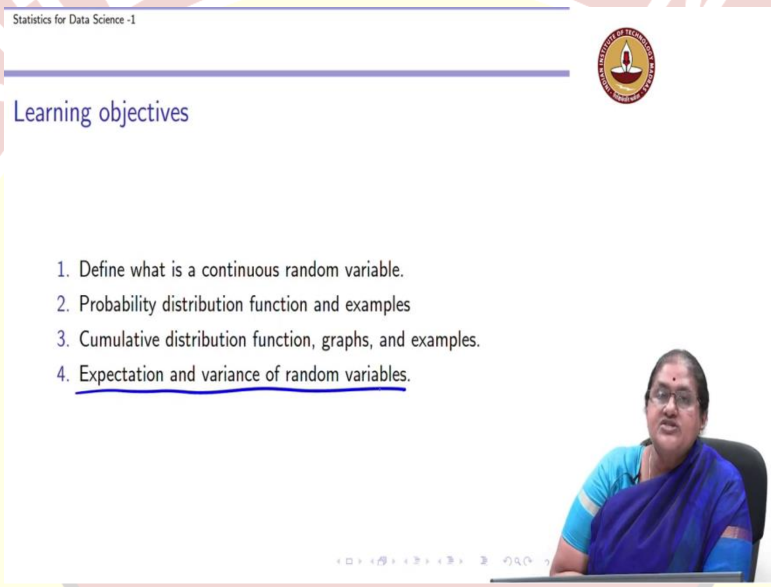
IIT Madras

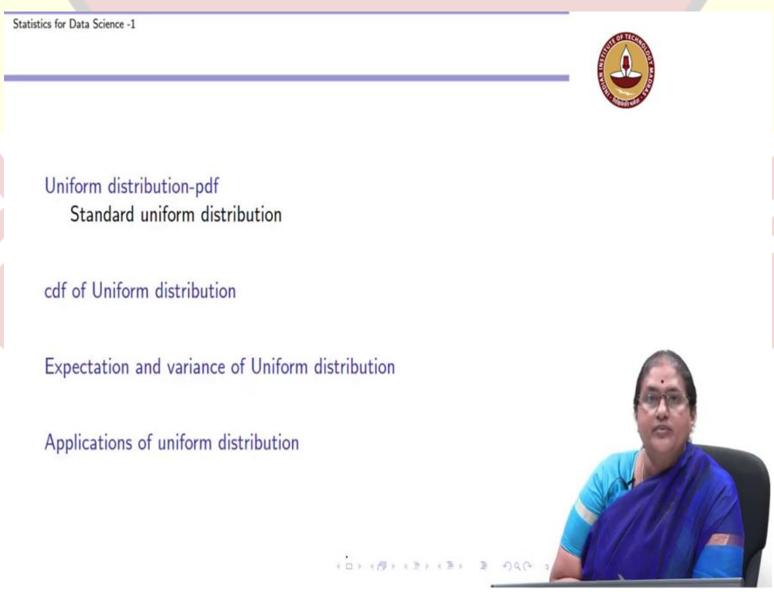
ONLINE DEGREE

Statistics for Data Sciences - 1
Professor. Usha Mohan
Department of Management Studies
Indian Institute of Technology, Madras
Continuous random variable - Uniform distribution

So, in today's session we are going to learn about a few important continuous random variables. We first begin with the uniform distribution.

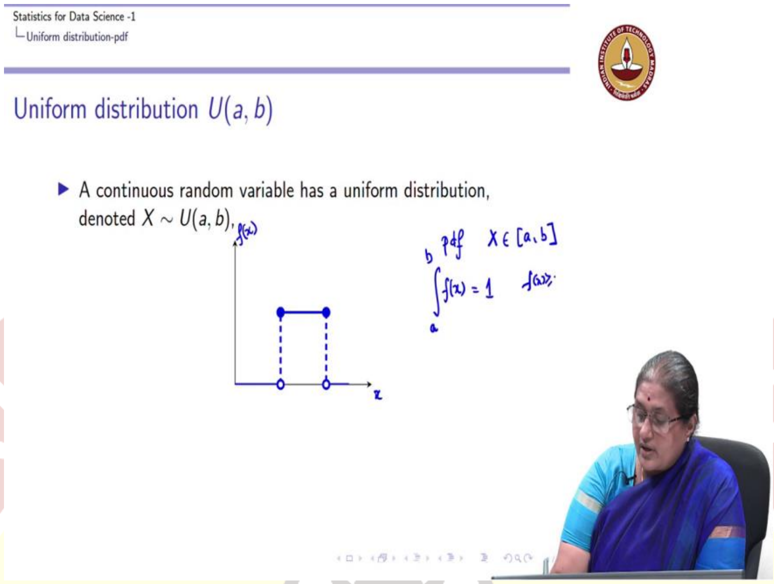
(Refer Slide Time: 0:27)

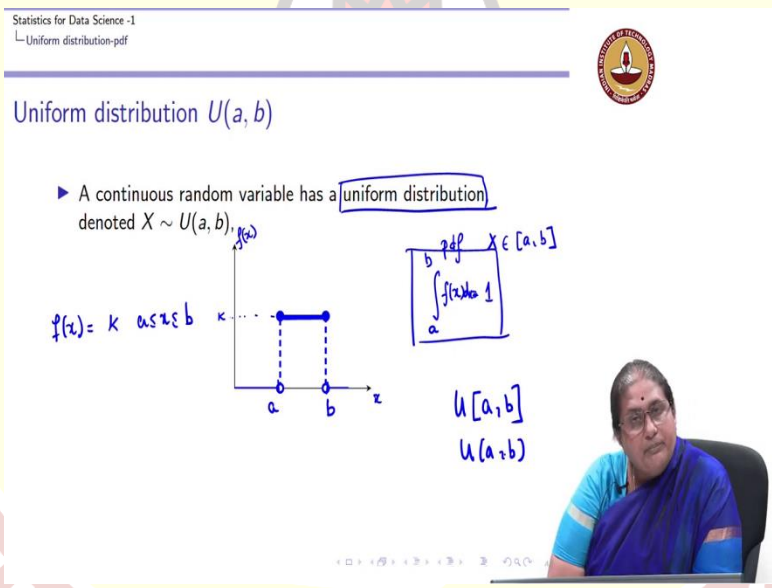


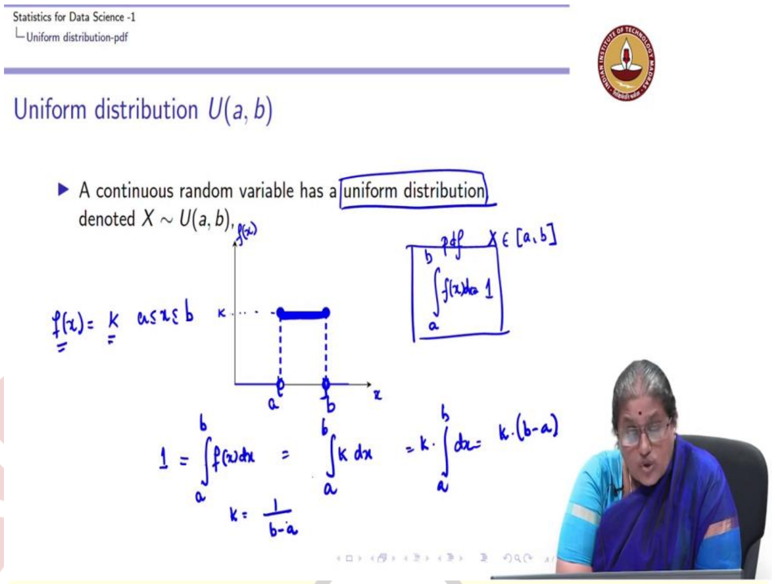


Now, when we look at a uniform distribution again this is the important thing which you need to understand as to what is a uniform distribution.

(Refer Slide Time: 0:36)







So, let us go back and understand what is a uniform distribution. So, if you remember if this is my x, this is going to be my PDF $f(x)$, so we introduced the concept of a probability density function in the last class and we said that if $f(x)$ is a probability if X is a random variable which takes values in a particular range, then over that range my $f(x)$ in other words if X $\epsilon[a, b]$; we said that $\int_a^b f(x)dx = 1$ and $f(x)$ should be non negative in that particular interval.

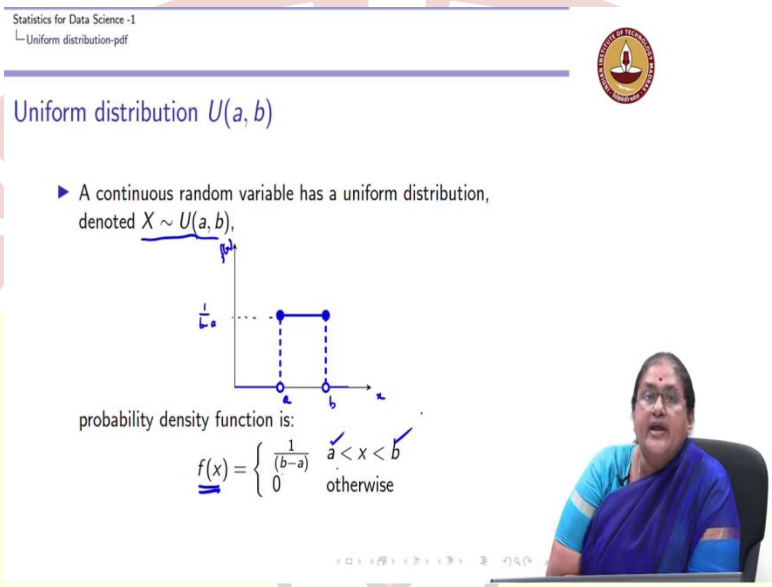
This is how we defined a PDF, so now this is a very important property that the probability of x belonging to that interval is going to be equal to 1 and hence $\int_a^b f(x)dx = 1$, this is what we discussed last time. So, now what is a uniform distribution? As the name suggests if I have any interval [a,b], then $f(x)$ is uniformly distributed in the other sense $f(x)$ takes a constant value throughout this interval.

So, I can say that $f(x)$ is a constant, some constant k for $a \le x \le b$. Now, remember again in the last session we said that x taking a particular value, $f(x)$ is a continuous random variable does not make sense because it is continuous, so $f(x)$ if $a < x < b$ is equivalent to $a \le x \le b$.

Hence, some books refer to a uniform random variable on a closed interval, some books refer to it as an open interval, we are going to assume that both of them mean the same. So, what does this mean that uniformly distributed? Uniformly distributed means that $f(x)$ has constant density in this interval (a, b).

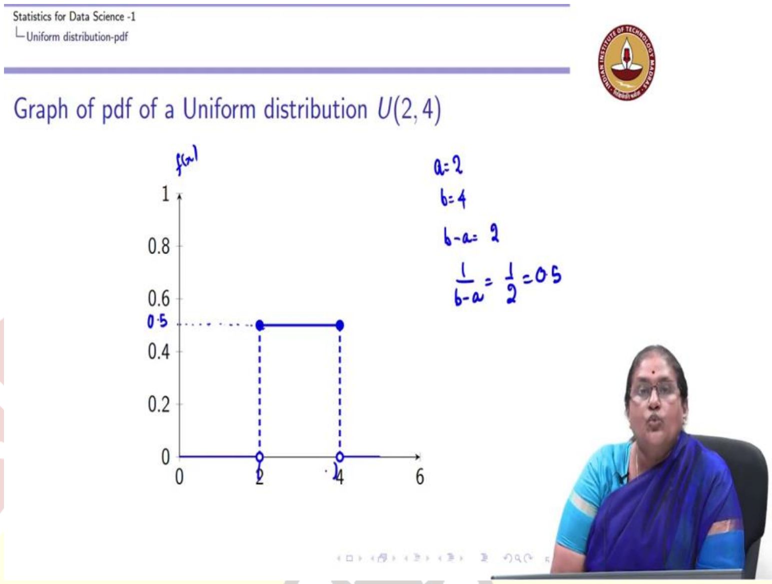
So, now what is this constant density? Now, if it is a constant density and if $f(x)$has to be a probability density function, then I know $\int_a^b f(x)dx = \int_a^b k\,dx = k\int_a^b dx = k(b-a)$ I know if this is a PDF I know this should be equal to 1 which gives me $k = \frac{1}{(b-a)}$.

(Refer Slide Time: 3:54)



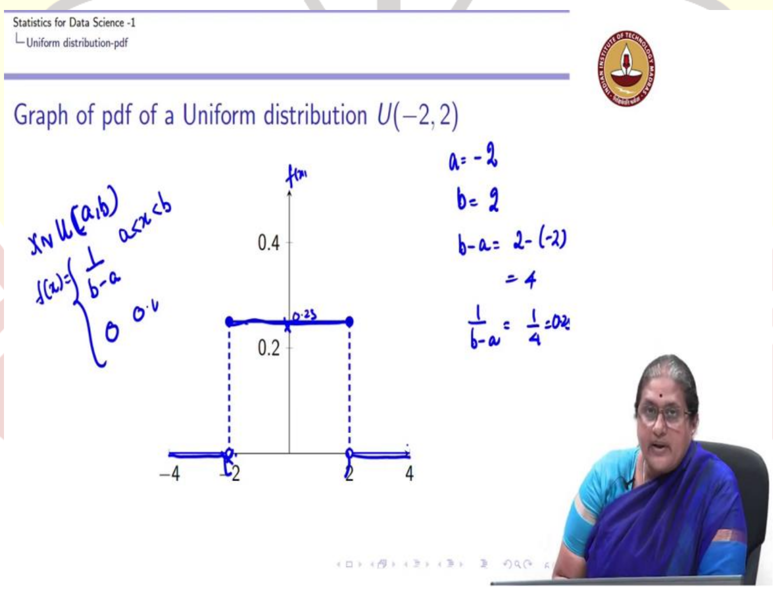
So, what have we achieved here is $f(x)$ is between a and b and this is x so if $X \sim U(a, b)$, then the probability density function we have obtained the $k = \frac{1}{(b-a)}$ hence the probability density function is $\frac{1}{(b-a)}; a < x < b$. So, this point is $\frac{1}{(b-a)}$, and it is 0 otherwise for any real value of a and b.

(Refer Slide Time: 4:40)



So, let us look at a few examples. Now, suppose my a = 2 and b = 4, I know b-a = 2 which gives $\frac{1}{(b-a)} = \frac{1}{2}$ which is a 0.5, so you can see that $f(x)$ is a constant with f of, so if this is my $f(x)$, it is taking a constant value which is equal to 0.5 for any value of x between 2 and 4.

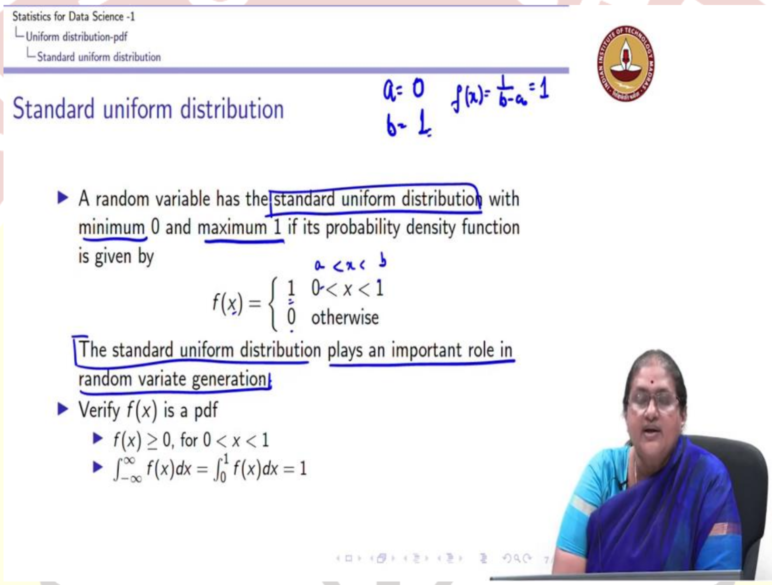
(Refer Slide Time: 5:12)



Let us look another case where a = -2 and b = 2 so my b - a = 2-(-2) = 4 which gives $\frac{1}{(b-a)} = \frac{1}{4}$ which is 0.25, so you can see in the interval (-2 , 2) $f(x)$ is taking the value 0.25 and this is a constant in the interval. So, in summary if I have X is a uniformly distributed random variable in an interval a to b again I am not making a distinction, I am going to represent it as (a,b), some books represented as [a,b].

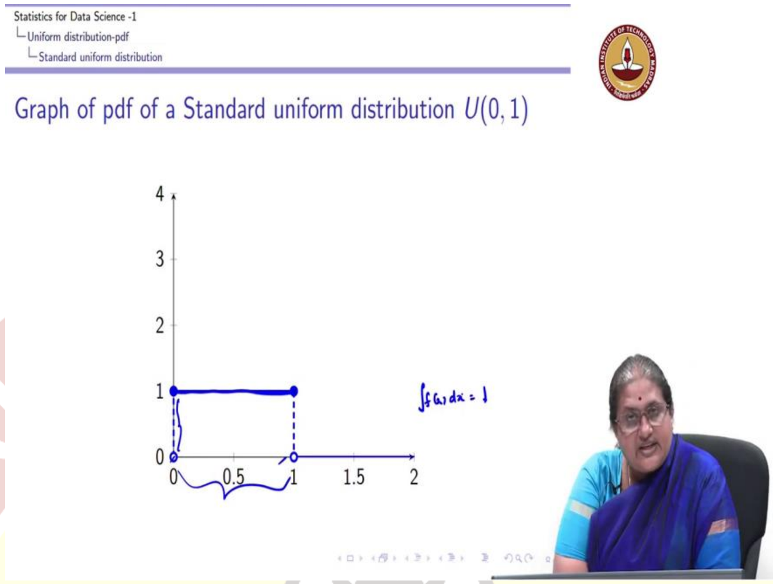
I have a probability density function which takes the value $\frac{1}{(b-a)}$ for $a < x < b$ and 0 otherwise. So, it is taking the value 0 outside the interval and within the interval it is taking the value $\frac{1}{(b-a)}$.

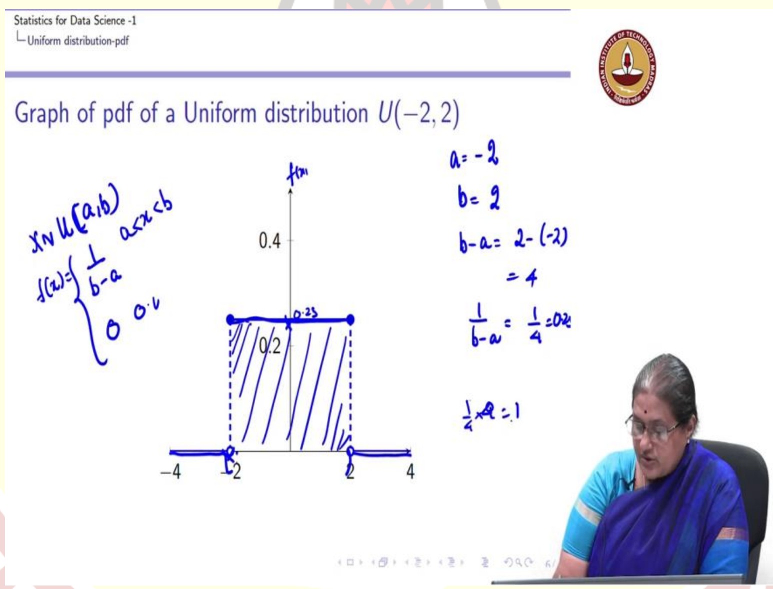
(Refer Slide Time: 6:26)

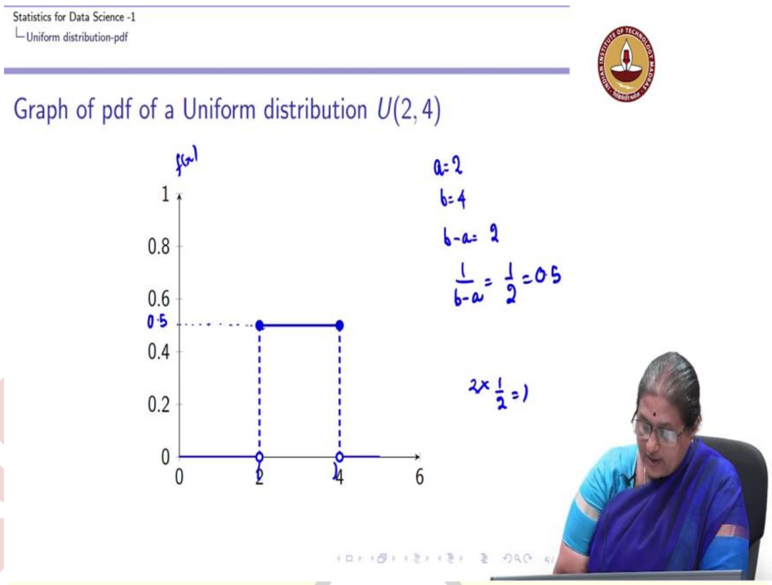


A special case of the uniform distribution is when a takes the value 0 and b takes the value 1 that is minimum is 0 and maximum is 1 and this is referred to as a standard uniform distribution, I repeat, a special case of a uniform distribution where the minimum value is 0 and maximum value is 1 is referred to as a standard uniform distribution and for that so with the density it is again going to be $\frac{1}{(b-a)}$, b = 1 and a = 0, which is equal to 1, so my density is 1 for $0 < x < 1$, 0 otherwise.

(Refer Slide Time: 7:24)



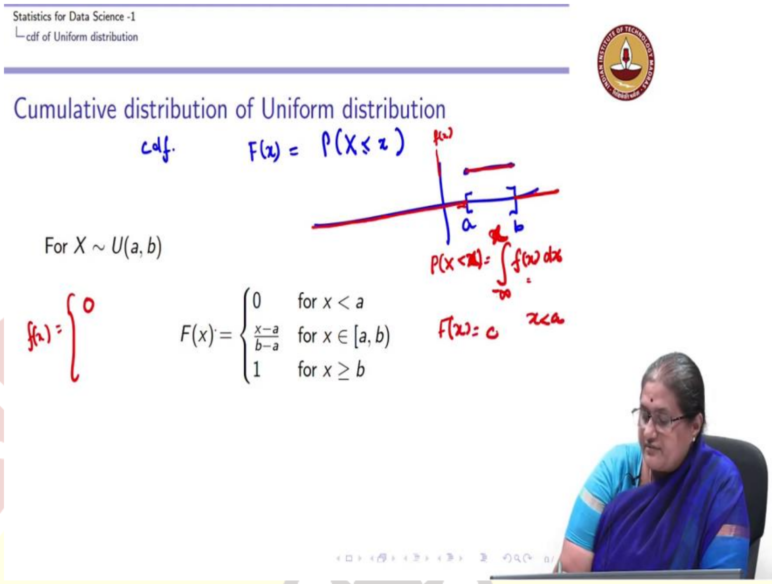


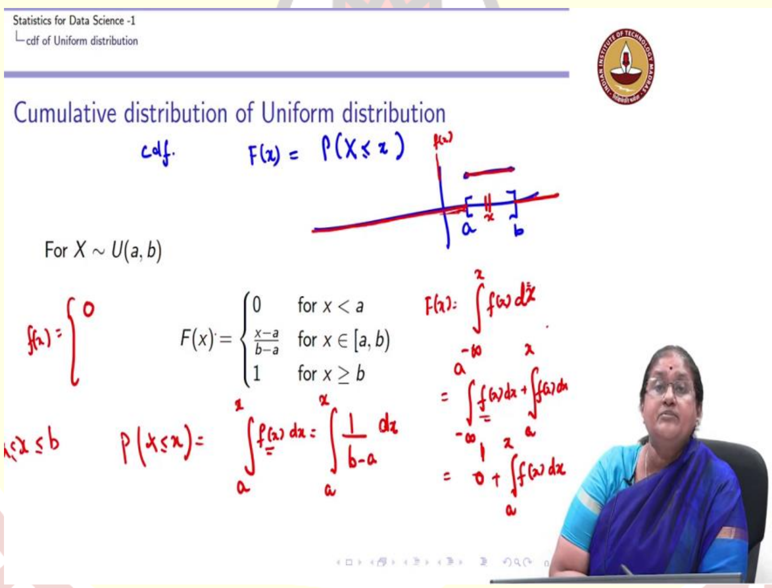


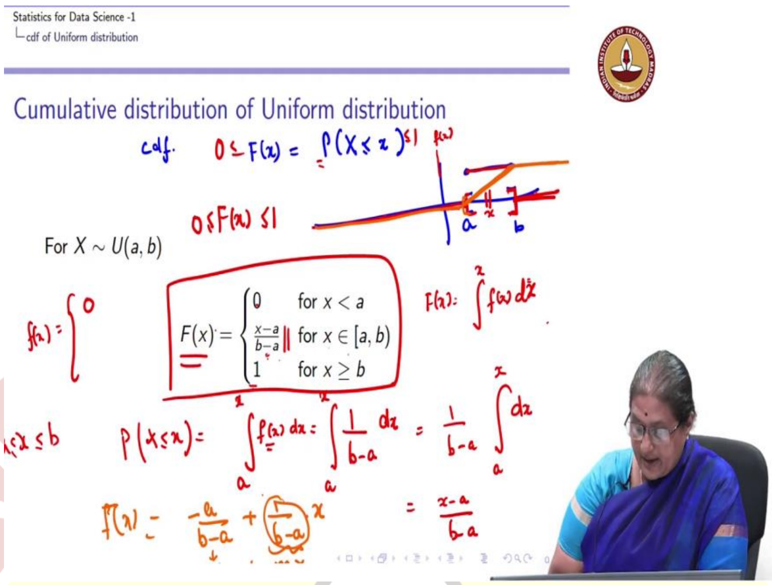
So, the density function will take the value 1 in the interval 0 to 1 and it is 0 otherwise. Now, we also can confirm that the area under the density function is nothing but -2 to 2 with 0.5 and you can see that is $\frac{1}{4} * 4 = 1$. Here it is going to be $\frac{2*1}{2}$ which is equal to 1. And in the case of the standard uniform this is 1, this is 1 so the area integral $\int f(x)dx = 1$ all of it hence verifying that this $f(x)$ or the PDF is indeed a probability density function.

So, now we can verify for all of that that this is a probability density function, this standard uniform distribution, it plays an extremely important role in random variate generation which you are going to come across in your advanced courses, especially when you are going to simulate random numbers. This is about the probability density function of a uniform distribution and a standard uniform distribution.

(Refer Slide Time: 8:58)







Now, what is the cumulative distribution of a uniform distribution? Recall the cumulative distribution function or $\mathrm{F}(x) = P(X \leq x)$. Again, I know that this if I am given a to b, I have this is my probability density function, it takes the value so it is taking a value 0 till it hits this $f(x)$ is going to take and it is 0 elsewhere and it is going to be $f(x)$ here, this is my probability density function.

So, for $\mathrm{x} < \mathrm{a}$. So, whenever $\mathrm{x} < \mathrm{a}$, $\int_{-\infty}^{a} f(x)dx$, I also know $f(x)$ takes the value 0 outside this interval. So, this is going to be 0, hence, $\mathrm{F}(x)$ or if $P(X \leq x)$ for $\mathrm{x} < \mathrm{a}$ $\mathrm{F}(x)$ is going to take the value 0.

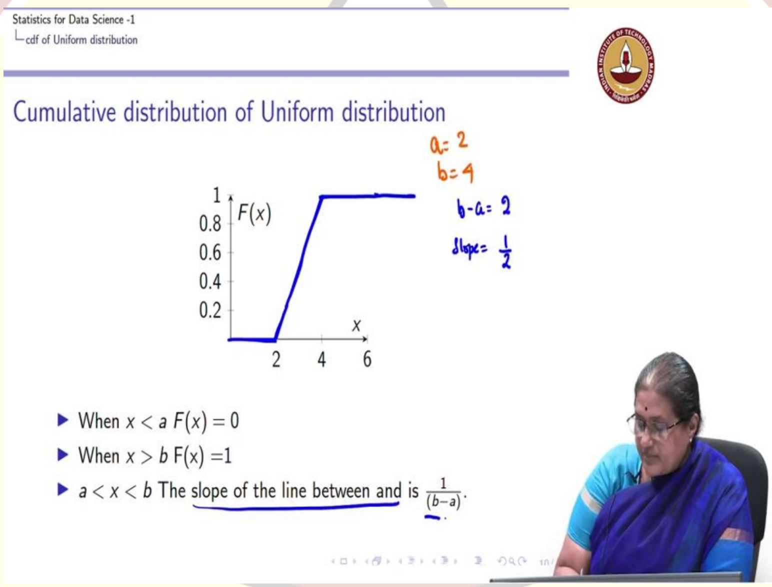
Now, if x is equal to a point which is between a and b. Suppose x is here, now for this point my $f(x)$ is going to be $\int_{-\infty}^{x} f(x)dx$ which is going to be $\int_{-\infty}^{a} f(x)dx + \int_{a}^{x} f(x)dx$. Now, I know that $f(x)$takes the value 0 for $\mathrm{x} < \mathrm{a}$, so this will contribute a value of 0 and this is going to be $\int_{a}^{x} f(x)dx$. So, for x lying between a and b I have $P(X \leq x) = \int_{a}^{x} f(x)dx$.

Now, x is lying between a and b I know $f(x)$ when it is lying between a and b is $\frac{1}{(\mathrm{b}-\mathrm{a})}$. So, I have a , $\int_{a}^{x} \frac{1}{(b-a)} dx$. Now, this $\frac{1}{(b-a)}$ is a constant, so what I can do is this $\frac{1}{(b-a)}$ can be taken outside and this becomes $\frac{1}{(b-a)} \int_{a}^{x} dx$ which I know is nothing but $\frac{(x-a)}{(b-a)}$.

Hence, I have $F(x) = \frac{(x-a)}{(b-a)}$ for $x\epsilon[a, b)$. Now, what happens for x is here, so at x = b I know $f(x) = 1$, so for any $x \geq b$,$F(x)$ is going to remain to be 1 because I know a cumulative distribution function is a probability and that lies between 0 and 1.

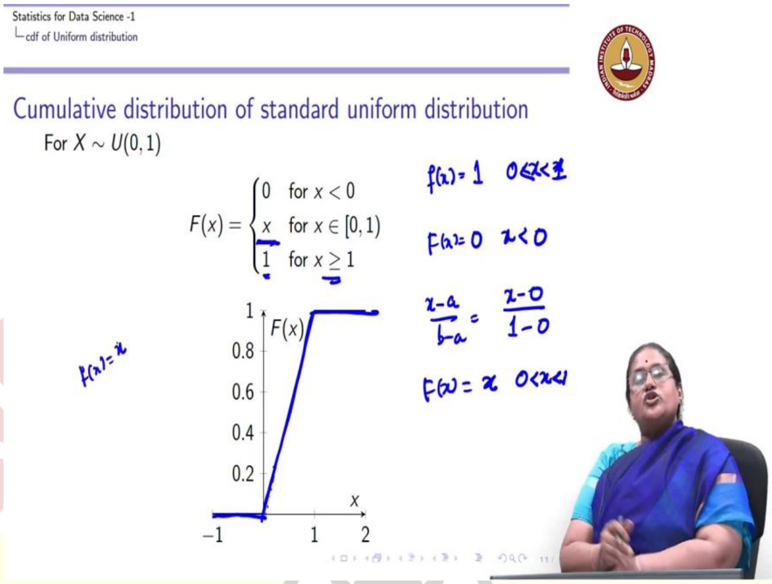
Hence, the cumulative distribution function of my uniform distribution is given by the following which takes the value 0 till it hits a, between a and b it would be a straight line because I can write this $\frac{(x-a)}{(b-a)}$ as $\frac{(-a)}{(b-a)} + \frac{1}{(b-a)} * x$ so I get $F(x)$ is , I can recognize that this is some constant plus a slope into x. So, this is a straight line with slope $\frac{1}{(b-a)}$ x and beyond this it is going to remain at 1.

(Refer Slide Time: 13:29)



So, the CDF or the cumulative distribution function of a uniform distribution would be 0 till it is at a, it would be 1 beyond b and between a and b it is going to be a equation of a straight line with slope $\frac{1}{(b-a)}$. So, again going back to our examples where a was 2 and b was 4, I can see that for this example b - a = 2, so the slope is going to be $\frac{1}{2}$ which is $\frac{1}{(b-a)}$, for x<2 $F(x)$was 0, for x >4 $F(x)$is going to continue to be 1 and between 2 and 4 we can see that it is a straight line with slope $\frac{1}{2}$. So, it is always the slope of the line between b and a is $\frac{1}{2}$.

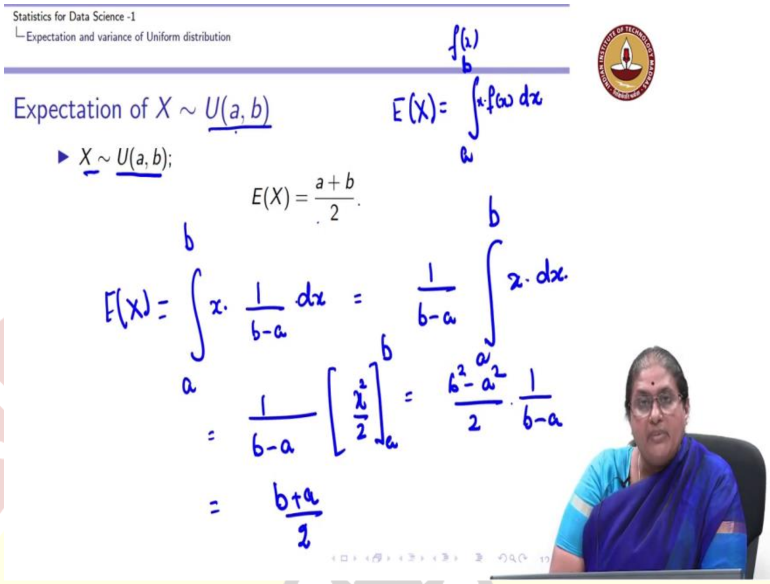
(Refer Slide Time: 14:35)

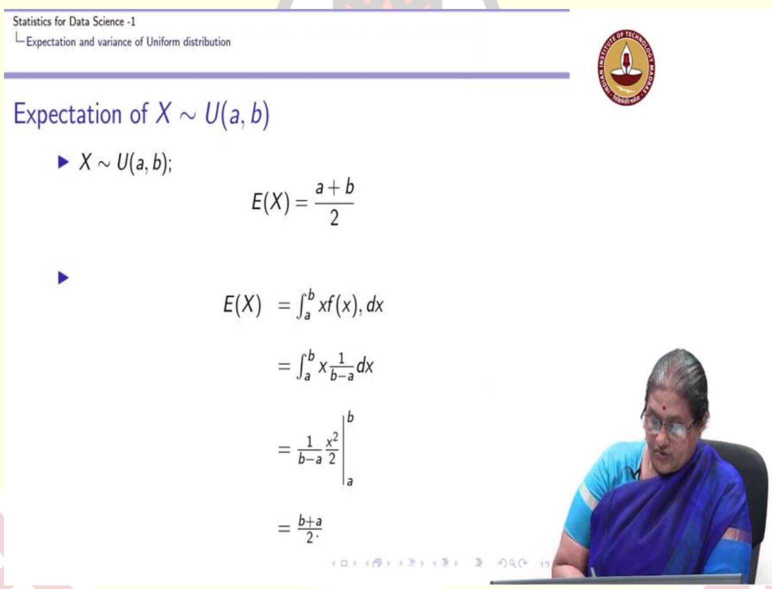


So, if you go to this example of a standard uniform, we know that $\mathrm{f}(x) = 1$ for $0 \leq \mathrm{x} < 1$. so my $\mathrm{F}(x) = 0$ for x < 0, for, so $\frac{(x-a)}{(b-a)}$ is the same as $\frac{(x-0)}{(1-0)}$. So, $\mathrm{F}(x)$ takes the value x for $0 \leq \mathrm{x} < 1$ and that is what we have here and $\mathrm{F}(x)$ takes the value 1 for x ≥1, and you can see that this is what for x < 0 it is 0, for x >1 it is 1, and between x lying between 0 and 1, $\mathrm{F}(x)$ takes the value x. So, $\mathrm{F}(x)$ takes the value x. So, this is about the cumulative distribution function of a uniform and a standard uniform distribution.

Now, we will look at the important properties of a uniform distribution. Recall whenever we are talking about a random variable we always wanted to define the random variable give the probability distribution function of the random variable, come up with a cumulative distribution function of the random variable and look at the expectation and variance of these random variables.

(Refer Slide Time: 16:28)



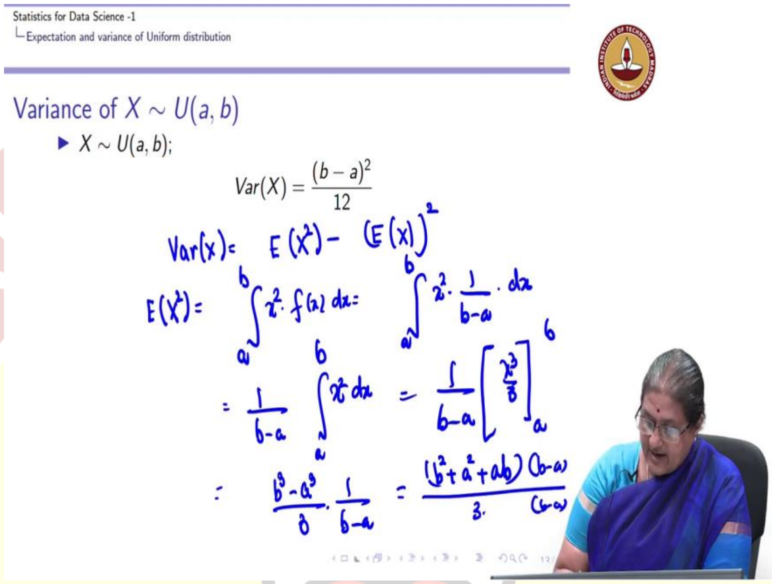


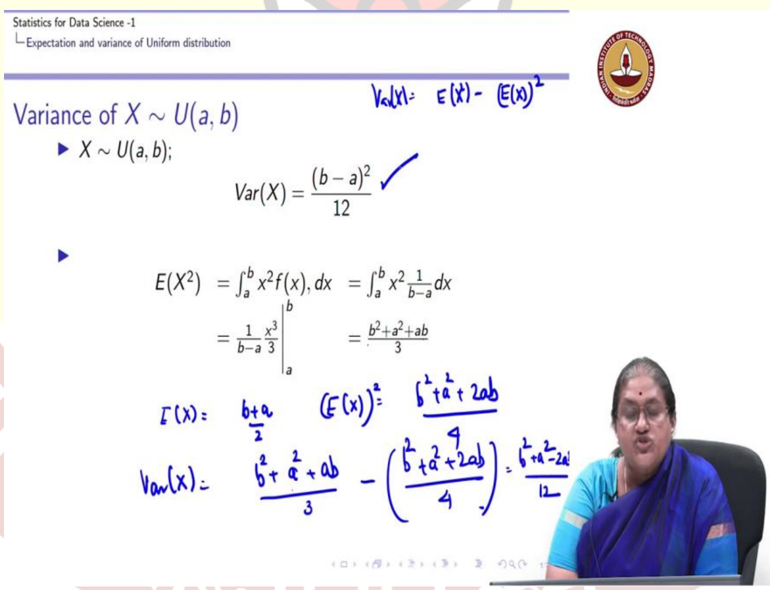
So, we are going to look now at the expectation, so we are going to look now at the expectation and variance of the uniform random variable which is uniformly distributed over an interval a to b. We all know that expectation of a random variable with a PDF $f(x)$ is over the range, so the range here of definition is $\int_a^b xf(x)dx$.

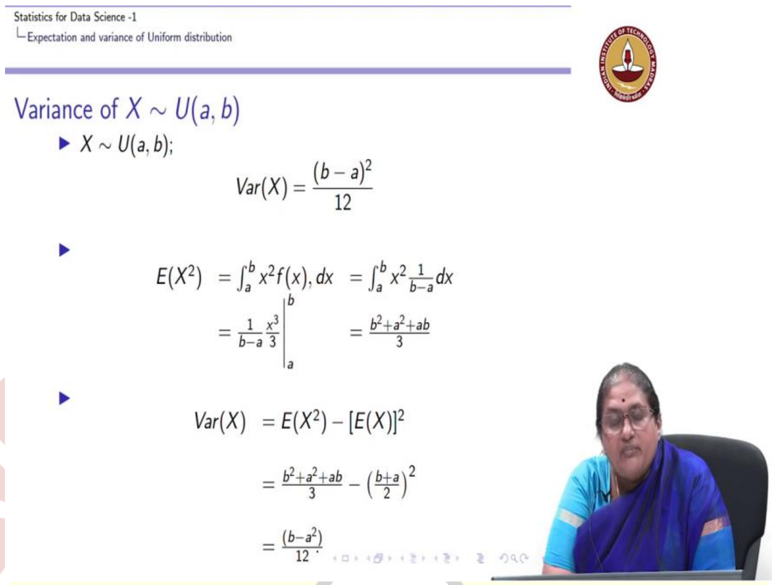
So, I have a to b , $E(X) = \int_a^b xf(x)dx = \int_a^b x \frac{1}{(b-a)} dx = \frac{1}{(b-a)} \int_a^b xdx = \frac{1}{(b-a)} \left[\frac{x^2}{2}\right]_a^b = \frac{b^2-a^2}{2} \frac{1}{(b-a)}$

$= \frac{(b+a)}{2}$. So, the expectation of a random variable which is uniformly distributed over a and b is $\frac{(a+b)}{2}$. So, we can see that it is $\frac{(a+b)}{2}$.

(Refer Slide Time: 17:59)







Now, let us look at the variance, to compute the variance we use the computational formula which is $\mathrm{V}(X) = E(X^2) - (E(X))^2$. Now, let us compute $E(X^2) = \int_a^b x^2 f(x)dx = \int_a^b x^2 \frac{1}{(b-a)} dx = \frac{1}{(b-a)} \int_a^b x^2 dx = \frac{1}{(b-a)} \left[\frac{x^3}{3}\right]_a^b = \frac{b^3-a^3}{3} \frac{1}{(b-a)} = \frac{(b^2+a^2+ab)(b-a)}{3(b-a)}$ $= \frac{b^2+a^2+ab}{3}$.

Now, $E(X) = \frac{(b+a)}{2}$ which would give me $(E(X))^2 = \frac{b^2+a^2+2ab}{4}$. Hence, my

$\mathrm{V}(X) = \frac{b^2+a^2+ab}{3} - \left(\frac{b^2+a^2+2ab}{4}\right)$, this is what we have been applying throughout.

$= \frac{b^2+a^2-2ab}{12} = \frac{(b-a)^2}{12}$. So, we can see that the $\mathrm{V}(X) = \frac{(b-a)^2}{12}$.