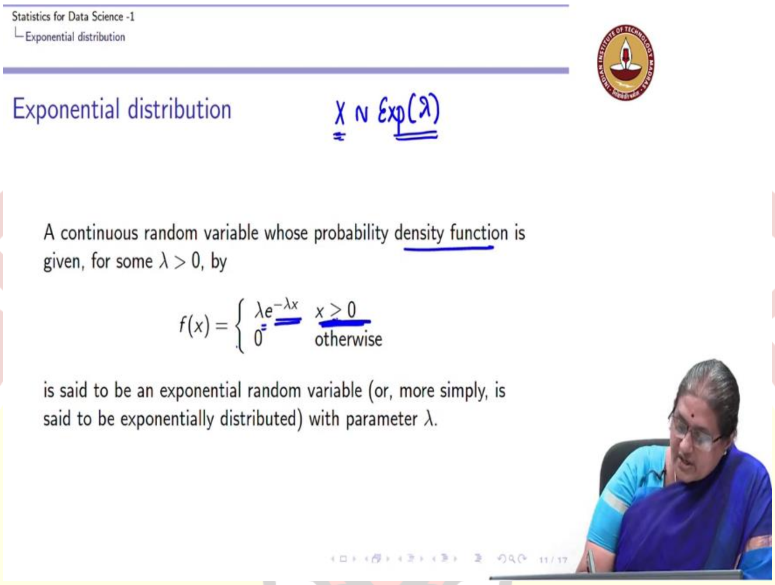
IIT Madras

ONLINE DEGREE

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

(Refer Slide Time: 0:14)

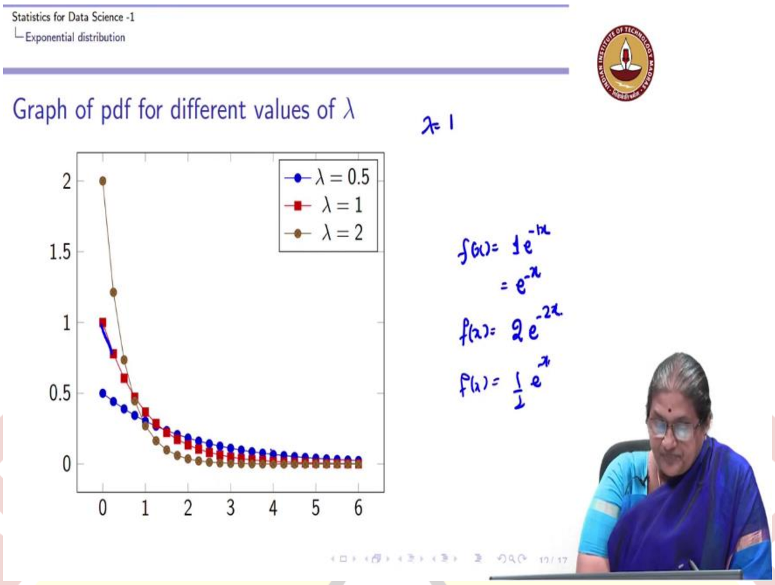


The next distribution we are going to discuss when we are talking about continuous distributions is what we refer to as a exponential distribution. Again, exponential distribution has a lot of application. Typically, exponential distributions are used to model inter-arrival times in a queuing system, and service times also when we are having queues.

Not only queues wherever you have the model system where you have a sequence of people arriving into any particular system or it could be a bank, it could be a hospital, then the inter-arrival times between the people who are arriving and the service times typically are modeled using an exponential distribution. So, we are going to spend some time to understand about this exponential distribution.

So, a continuous random variable X, so a random variable X is said to be a exponential distribution with parameter λ. So, X is a random variable, it could measure anything, it could measure the amount of time, but it is said to be exponential distribution with parameter λ if its probability density function is given as $f(x) = \begin{cases} \lambda e^{-\lambda x} & x \geq 0 \\ 0 & otherwise \end{cases}$. So, this is called the parameter of the exponential distribution.

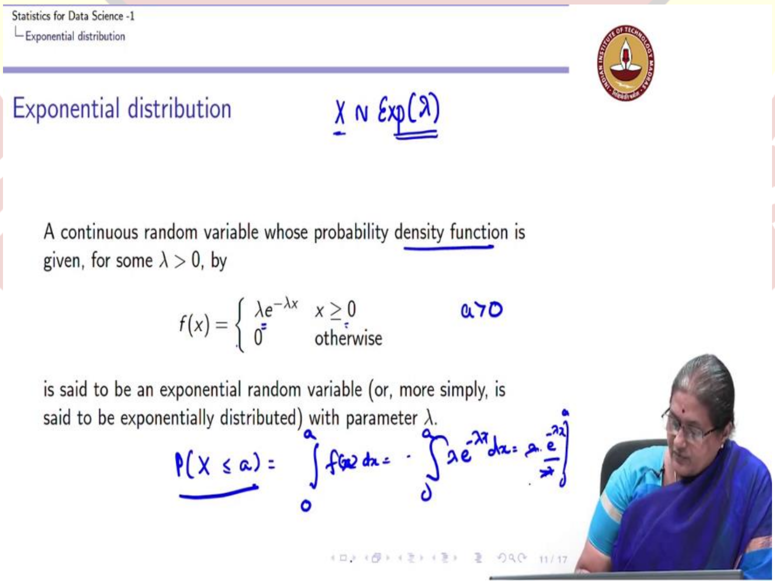
(Refer Slide Time: 1:59)

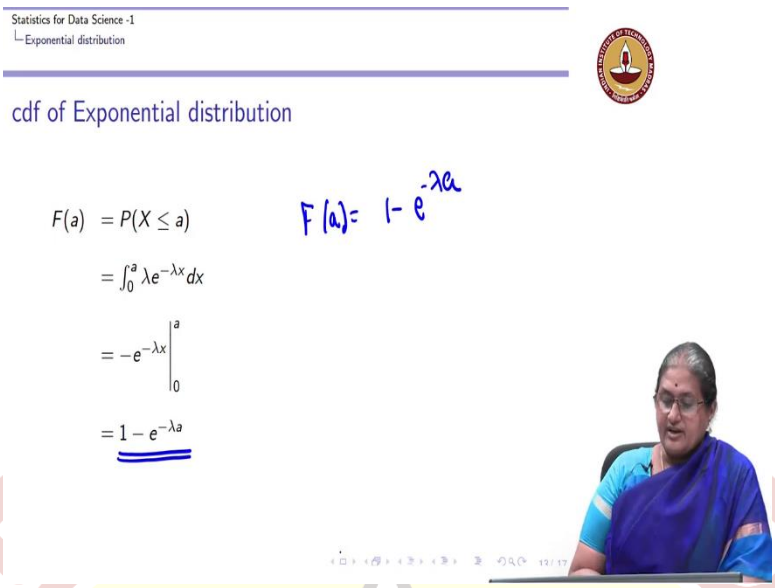


So, for different values of the parameter. So, if your λ =1, you see that this $f(x)$ just reduces to e^{-x} , so for λ =1 you can see the red curve, it is an exponential curve. For λ =2 you can see that it is $2e^{-2x}$ which is the blue brown curve, and for λ =1/2, $f(x) = \frac{1}{2} * e^{-0.5x}$, which is the blue curve.

So, you can see that depending on what is your parameter the PDF of this exponential distribution is going to take the different curves.

(Refer Slide Time: 2:53)



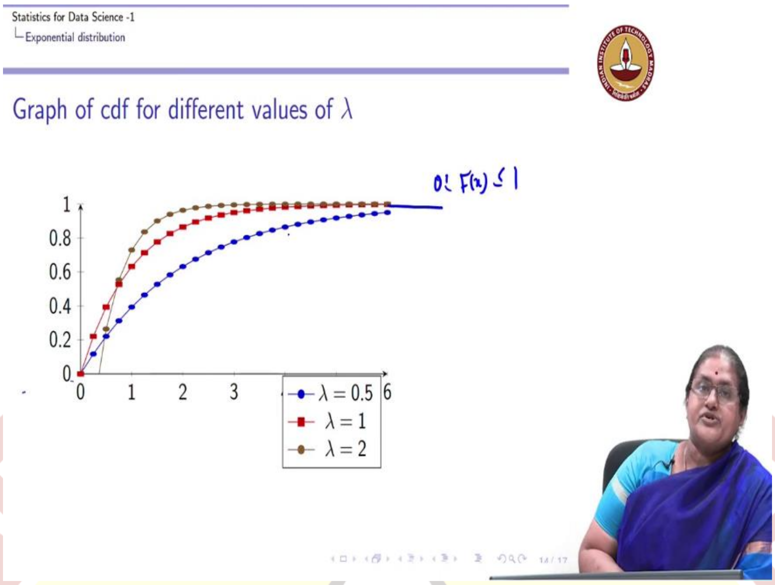


Now, what is the CDF or the cumulative distribution function for the exponential distribution? Recall the CDF, $P(X \le a) = \int_{-\infty}^{a} f(x)dx$. This is how we define for a continuous random variable.

So, this for X is less than 0, $f(x)$ =0, so $F(X \le a) = \int_{-\infty}^{a} f(x)dx = \int_{-\infty}^{0} 0dx + \int_{0}^{a} f(x)dx$. Now, this term is going to vanish because it is just going to be 0 and it is going to be $\int_{0}^{a} f(x)dx$. So, if a is less than so for any value since x greater than 0, it is going to be $f(x) = \lambda e^{-\lambda x}$ so $P(X \le a) = \int_{0}^{a} f(x)dx$.

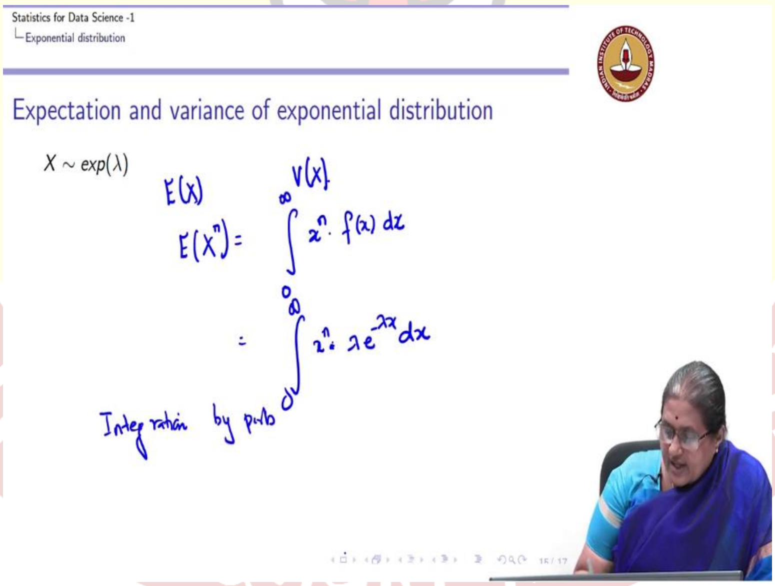
So, now we can see that this $P(X \le a) = \int_{0}^{a} f(x)dx = \int_{0}^{a} \lambda e^{-\lambda x}dx$, again, I can remove λ out and I get this is $\lambda\frac{e^{-\lambda x}}{-\lambda}|_{0}^{a}$, $-\lambda$ cancels out which will give me the CDF, is nothing $F(a) = \int_{0}^{a} \lambda e^{-\lambda x}dx = e^{-\lambda x}|_{0}^{a} = 1 - e^{-\lambda a}$.

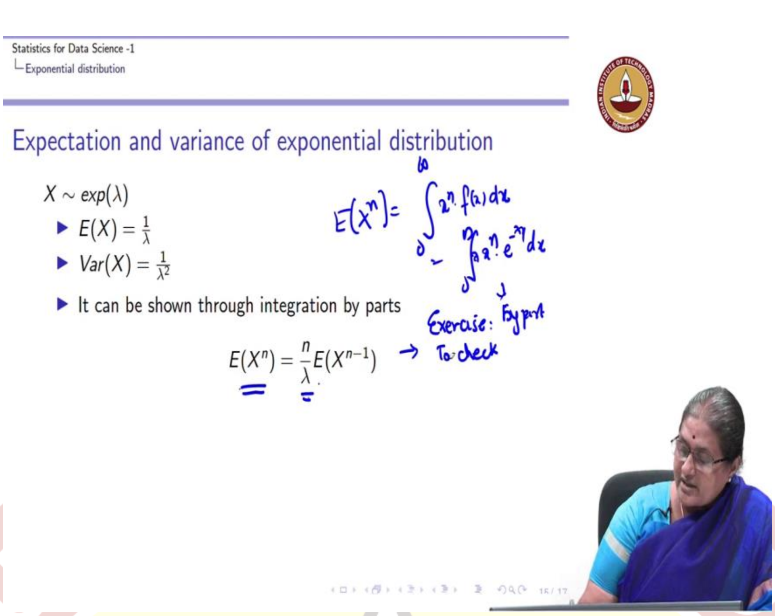
(Refer Slide Time: 4:56)



Again for different values of λ, we know that F(x) will always lie between 0 and 1. So, for different values of λ, you can see that for all below x equal to 0 is going to be 0 and then all of them taper off at 1 and these are the shapes of the cumulative distribution function.

(Refer Slide Time: 5:21)

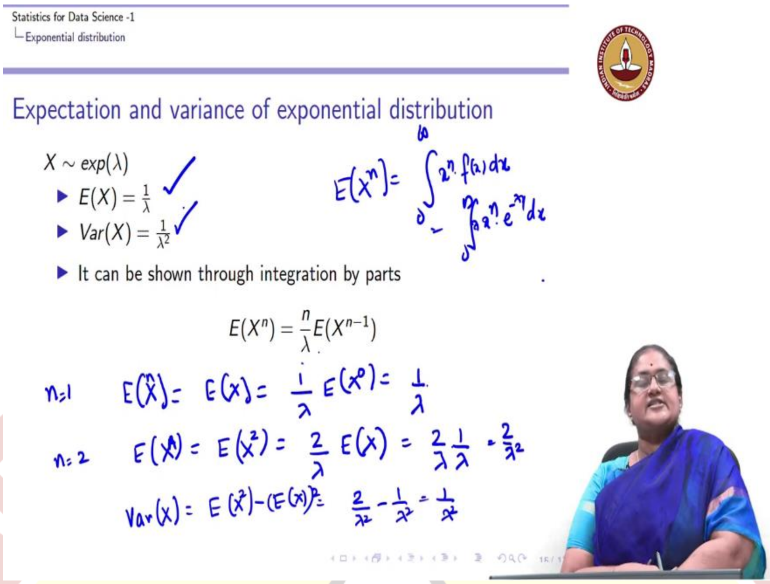




What is the expectation of this random variable? To get the expectation and variance of the exponential distribution, we will just introduce an $E[X^n]$, $E[X^n] = \int_0^\infty x^n f(x)dx = \int_0^\infty x^n \lambda e^{-\lambda x} dx$.

I am not going to do the entire working here, but you can strap that if you apply integration by parts, we can obtain the following that $E[X^n] = \frac{n}{\lambda} E[X^{n-1}]$, . This I leave it as an exercise to check using the formula $E[X^n] = \int_0^\infty x^n f(x)dx = \int_0^\infty x^n \lambda e^{-\lambda x} dx$, use expect integration by parts and check that $E[X^n] = \frac{n}{\lambda} E[X^{n-1}]$. I will leave it as an exercise.

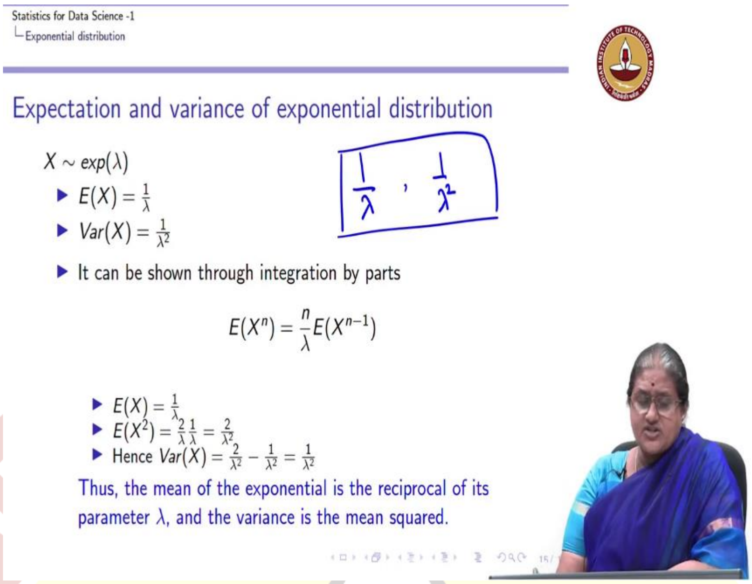
(Refer Slide Time: 6:56)



Now, once you accept this entity, or this identity, I can verify for n =1, $E[X^1] = \frac{1}{\lambda}E[X^0 = 1]$, x and E[1]=1, I get the $E[X] = \frac{1}{\lambda}$. For n = 2, I have $E[X^2] = \frac{2}{\lambda}E[X]$..

I already know $E[X] = \frac{1}{\lambda}$, so I get $E[X^2] = 2/\lambda^2$. Again, the computational formula for Var(X)= $E[X^2] - E[X]^2 = \frac{2}{\lambda^2} - \frac{1}{\lambda^2} = \frac{1}{\lambda^2}$,Hence, I have shown that variance of x is $\frac{1}{\lambda^2}$.

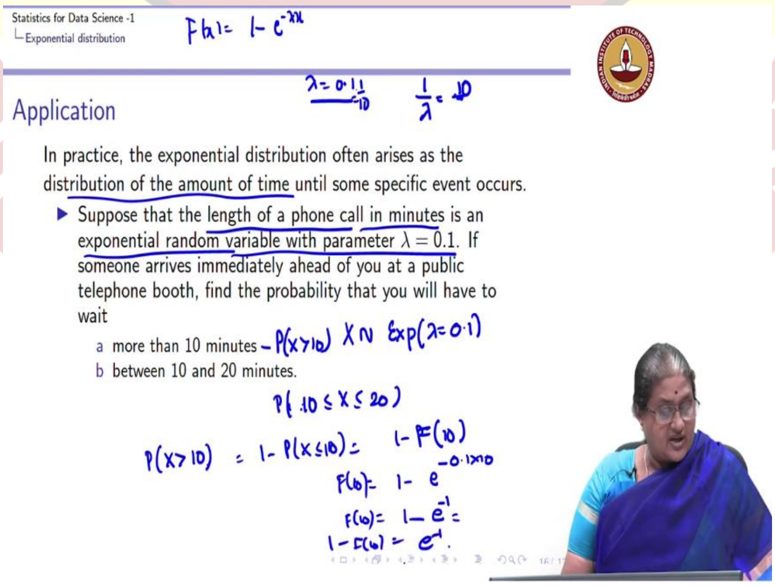
(Refer Slide Time: 8:17)



So, what does this mean? It tells us that the mean. So, if I have a exponential distribution with parameter λ, what it tells us is the mean of the distribution is just the reciprocal of the parameter $1/\lambda$ and the variances square of the mean. That is what it is an important observation and this property actually helps when you come for application or modeling situations using an exponential distribution.

So, let us look at the application of the exponential distribution. When we look at an application of the exponential distribution, let us go to the following problem.

(Refer Slide Time: 9:09)



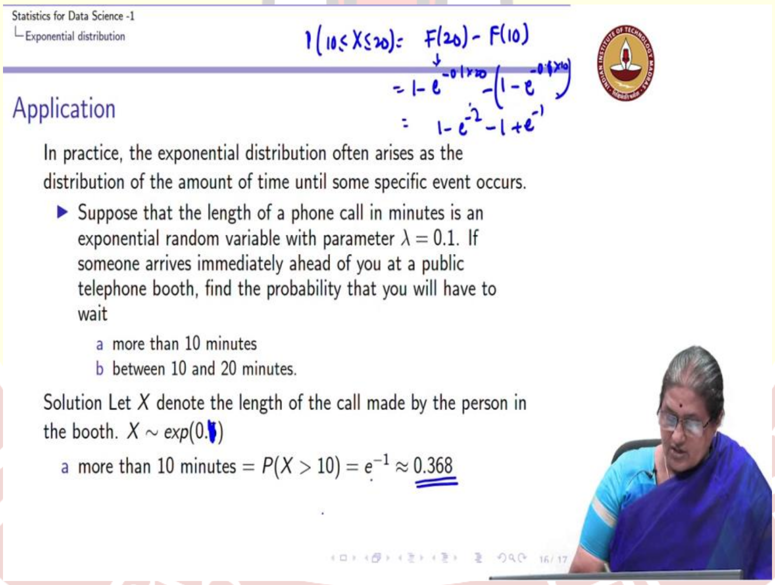
So, the exponential distribution arises, as the distribution of the amount of time. Again, remember we are measuring time. Suppose the length of a phone call in minutes is an exponential random variable with a parameter λ =0.1. So, if the λ =0.1, I know the mean=1/ λ, which is lambda = 0.1, mean is 1/0.1=10. In other words, I can say that every 10 minutes of phone call comes or it is the mean or an average, the distribution of time.

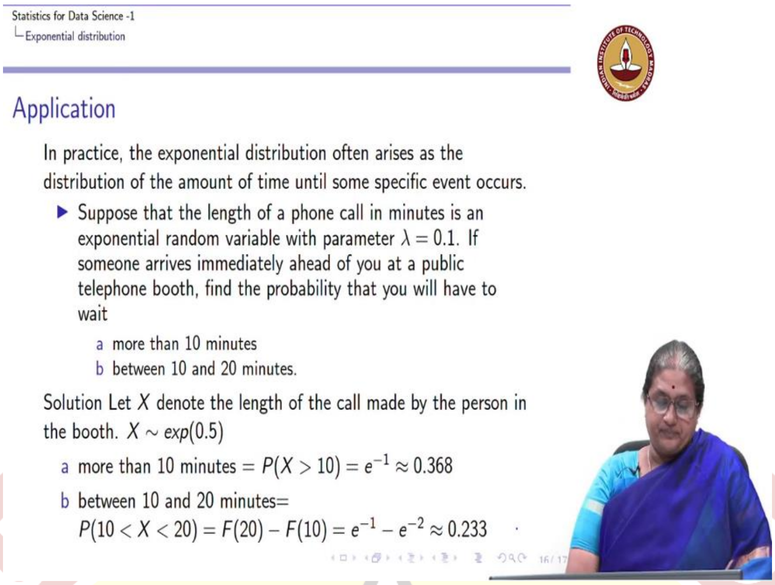
If somebody arrives immediately ahead of you at public telephone both, what is the probability that you will have to wait for more than 10 minutes? So, I know X is an exponential distribution with λ=0.1. So, I am going to wait for more than 10 minutes is $P(X > 10)$, and between 10 and 20 minutes is going to be $P(10 \leq X \leq 20)$,.

Let us answer each one of them. $P(X > 10) = 1 - P(X \leq 10) = 1 - F(10)$, Recall, $F(x) = 1 - e^{-\lambda x}$. And so, this is λ=0.1, $F(10) = 1 - e^{-\lambda x} = 1 - e^{-0.1*10} = 1 - e^{-1}$, and hence I know that $P(X > 10)$ is equal to…

So, $1 - P(X \leq 10) = 1 - F(10) = 1 - (1 - e^{-1}) = e^{-1} = 1/e$, which is approximately about 0.368.

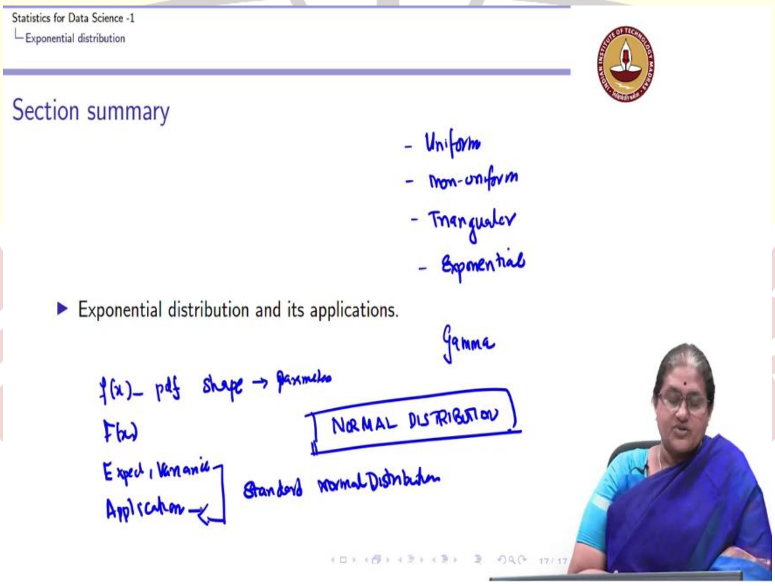
(Refer Slide Time: 11:27)





Similarly, if I want to know $P(10 \leq X \leq 20) = F(20) - F(10)$. Again, $F(20) = 1 - e^{-0.1*20} = 1 - e^{-2}$. $F(10) = 1 - e^{-0.1*10} = 1 - e^{-1}$. So, this is $1 - e^{-2} - (1 - e^{-1})$. So, the probability that I have to wait for more than, so the probability I will have to wait for more than between 10 and 20 minutes is equal to $e^{-1} - e^{-2}$, which is approximately point 0.233.

(Refer Slide Time: 12:51)



So, in summary, what we have seen so far is we looked at certain continuous distributions, and in particular, we looked at the uniform distribution. We looked at a case of a non-uniform distribution, we looked at an example of a triangular distribution, we looked at the PDF, the CDF the expectation and the variance of an exponential distribution. These are commonly arising continuous distributions. There are many more continuous distributions. For example, the gamma distribution is another important distribution, which arises a lot in modeling, but the most important distribution is, what we refer to as the normal distribution.

So, the next week, we are going to spend our time to understand the normal distribution. In particular, we will again start with what is the PDF of a normal distribution, the shape of the normal distribution and how it varies with respect to the parameters of the distribution, the CDF of the normal distribution, the expectation and the variance of a normal distribution.

And finally, we are going to spend a lot of time in understanding about applications. Before we go to applications, we will introduce what is a standard normal distribution and then we are going to look at how we will answer a lot of questions based on the distribution. So, this is the last topic of the course.