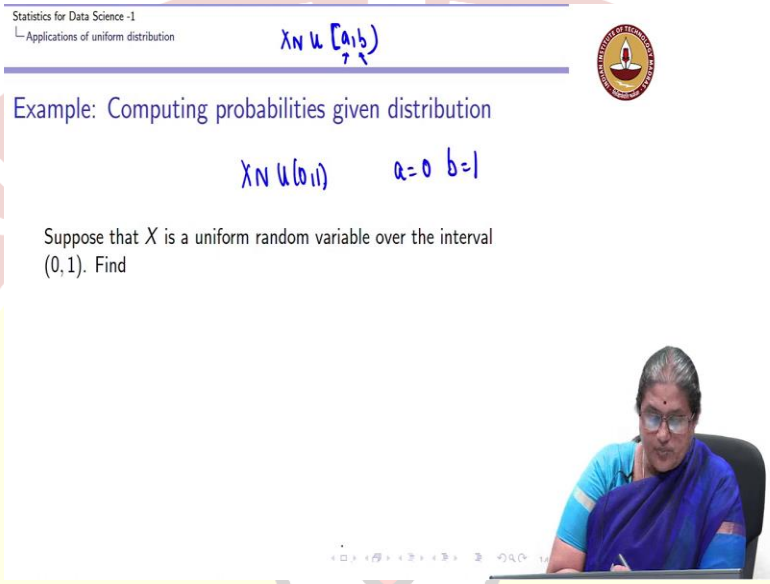
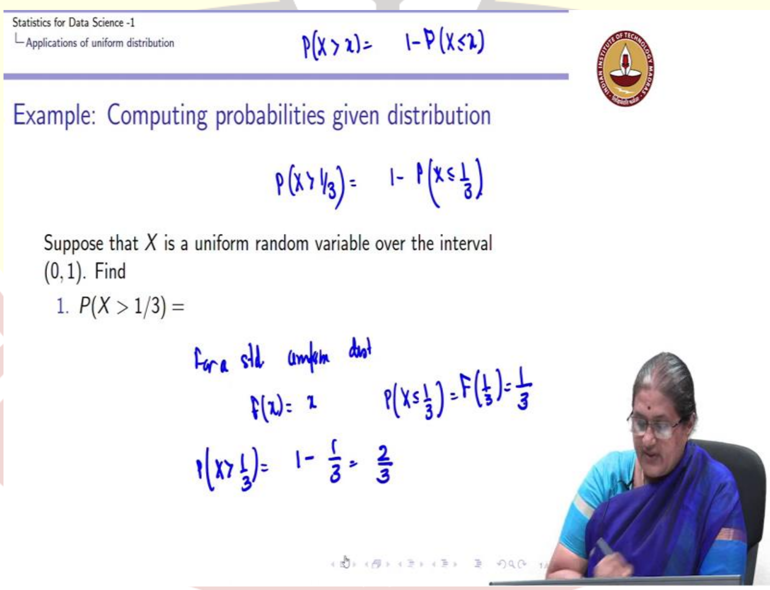
IIT Madras

ONLINE DEGREE

Statistics for Data Science – 1
Professor Usha Mohan
Department of Management Studies
Indian Institute of Technology, Madras
Continuous random variable - Uniform distribution: applications

(Refer Slide Time: 0:14)





Let us apply these concepts to compute probability given the distributional parameters. By parameters, I mean that when I say X is a uniform distribution between a and b, a and b are specified as the parameters of my distribution. So, suppose X is a standard uniform random variable over the interval (0, 1). So, X is uniform 0, 1, my a is 0, b is 1, what is probability? So, these are what I want to find out $P(X > \frac{1}{3})$, we know that $P(X > x) = 1 - P(X \leq x)$

That is because I know that the probability is going to be both

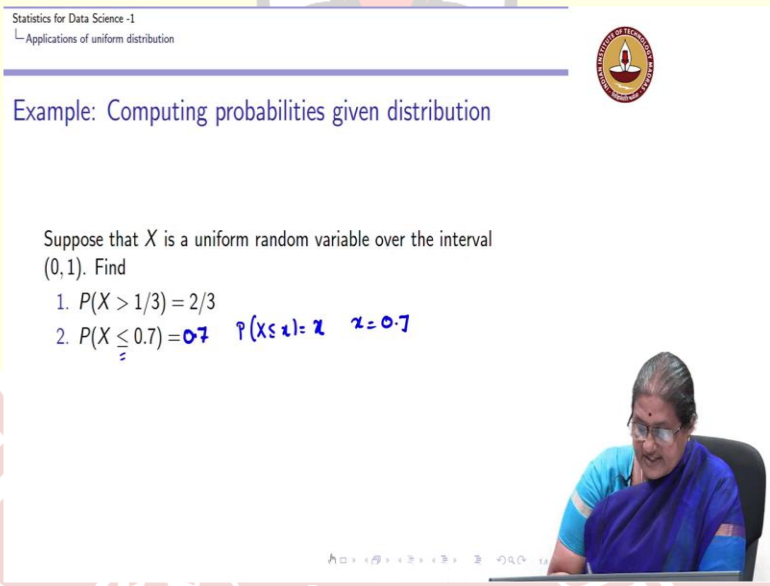
$$P(X > x) + P(X \leq x) = 1$$

$$P\left(X > \frac{1}{3}\right) = 1 - P(X \leq \frac{1}{3})$$

So, this is 1 - your probability X is less than or equal to 1 by 3. And I know that for a standard uniform distribution, $f(x)$ is equal to x, this is what we saw. And hence I have a probability X is less than or equal to 1 by 3, which is equal to $f(1/3)$, which is equal to 1 by 3,

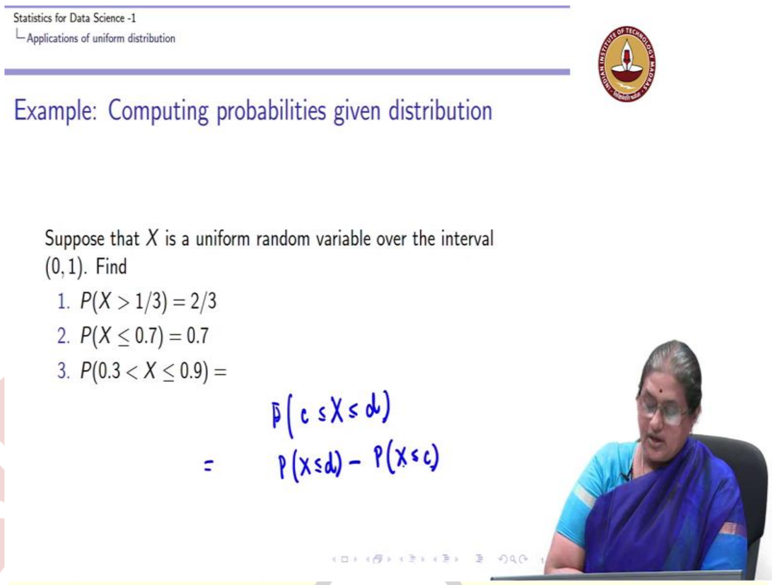
$$P\left(X > \frac{1}{3}\right) = 1 - \frac{1}{3} = \frac{2}{3}$$

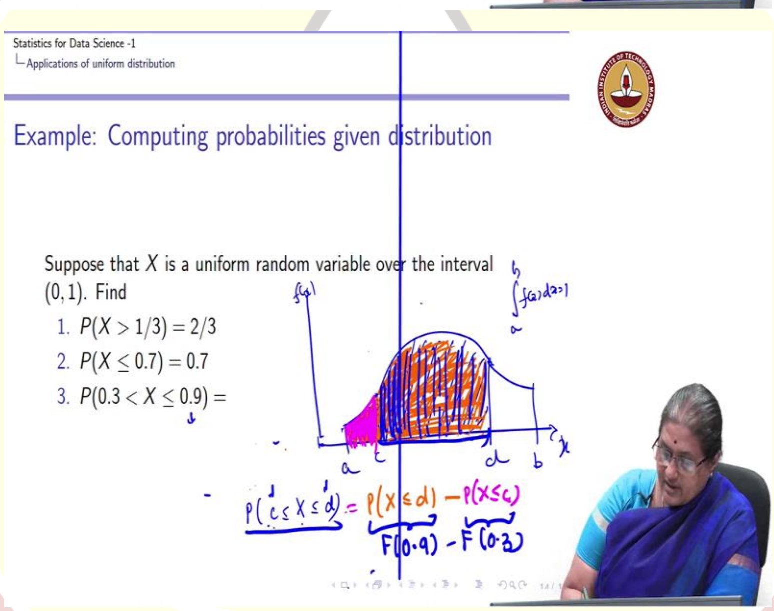
(Refer Slide Time: 2:05)



So, I have X is greater than 1 by 3 is 2 by 3, $P(X \leq 0.7)$ is trivial $P(X \leq x) = x$. Here, my small x is 0.7. So, $P(X \leq 0.7)$ is 0.7.

(Refer Slide Time: 2:26)





Now,

$$P(c \leq X \leq d) = P(X \leq d) - P(X \leq c)$$

Now, why is this true? So, let us look at the understanding we have, I have this as my density function, which is defined between c and d is what I want to know. So, this I have a to b, so, a to b is going to be my entire probability and I know.

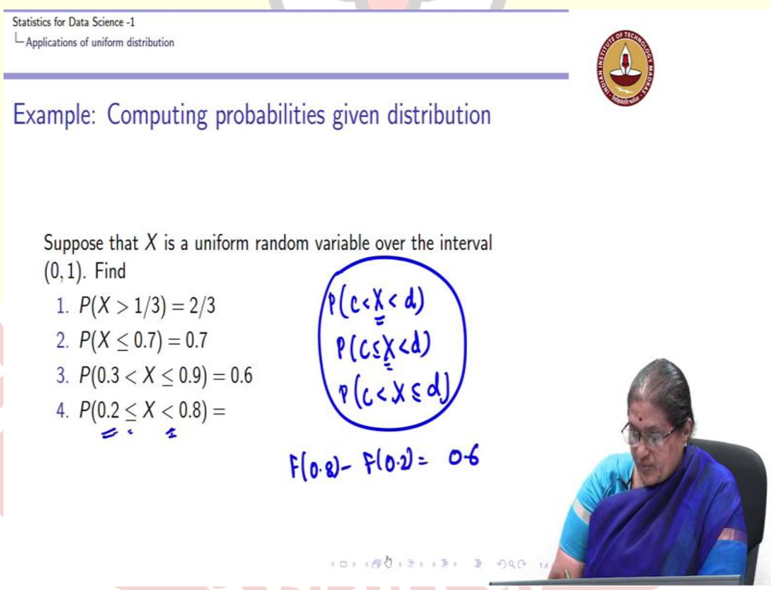
$$\int_a^b f(x)dx = 1$$

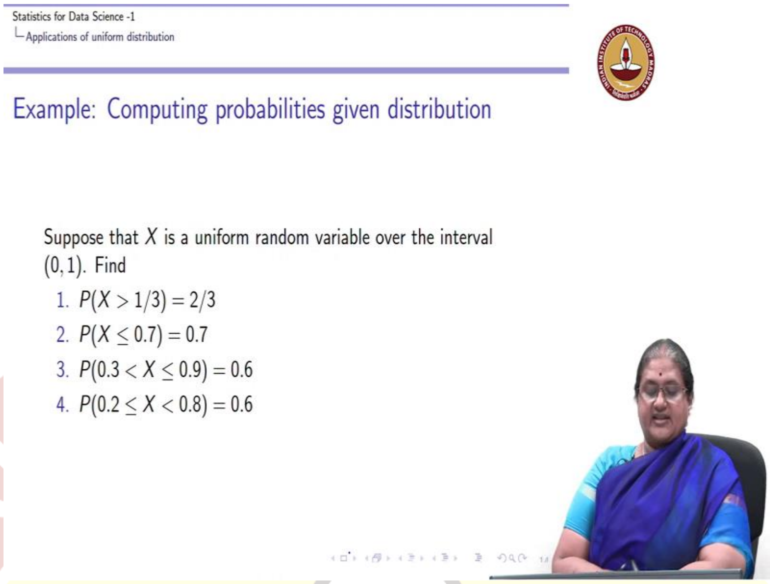
So, given any density function now, if I am interested in knowing $P(c \leq X \leq d)$, what effectively asking is this blue area that is what I want to know. Now, if I want to know the blue area, that blue area is the same as I can find out what is this entire area, the orange area - the purple area? The blue area would be the orange area - the purple area. The orange area is nothing but, the orange area is nothing, but $P(X \leq d)$. That is my orange area.

If I look at the orange area, the orange area is going to be this orange shaded region is $P(X \leq d)$. Because this is my d, the entire orange shaded region is $P(X \leq d)$. The purple shaded region, the only the purple shaded region is going to be $P(X \leq c)$ and the difference in these two areas is what I originally had as my blue shaded region. And that is what we wanted the blue shaded region is what we wanted was the area enclosed between c and d.

So, if I am interested in going for any two constants, c and d, $P(c \leq X \leq d)$, I find out $P(X \leq d) - P(X \leq c)$. Now my d here is 0.9. So, this would be $f\ (0.9) -\ 0.3$, which I am going to get 0.6.

(Refer Slide Time: 5:42)



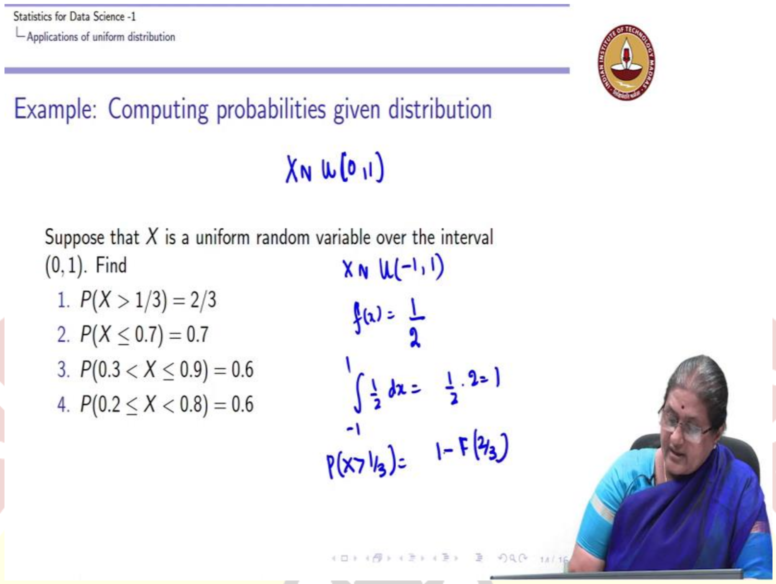


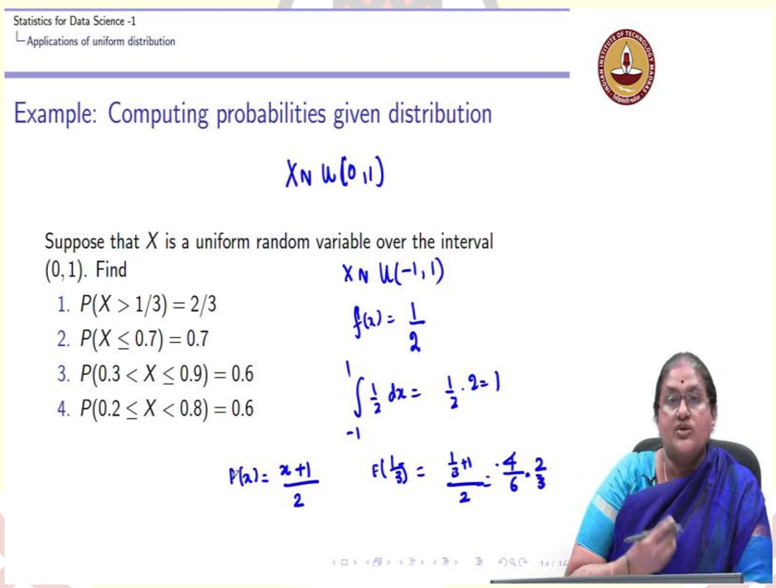
So, similarly, if I am looking at 0.2 is less than or equal to point, so what I want to tell here is

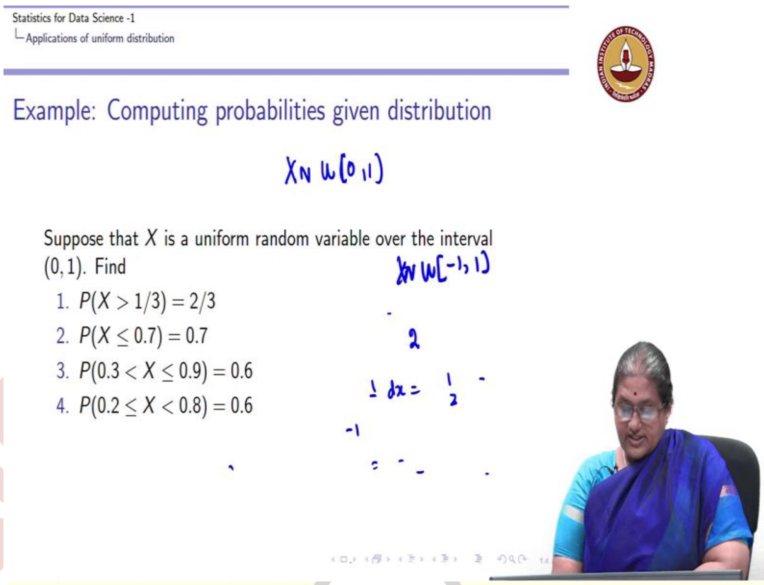
$$P(c \leq X \leq d) = P(c \leq X < d) = P(c < X \leq d)$$

Again, because X is a continuous random variable, these probabilities are the same, because in a uniform distribution, this is how I have defined it. So, what is $P(0.2 \leq X \leq 0.8)$. Again, $f(0.8) - f(0.2)$ which I again get is 0.6. And that is displayed on the screen.

(Refer Slide Time: 6:36)







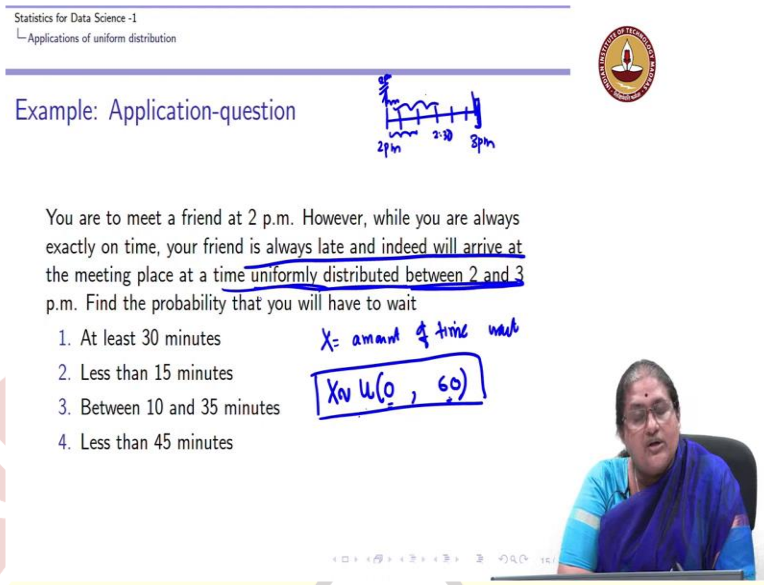
So, given, so this was when X is a uniform between 0 and 1, I can do the same exercise, if it was a uniform between - 1 and 1, what would happen, suppose I have a X, which is uniformly distributed between - 1 and 1, my $f(x)$ again is 1 by 2 and I can verify that

$$\int_{-1}^{1} \frac{1}{2} dx = \frac{1}{2} * 2 = 1$$

So, in which case, $P\left(X > \frac{1}{3}\right) = 1 - f(\frac{2}{3})$

Now, what is your $f(x)$ here? Your f (x) is going to be your $\frac{x+1}{2}$, that is what your $f\ (x)$ is going to be. So, $f(1/3)$ is going to be your $\frac{\frac{1}{3}+1}{2}$, that is what is your 4 by 6, which is my 2/3 , and 1 - 2 by 3 is 1 by 3. So, that is how I get a $f(X \geq \frac{1}{3})$. For the other problems, I just leave it as an exercise for you all to verify, what are the problems when x is equal to a uniform - 1 to 1.

(Refer Slide Time: 8:08)



Now, let us look at an application of a uniform distribution. Suppose I am asked to meet a friend at 2 pm. I am always on time, but my friend is always late. And he can arrive at a meeting place any time. And the chance that he will arrive at any time between 2 pm and 3 pm is the same.

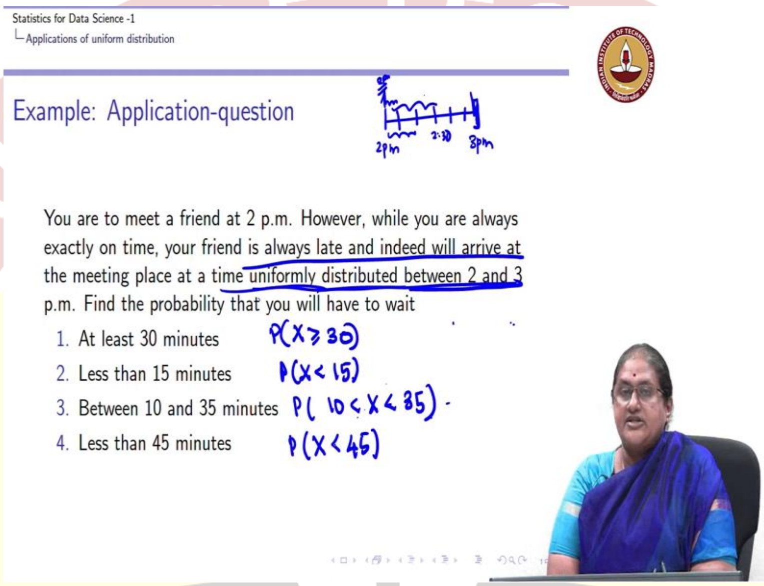
In other words, if X is the amount of time I am going to wait for my friend, if he arrives at 2 pm, because I have arrived at 2 pm, I do not wait. If he arrives at 2:10, I will have to wait for 10 minutes. If he arrives at 2:20. I wait for 20 minutes. If he arrives at 2:30, I will have to wait for 30 minutes, and 2:40, 2:50. And if he arrives at 3, I wait for 60 minutes.

So, if X is the amount of time I am going to wait, I know time is being measured. The amount of time I am going to wait is uniformly distributed whether I am waiting for 10 minutes or 20 minutes or 30 minutes. It is uniformly distributed in the interval 2 to 3. So, what is the first thing so I either wait the time I am going to so the total time I wait, the minimum is 0. The maximum is 60, I do not wait for anything when he comes at 2. I wait for 60 minutes when he comes at 3:00. And that so I can look at the time of waiting to be a uniformly distributed random variable between 0 minutes and 60 minutes.

Now, this is the key thing to identify in modelling problems. So, I am saying that this and this X belong, X is distributed, 0 to 60 comes from this information, that is this person will arrive at a time that is uniformly distributed between 2 pm and 3 pm. Hence, my waiting time, the time I wait is either 0 minutes or the maximum time I wait 60 minutes, hence, the time I wait is going to be uniformly distributed between 0 minutes and 60 minutes.

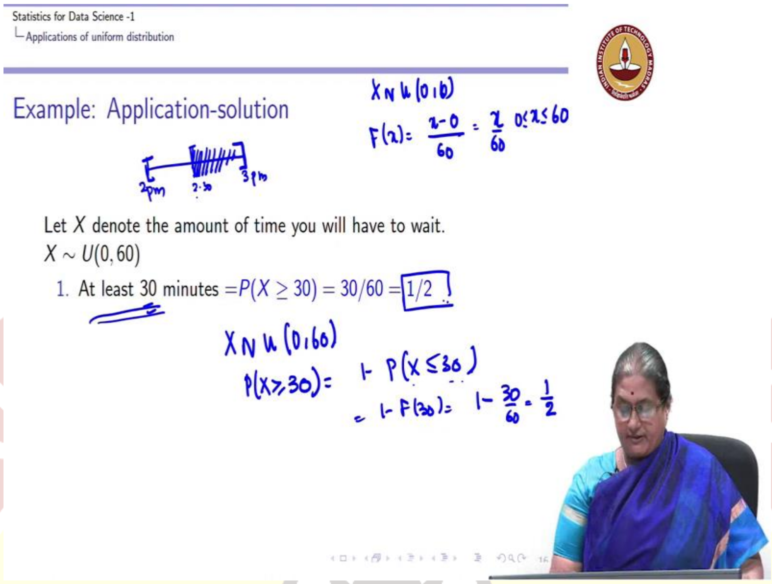
Why is it a continuous random variable because I am measuring time and you also recall that whenever I am measuring and not counting, I will discuss, or I will refer to it as a continuous random variable. So, the question now first I have understood that exercise a uniform between 0 and 60. So, let us translate these questions first in terms of X.

(Refer Slide Time: 11:23)



So, the first thing is, X is my waiting time, it is at least 30 minutes, I have to find out $P(X \geq 30)$, $P(X < 15)$, $P(10 < X < 35)$. And less than 45 is going $P(X < 45)$. I again recall less than 45 and less than or equal to 45 does not make a difference when you are referring to continuous random variables.

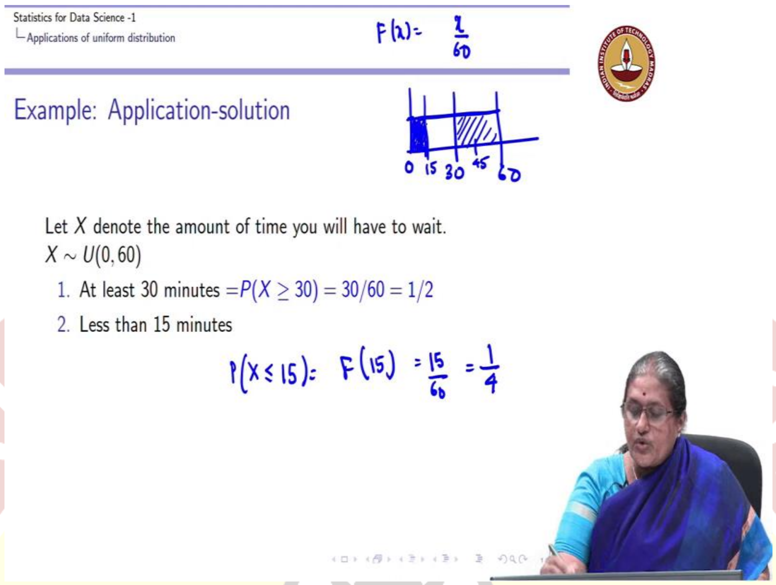
(Refer Slide Time: 12:03)



So, let us look at 1 by 1, $P(X \geq 30)$, Again, X is a uniform 0 to 60. I know $P(X \geq 30)$ is the same as 1 - $P(X < 30)$. And for X, which is a uniform 0 to 60, $f(x) = \frac{x-0}{60} = \frac{x}{60}$, for x lies between 0 and 60. Now, 30 lies between 0 and 60. So, $1 - f\ (30)$, is going to 1- $\frac{30}{60}$, which is going to be equal to half.

So, the chance that a person waits for at least 30 minutes is 1 by 2, which seems reasonable. Suppose I know a person is going to come in any time between 2 pm and 3 pm. And the chance that he arrives at any point of time in this interval is going to, is the same, then the way that you are going to wait for at least 30 minutes is going to be that he arise between 2:30 to 3, I know he is going to definitely arrive between 2 and 3. So, the chance he is going to arrive in this interval is a 50 percent chance and that is given by half.

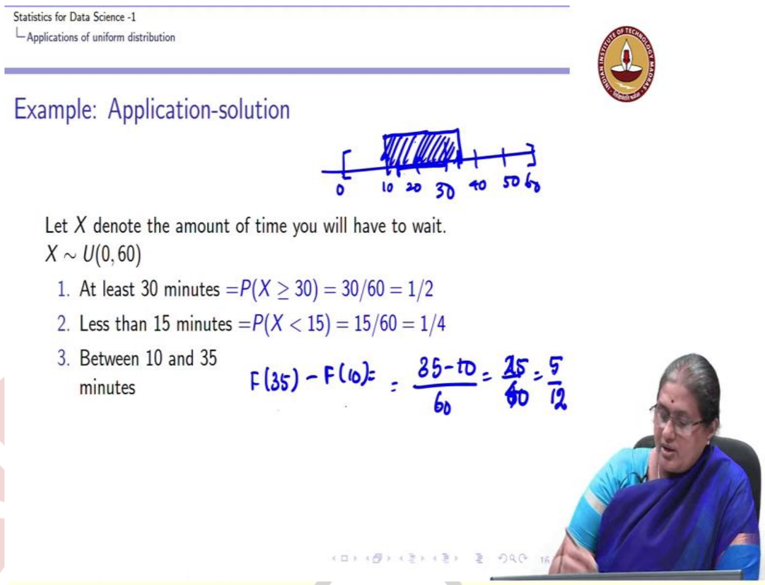
(Refer Slide Time: 13:52)



Less than 15 minutes again, let us look at it. So now again, what would be less than 15 minutes greater than 30 minutes so, this is 0 minute, this is 30 minutes, this is 60 minutes, this is going to be 15 minutes, this is going to be 45 minutes. So again, one way is greater than 30, we know was half, so less than 15 is going to be this. So, what do you think would be this, you can see that greater than 30 is half, this is one-fourth of this total area?

So, you can also verify that $P(X < 15)$, is f (15). Again, recall $F(x)$ is x by 60, which is going to be 15 by 60, which is 1 by 4. Again, if the area of this entire rectangle is 1, I know this area, the shaded blue region is going to be 1 by 4 or there is a 25 percent chance that you will wait for less than 15 minutes.

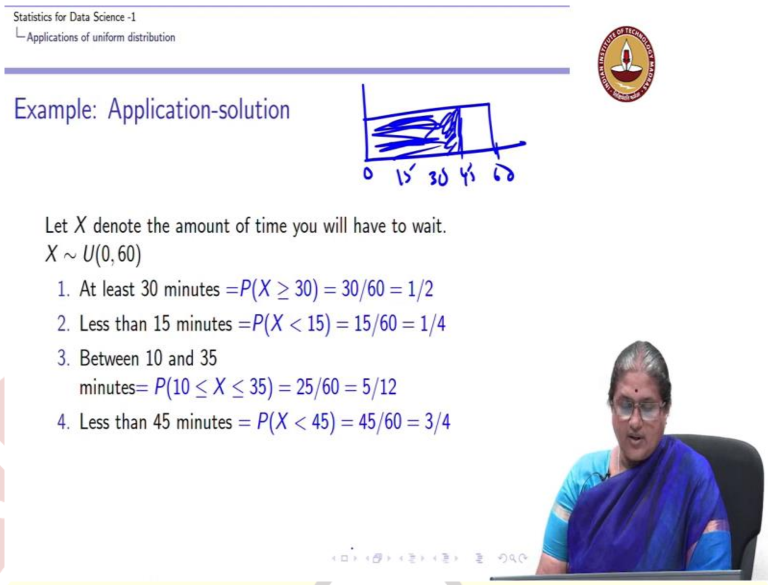
(Refer Slide Time: 14:59)



Now, let us look between 10 and 35 minutes again, let us go back, I have a 0 minute here I am just shifting this this 1 and this is 60 minutes. So, this is 30 minutes, this is 15 minutes, so this is going to be your 10th minute is here and your 35th minute is here. So, another way to look at it is let us look at 0 minute here, 60 minutes here, 30, so this is 10, 20, 40, 50 and I am interested in knowing I know it is the same probability across, so between 10 minute and 35 minutes.

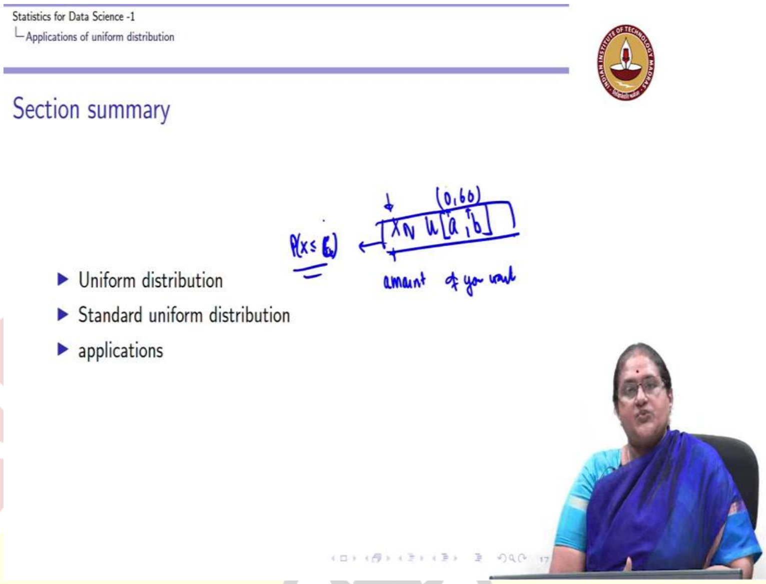
Now, between 15 and 0, I knew it was 1 by 4. So, this is also going to be 1 by 4. So, this is the area we are seeking now. So, this is $F(35) - F(10)$, which is $\frac{35-10}{60}$, which I get is 25 by 60, which is a 5 by 12 probability.

(Refer Slide Time: 16:19)



So, the last thing is less than 45 minutes, $P(X < 45)$,, this is straight forward, which is going to be 45 by 60. And I know that is three-fourth and that you can see that, if this is 0, and this is a 60, and this is a 30. This is a 45. This is a 15. I know the area, which is going to be less than or equal to this is three-fourth of the total area, the total area was 1. So, hence this is three-fourth, and the chance that you will be waiting for less than 45 minutes is a pretty high 75 percent chance.

(Refer Slide Time: 17:11)



So, what we have seen so far is we looked at what was a uniform random variable, what is a standard uniform distribution, and we looked at a simple application of a uniform distribution. The key point to note is when you are looking at a uniform distribution and an application, first recognize what are your parameters? So, we recognized what is the variable, define the variable.

So, in our application, we defined the variable to be the amount of time you wait. And since it was given to be that you are waiting equally between this time interval 2 pm to 3 pm, we recognized a as 0 and b as 60. And then once we find out and establish this we can, it is easy for us to answer questions regarding these whatever questions we need to know we are translating it in terms of probabilities.