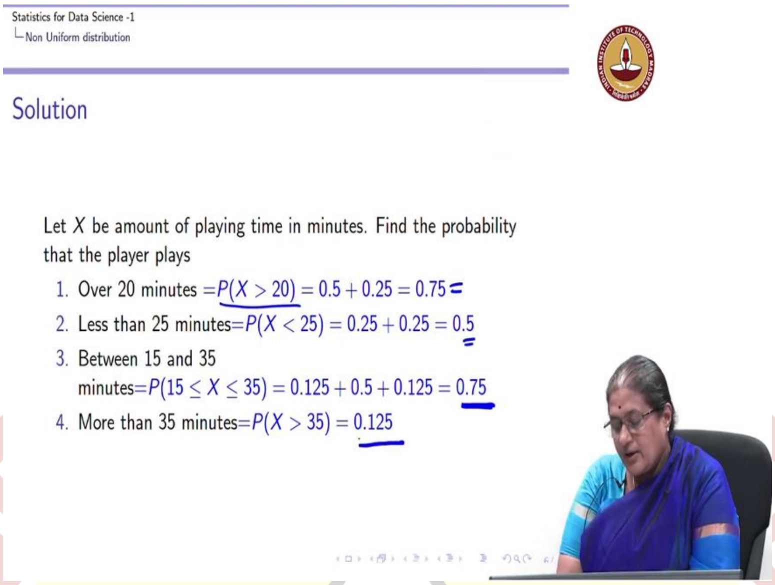


So, now let us look at the case with the number of minutes. Again, my random variable x is measured it is time measurement. And how am I measuring it? I am measuring it, as number of minutes of playing time, but here it is not a uniform random variable, it is following the following density curve. What it says is $f(x) = 0.025\ 10 < x < 20$, that is what is this portion.

$f(x) = 0.05\ 20 < x < 30$ which is this portion and for $30 < x < 40$ $f(x) = 0.025$ which is this portion, and it is 0 otherwise. So, as in every case, the first ratification is, yes, it is greater or equal to 0, that is clear because it is not going below the x axis. The second thing is over the entire range, what is f value? What does it add up to be?

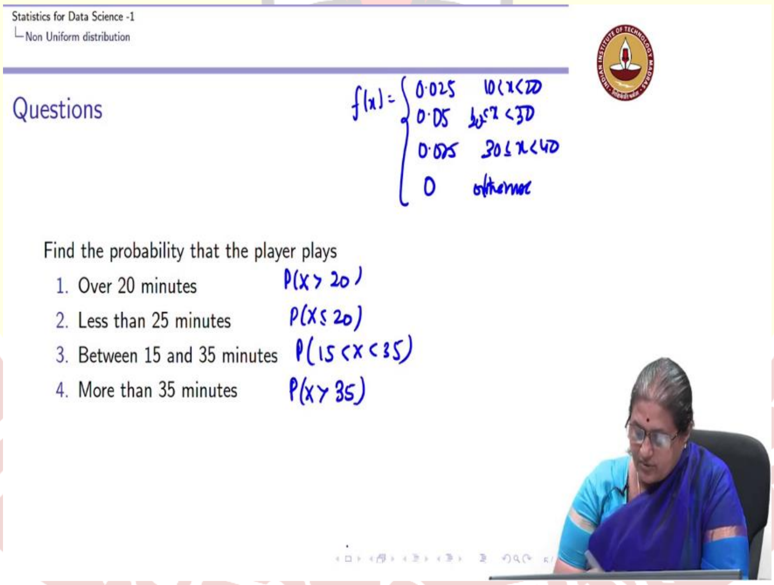
So, you can see that this is going to be $\int_{10}^{20} 0.025dx + \int_{20}^{30} 0.05dx + \int_{30}^{40} 0.025dx$, that is the total area under the curve. Now, this is nothing but the area of this rectangle plus the area of this rectangle, plus the area of this rectangle. Now, we can see that the area of the first rectangle is what we have here, which is going to be 0.25, 0.025*10. The area of this rectangle is 0.5.

(Refer Slide Time: 3:19)



And the area of this rectangle is again 0.25 giving us that the total area under the curve is equal to 1 and hence, this satisfies up PDF. So, the first verification is done.

(Refer Slide Time: 3:40)



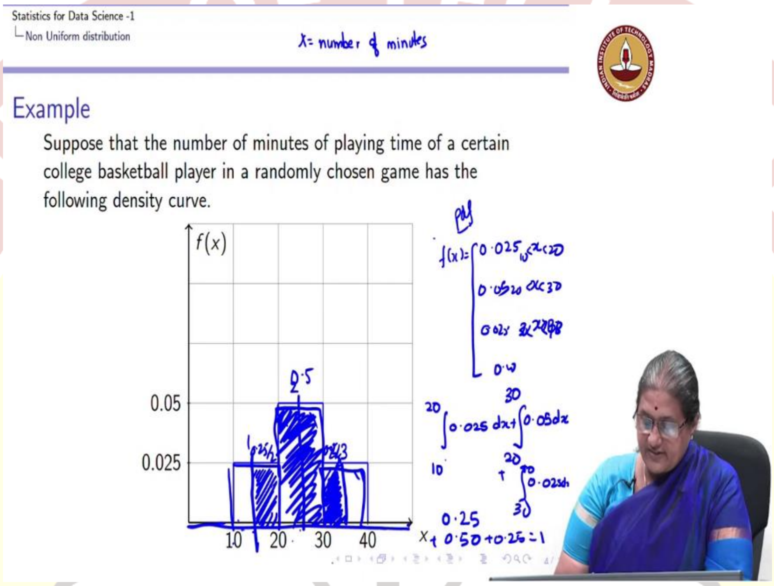
Now, the question that is asked is, I want to find that. So, I am given X now, I do not have it as a known distribution, but I can find know what is my PDF and my PDF is

$$f(x) = \begin{cases} 0.025 \ 10 < x < 20 \\ 0.05 \ 20 \le x < 30 \\ 0.025 \ 30 \le x < 40' \\ 0 \ otherwise \end{cases}$$

this is how I can define my PDF.

And I am asking now, what is the chance that X > 20, $P(X \leq 25)$, 15 < X < 35, and $P(X \geq 35)$. These are the questions you are asking. So, let us go back to the problem and you can see that here, probability x is greater than 20. Now, what is x is greater than 20? X is greater than 20 is going to be just this area. I already know this, so if I am calling my area 1, 2, and 3 probability X is greater than 20 is area 2 + area 3, area 2 is 0.5, area 3 is 0.25, so $P(X > 20) = 0.5 + 0.25 = 0.75$.

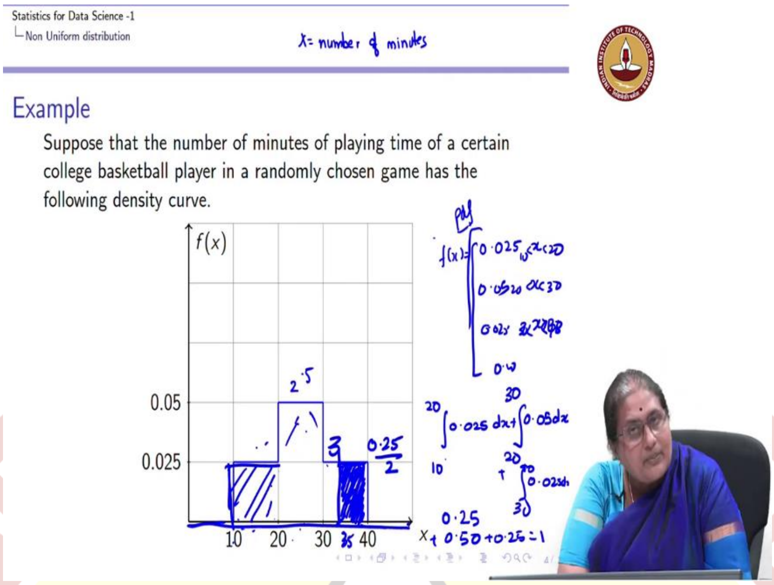
(Refer Slide Time: 5:07)



$P(X \leq 25)$, Again, 25, 5 lies between this, so if area of 2 is 0.5, the half of that area is 0.25 X is less than 25 is area 1 + half of that area 1 is 0.25 this is also 0.25, making the area total of this area to be point 0.25, + 0.25=0. Between 15 and 35. So again, you can see that 15 and 35, between 15 and 35, 15 lies between 10 and 20. Hence, this area between 10 and 20 is going to be 0.25 divided by 2.

Similarly, 35 lies between 30 and 40, so this area is again, 0.25/2. And this is going to be already we know this area is 0.5. So, this is a 0.5, this is 0.5/2 and this is 0.5 by 0.25 by 2, hence, the total area is 0.25+ 0.5= 0.75.

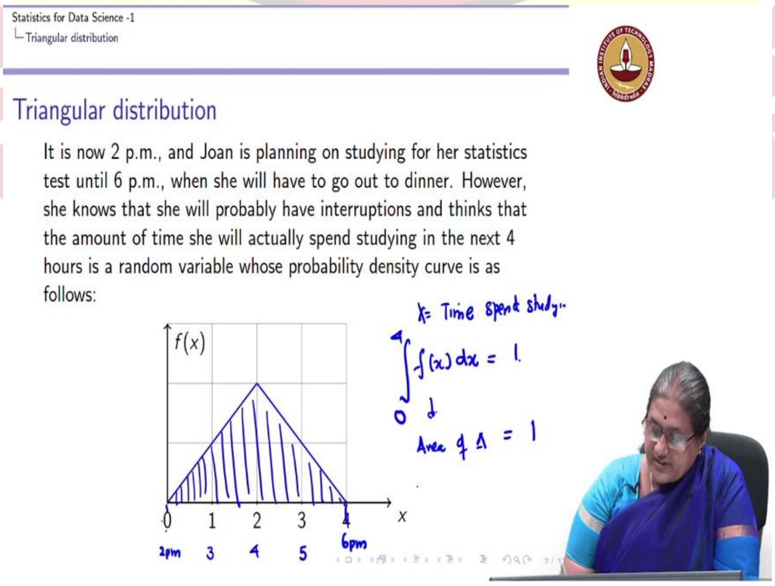
(Refer Slide Time: 6:34)



Now greater than 35 minutes, that is the last thing which we need to see. Again, we go back here. I know, 35 lies here, which is exactly at half greater than 35 minutes is going to be this area. I know, that the total area of 3 is 0.25, so the half of this area is going to be 0.25/2=0.125.

So now we, what we have established is I do not need a uniform distribution, but I need to understand what is my density? Given this problem, first, I find out what is my actual whether it is a probability density function and I apply the formula we have got to find out what was the probabilities that were being asked just based on the problem.

(Refer Slide Time: 7:38)

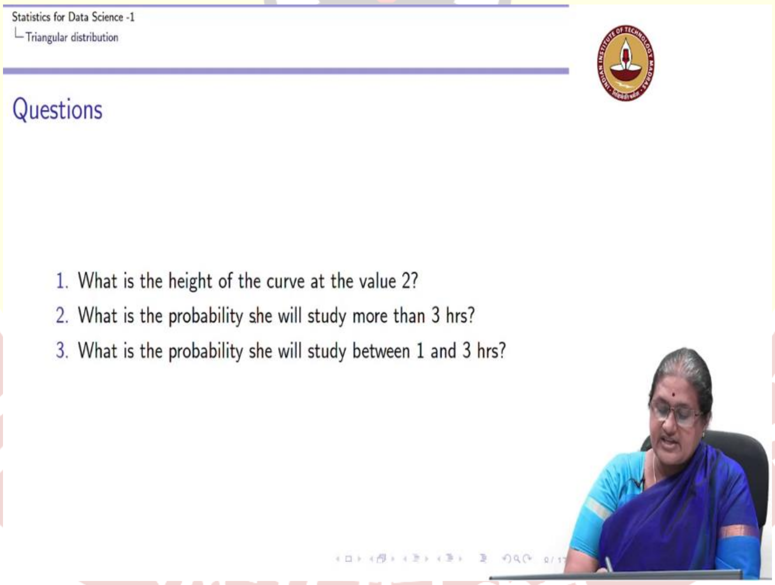


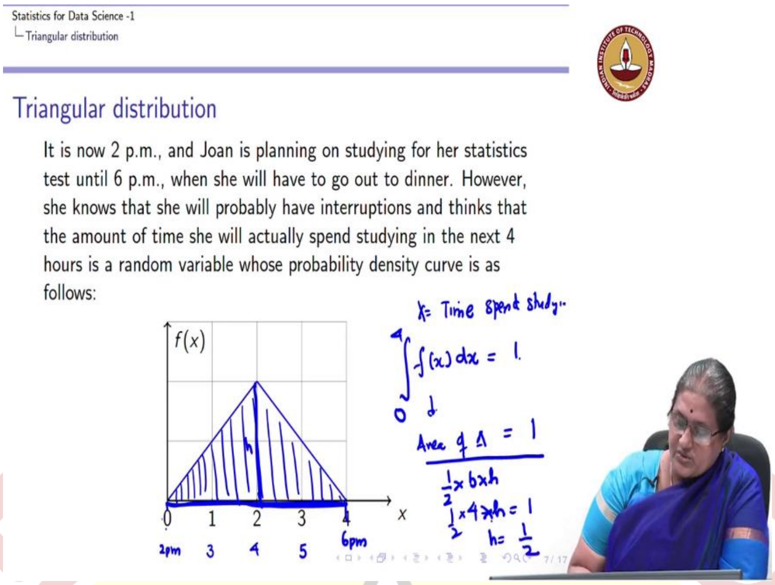
Now let us look at another distribution. It is now 2 p.m. and Joan is planning to study for her statistics test until 6 p.m. So, if I set 2 p.m. to be my origin and this is my 6 p.m, I note, this would be 3 p.m this would be 4 p.m, and this would be 5 p.m, and each of this is 1 hour. But however, she knows that she will probably have interruptions and she thinks that the amount of time she actually spends would be a random variable with the following probability distribution curve.

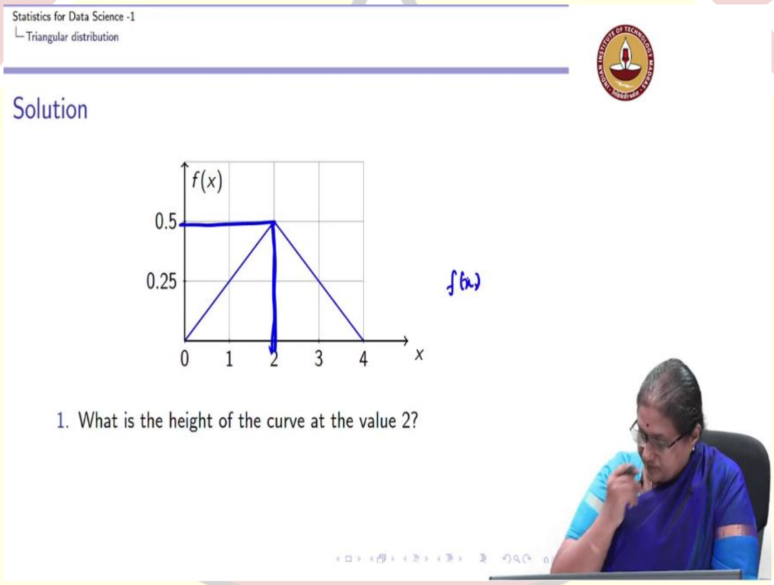
So, if this is a probability distribution curve, and this is 1 to 4, this is X is the time. Again, we are measuring time and this is the time spent studying. So, I am going to spend anything between 0 and 4 hours and this is given by the following density. For this to be a density function, so $f(x)$ is a density function. I know $\int_0^4 f(x)dx = 1$

Now this density function has a triangular form, and I know that area under this triangle should be equal to 1. So, $\int_0^4 f(x)dx$ is nothing but area of the triangle, and I know that area of the triangle should be equal to 1. So, what is the density function?

(Refer Slide Time: 9:33)

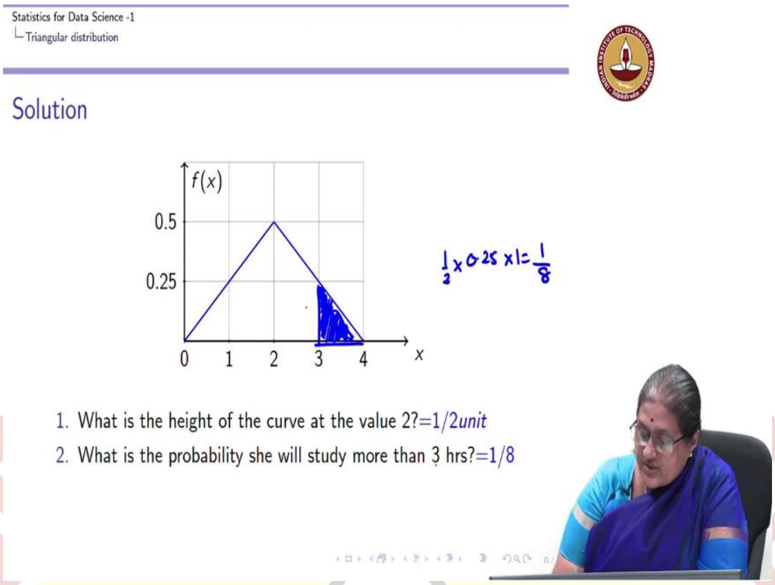






If I need to know what is the density function, the first question we are asking is what is the height of the curve at values 2. Recall the area of a triangle is 1 and I also know the area of a triangle $=\frac{1}{2}b * h$. So, the breadth (b)= 4. I need to know what is my height, and I know that the height is 1/2, because $\frac{1}{2}4 * h = 1$, I can establish that the height of the curve at where you 2 equal to 0.5, because the area of the triangle has to be equal to 1, because we have a density function, $f(x)$.

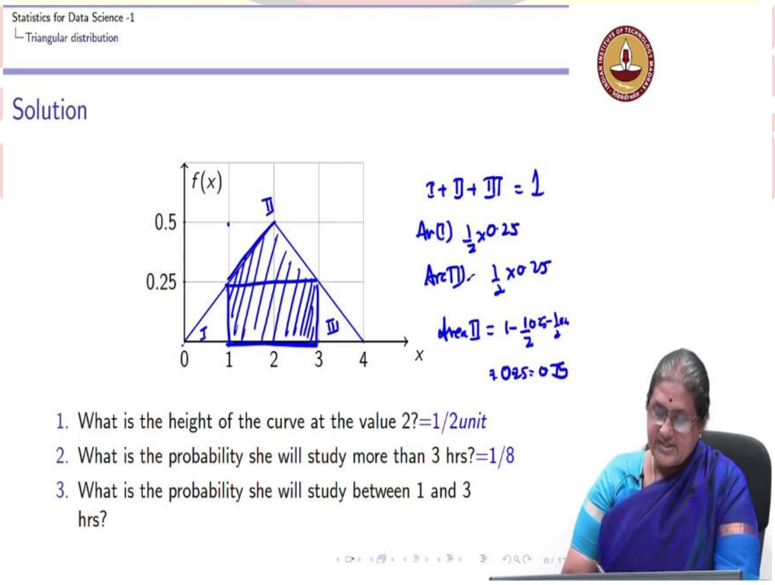
(Refer Slide Time: 10:34)



What is the probability that she will study for more than 3 hours? So, the probability she will study for more than 3 hours is going to be this shaded region. So, there are a couple of ways we can find this area. So, this height is 0.5, so this height is 0.25, this breadth is 1 unit. So, it is just this area. So, the probability she is going to study for greater than 3 hours is 1/8.

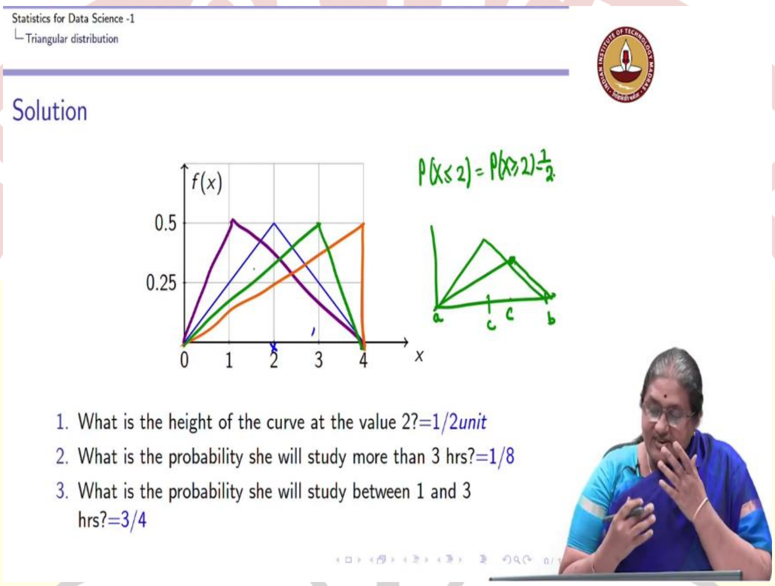
Another way to find out is 1/8 because this is 0.25 *1 *1/2, which is 1/8 of an hour. The probability she is going to study for greater than 3 hours is going to be 1 minus the probability she will study for less than 3 hours. Again, less than 3 hours, you can check that that would be the same as 1/8.

(Refer Slide Time: 11:51)



Now, the third probability is, she will study between 1 and 3 hours. So, that area is going to be this, which is going to be the shaded region here. There are two ways to find this out. Again, I can do 1, 2, 3, I can break this down and do it or I know the area of 1, the area of 2, which is this area, and 3 area of 1 plus area of 2 plus area of 3 is equal to 1. Area of 1 is again, 0 to 1.25, which is ½*0.25. Area of 3 is again half into 0.25, so I can find out area of 2 is 1 – ½* 0.25 – ½* 0.25 = 1 - 0.25 = 0.75.

(Refer Slide Time: 12:57)

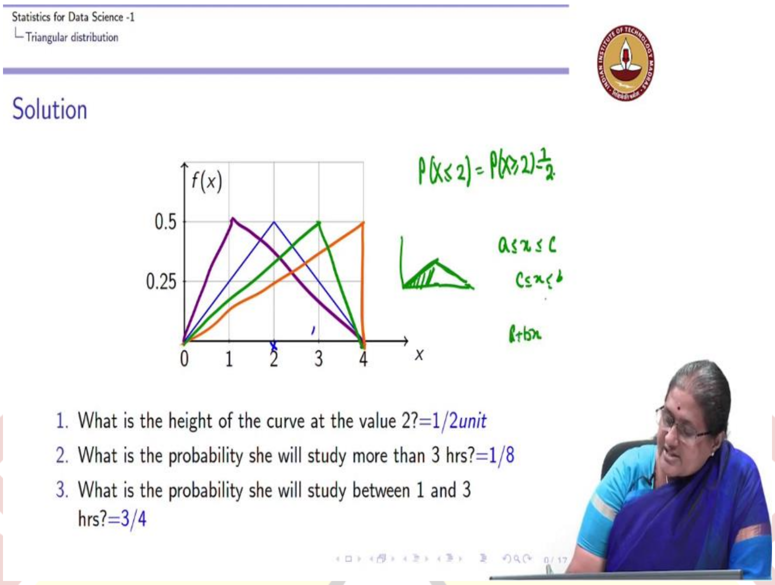


Hence the area of 2 is 3/4, which is 0.75. So here, you can extend this that one way to look at a triangular distribution is that here I had a symmetry, but I do not need to have a symmetry, I could have a distribution where the triangle is of this kind also or it could be of this kind or it could be of this kind.

What do we need to ensure is, the area under these curves are the same. In this case, the blue line it was a symmetric distribution in the sense that the $P(X \leq 2) = P(X \geq 2) = 0.5$ because the area under the blue triangle has to be equal to 1.

It need not be the case, it could be again, this is with respect to this. So, a triangular distribution basically, is a distribution which has endpoint, and it has a point c, with the slope between a and c to be 1 and the slope between c and b to be another, if c is a midpoint and this slope is equal to the negative of the slope, this is my blue triangle.

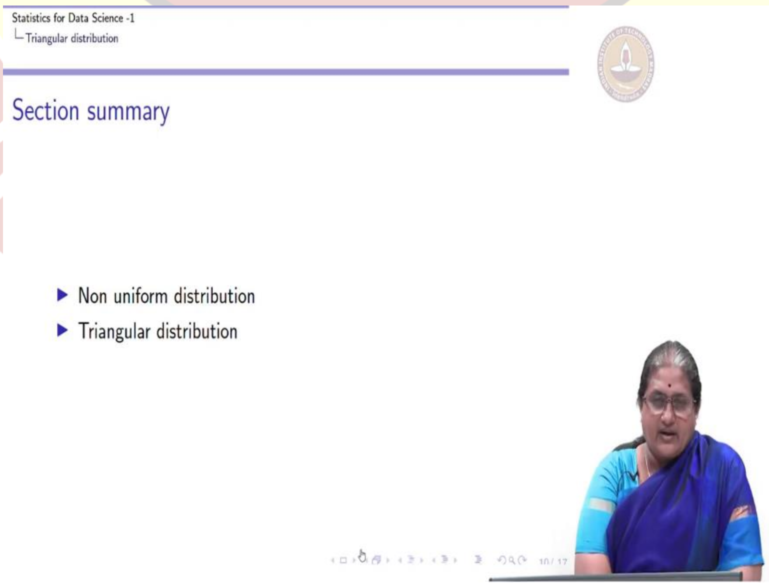
(Refer Slide Time: 15:07)



So, in general, we can write down the density function of a triangular distribution using this notion that is for $a \leq X \leq c$, it would be a upward sloping curve and for $c \leq X \leq b$ is going to be a downward sloping line, we need to ensure. So, when I am giving the equation of a line, I will get $a + bx$ or $y = mx + c$. So, this would be a positive slope and this would be a negative slope.

So, we demonstrated this at a foundational level just using a graph. So, in advanced course, you will actually get mathematical expression for triangular distribution.

(Refer Slide Time: 16:00)



So, in summary, what we have seen here is without going into the details of the mathematical expressions, we saw that they could arise other continuous distributions and we saw the example of a non-uniform distribution and a triangular distribution where the density function was a triangle.