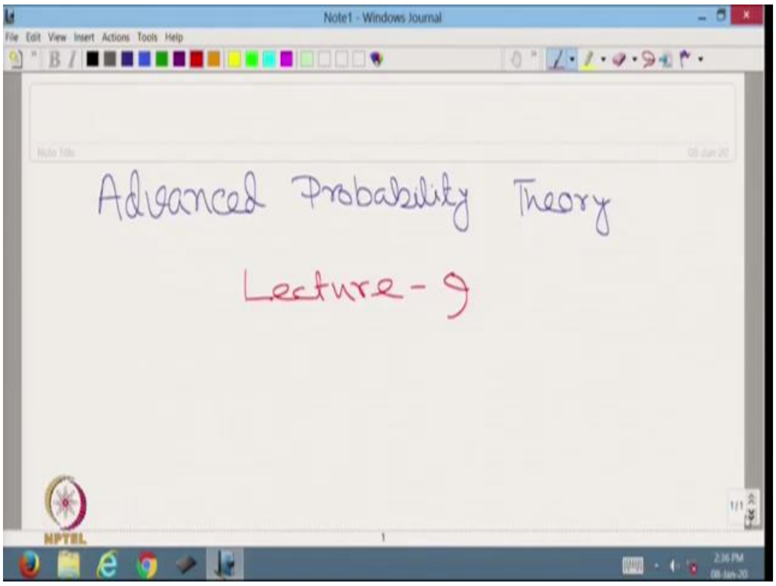
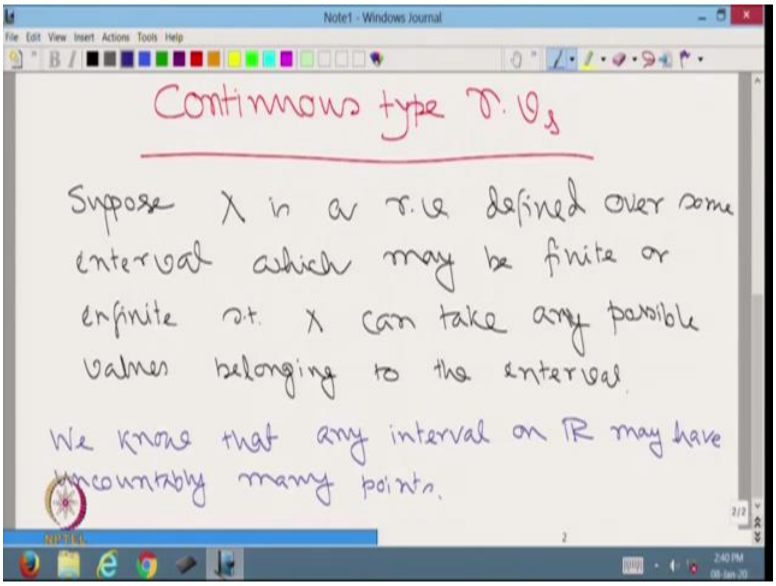
Advanced Probability Theory.
Professor Niladri Chatterjee.
Department of Mathematics.
Indian Institute of Technology, Delhi.
Lecture 9

(Refer Slide Time: 0:27)



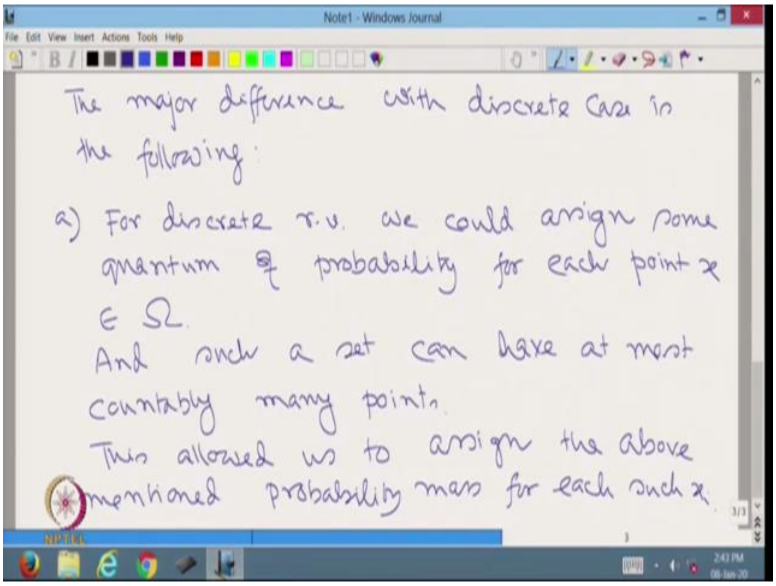
Welcome students to the MOOC's lecture series on Advanced Probability Theory. This is lecture number 9. In the last few classes, we have studied discrete random variables and in particular, we have seen Bernoulli, Binomial, Poisson, negative binomial and hypergeometric distributions, which are discrete distributions on a finite interval on the real line or the entire positive side of the real line that we have seen. In today's class, we shall start with continuous type random variables and that we will continue for next lecture as well.

(Refer Slide Time: 1:21)



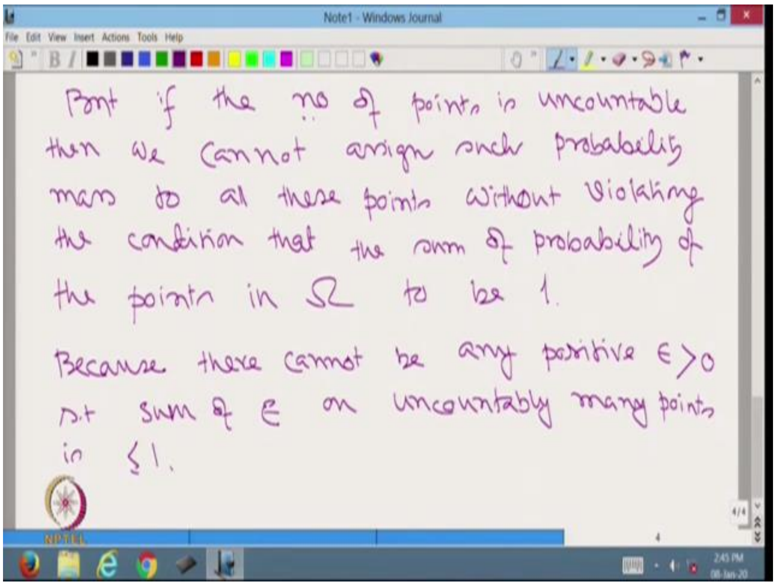
So, we are looking at continuous type random variables. Suppose, x is a random variable defined over some interval which may be finite or infinite, such that x can take any possible values belonging to that interval. Now, we know that any interval on R may have unaccountably many points. So, what difference does it make?

(Refer Slide Time: 3:23)



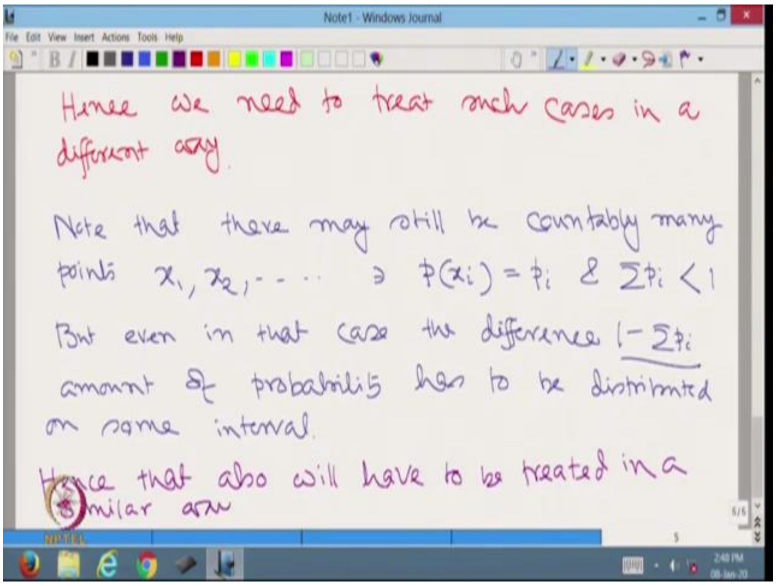
The major difference with discrete case is that, for discrete random variable we could assign some quantum of probability for each point x belonging to omega and such a set can have at most countably many points. So this allowed us to assign the above mentioned probability mass for each such xi.

(Refer Slide Time: 5:34)



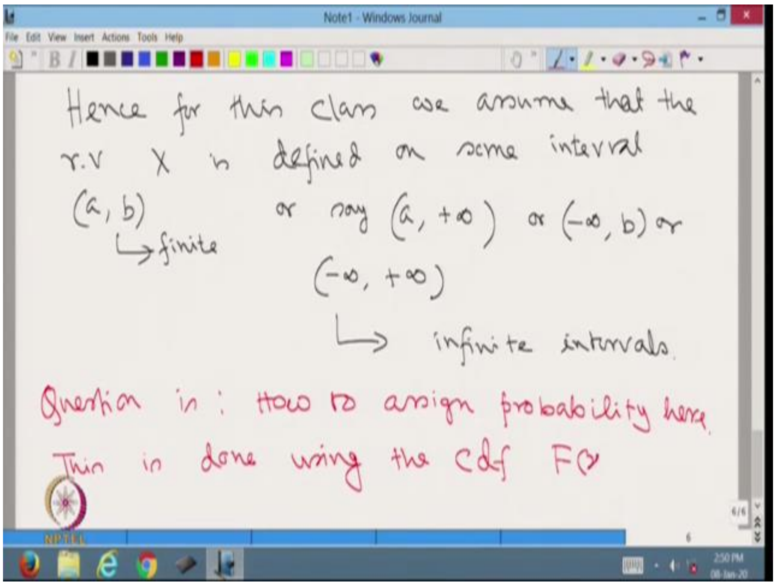
But, if the number of points is uncountable, then we cannot assign such probability mass to all these points without violating the condition that the sum of probabilities of the points in omega to be 1, because there cannot be any positive quantity or epsilon greater than 0, such that sum of an epsilon on uncountably many points is less than equal to 1.

(Refer Slide Time: 7:53)



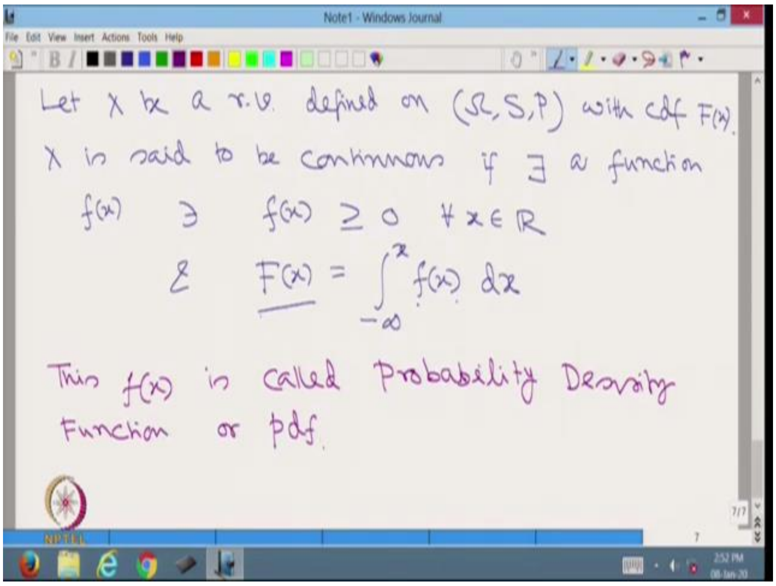
Hence, we need to treat such cases in a different way. Note that, there may be countably many points X1, X2… such that, probability of xi is equal to Pi and sigma Pi is less than 1, but even in that case the difference 1 - sigma Pi, this quantity has to be distributed on some interval. Hence, that also have to be treated in a similar way.

(Refer Slide Time: 10:13)



Hence, for this class, we assume that the random variable X is defined on some interval a to b which is finite or say some a to infinity or minus infinity to b or minus infinity to plus infinity, which are all infinite intervals. Question is, how to assign probability here? This is done using the cumulative distribution function Fx.

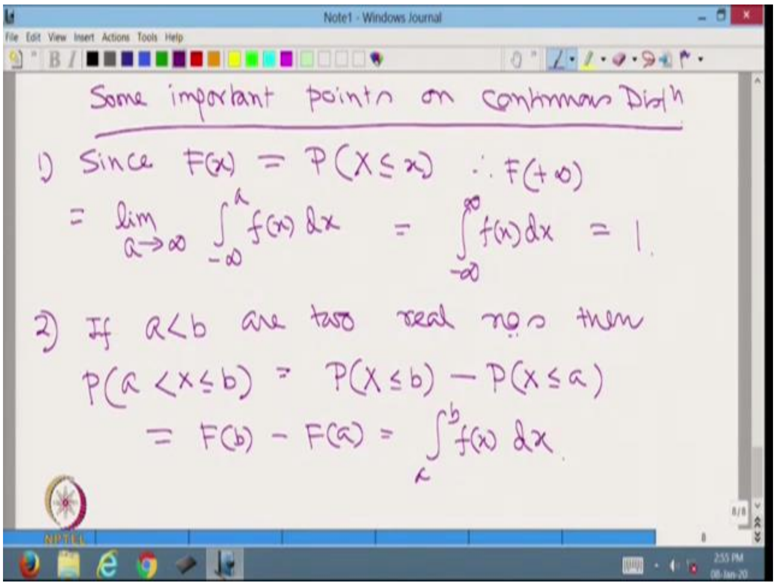
(Refer Slide Time: 12:13)



So, let X be a random variable defined on omega S, P, which we all know is the probability space with cdf Fx. X is said to be continuous, if there exists a function Fx such that, Fx greater than equal to 0 for all x belonging to R and Fx is equal to minus infinity to x Fx dx. That is, we get the cumulative distribution function by integrating the function small fx in the range minus infinity to x. This small fx is called probability density function or pdf.

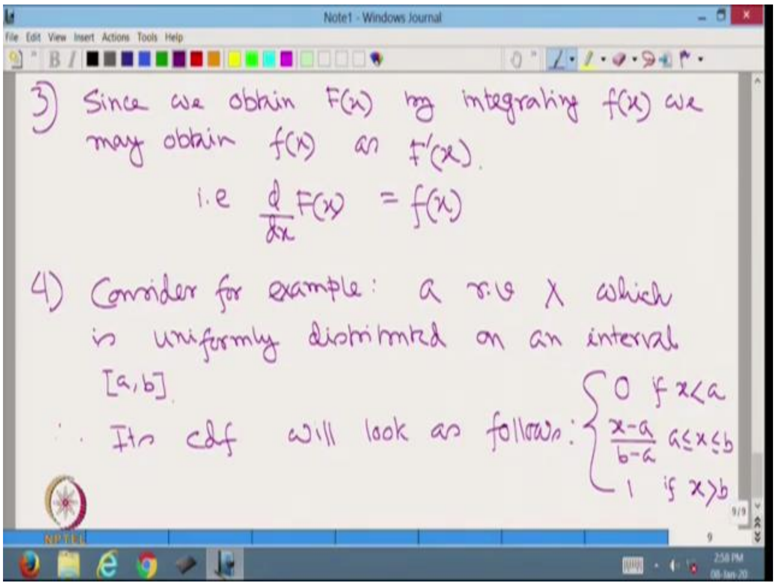
So, the major difference between probability mass function and probability density function is that, probability mass function associates a quantum of probability to the corresponding point x. Here, we are not assigning probability to any particular point x, but we are looking at the event that random variable X is taking value less than equal to x. And we are computing that probability by integration, by integrating the probability density function, small fx.

(Refer Slide Time: 14:57)



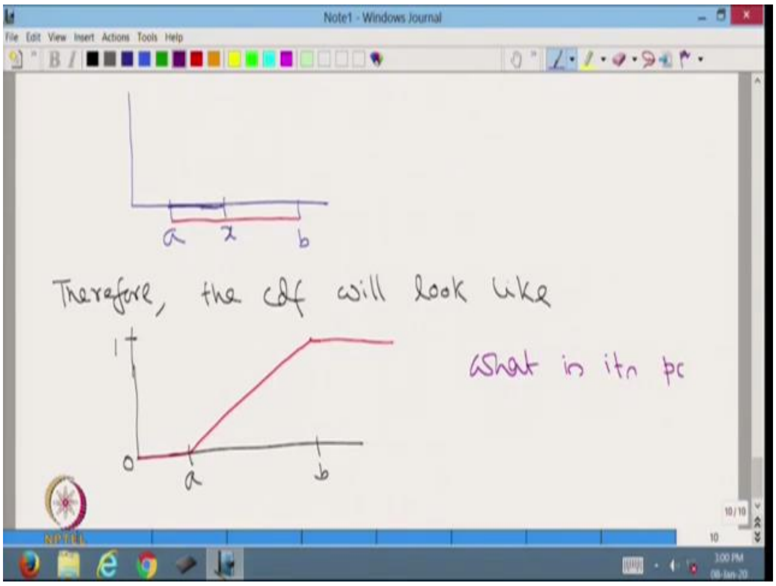
Now, let us discuss some important points on continuous distribution. One, since Fx is equal to probability X less than equal to x therefore, F of plus infinity is equal to limit a going to infinity minus infinity to a fx dx is equal to, we can write as minus infinity to infinity fx dx is equal to 1. Two, if a less than b are two real numbers, then probability a less than x less than equal to b is equal to probability x less than equal to b minus probability x less than equal to a is equal to F b minus F a is equal to a to b fx dx.

(Refer Slide Time: 16:51)



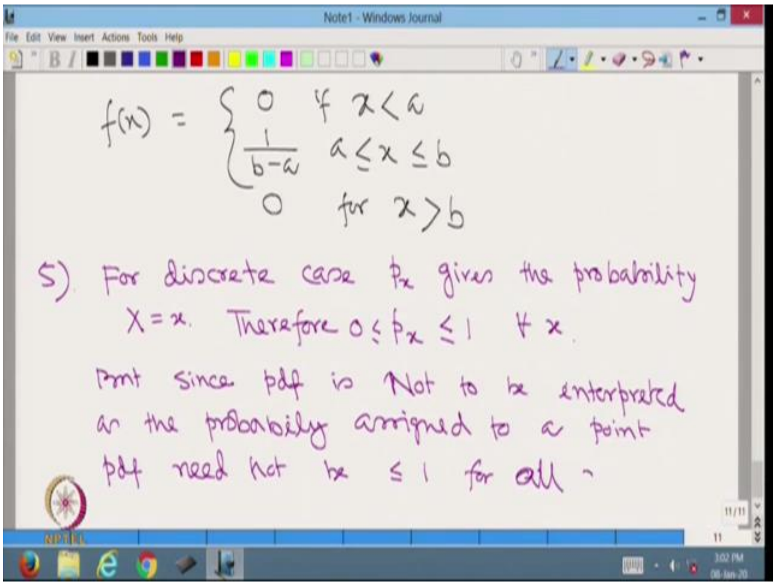
Three, since we obtain Fx by integrating small fx, we may obtain fx as the derivative of the cumulative distribution function capital Fx, that is d dx of Fx is equal to small fx. Four, consider for example, a random variable X which is uniformly distributed on an interval a to b. Therefore, its cdf will look like it is 0 if x less than a, it is x minus a upon b minus a for a less than equal to x, less than equal to b and it is 1 if x greater than b.

(Refer Slide Time: 19:09)



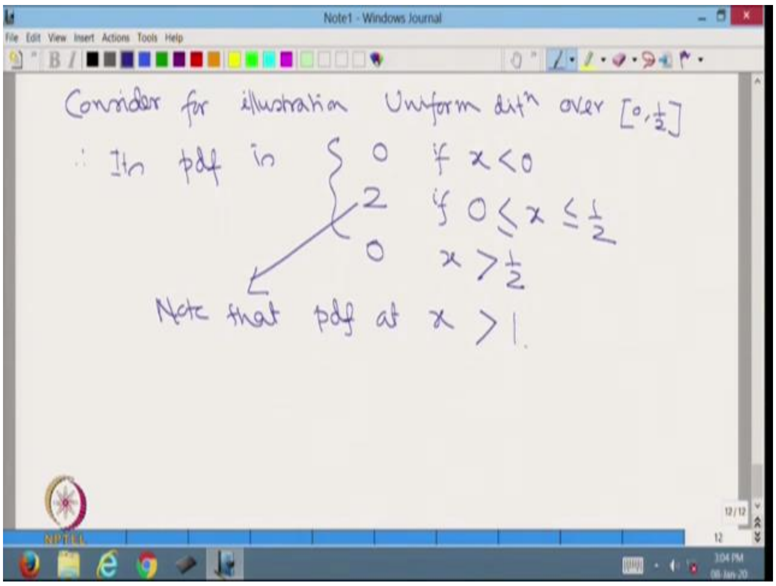
This is very obvious, because suppose this is my a and this is my b, therefore, for any point x the probability that the random variable X less than equal to this x is as we know that, is length of this interval upon length of the total interval. Therefore, the cdf will look like, below A it is going to be 0, from a to b, it is linearly increased to 1 and after b it will remain constant at 1. Question is what is its pdf?

(Refer Slide Time: 20:38)



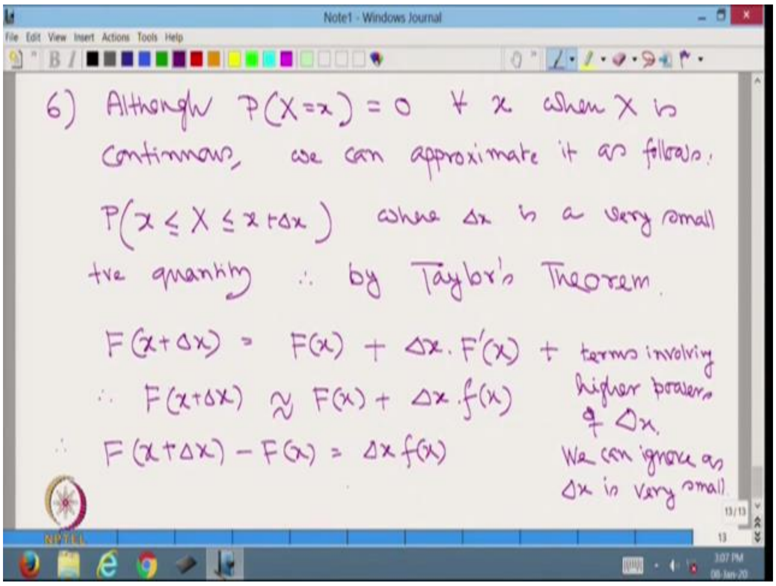
Very obvious, f of x is going to be 0 if x less than a, this is going to be 1 upon b minus a for a less than equal to x less than equal to b and 0 for x greater than b. For discrete case, P of x gives the probability X is equal to small x. Therefore, P x is less than equal to 1, greater than equal to 0 for all x, but, since pdf is not to be interpreted as the probability assigned to a point, pdf need not be less than equal to 1 for all x.

(Refer Slide Time: 22:29)



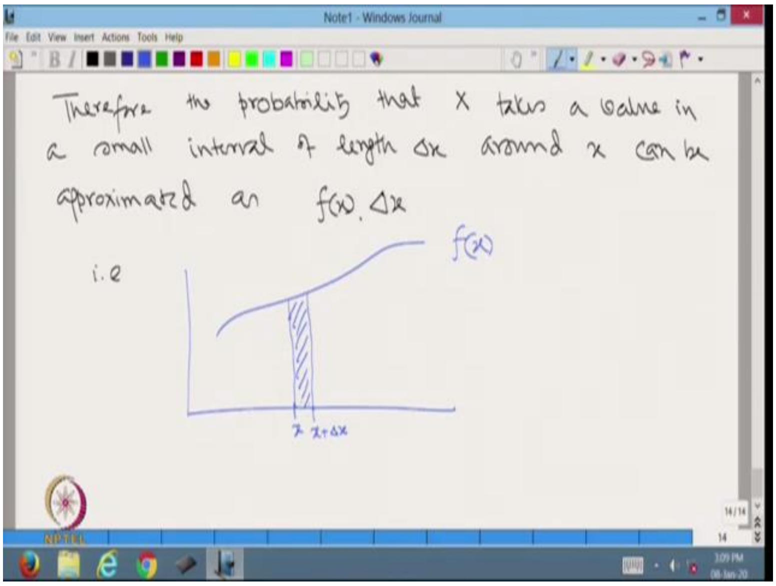
Consider for illustration, uniform distribution over 0 to half. So, since it is uniform over 0 to half, therefore, its pdf is 0 if x less than 0, it is 2 if 0 less than equal to x less than equal to half and is 0 if x is greater than half. So, note that pdf at x is greater than 1.

(Refer Slide Time: 23:38)



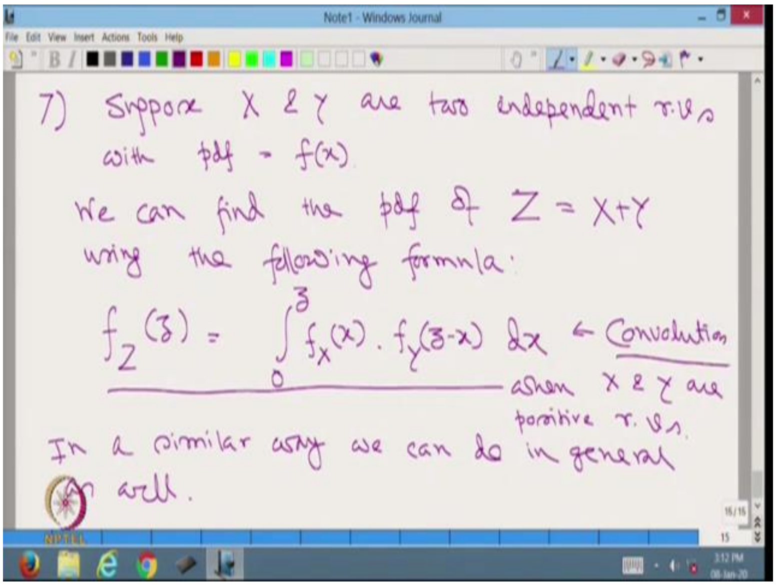
Six, although probability X is equal to x is equal to 0 for all x. When x is continuous, we can approximate it as follows. Probability X less than equal to x less than equal to x plus delta x, where delta x is a very small positive quantity. Therefore, by Taylor's theorem F at x plus delta x is equal to Fx plus delta x times F prime x plus terms involving higher powers of delta x. So, we can ignore them as delta x is very small. Therefore, F x plus delta x can be approximated as a Fx plus delta x times smaller fx, or F of x plus delta x minus Fx is equal to delta x times fx.

(Refer Slide Time: 26:09)



Therefore, the probability that x takes a value in a very small interval of length delta x around x can be approximated as fx times delta x, that is, suppose this is the curve of fx and here it is x, here it is x plus delta x. Therefore, this area will approximate the probability of X taking a value in the small interval x to x plus delta x.

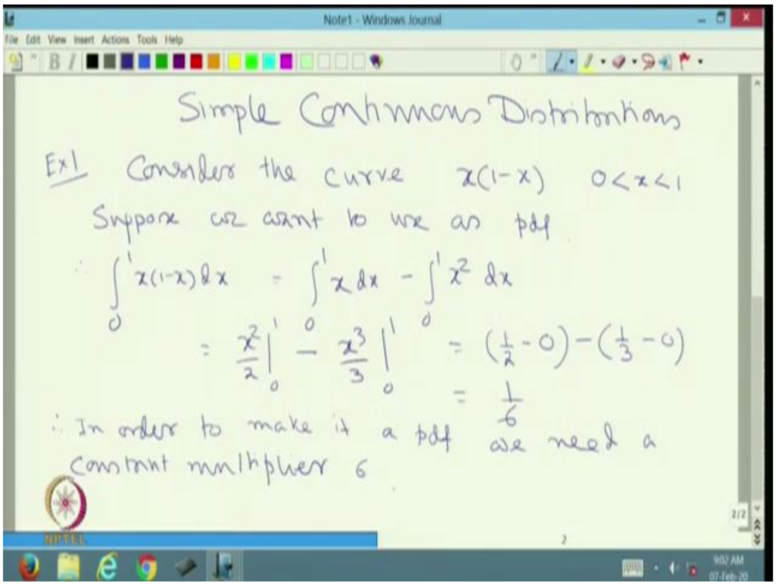
(Refer Slide Time: 27:42)



Seven, suppose x and y are two independent random variables with pdf is equal to f of x, we can find the pdf of Z is equal to x plus y using the following formula, F at Z at a point z is equal to integration 0 to z fx at x, f y at z minus x dx when x and y are positive random variables. In a similar way, we can do in general as well. So, this technique is not something very new for that one.

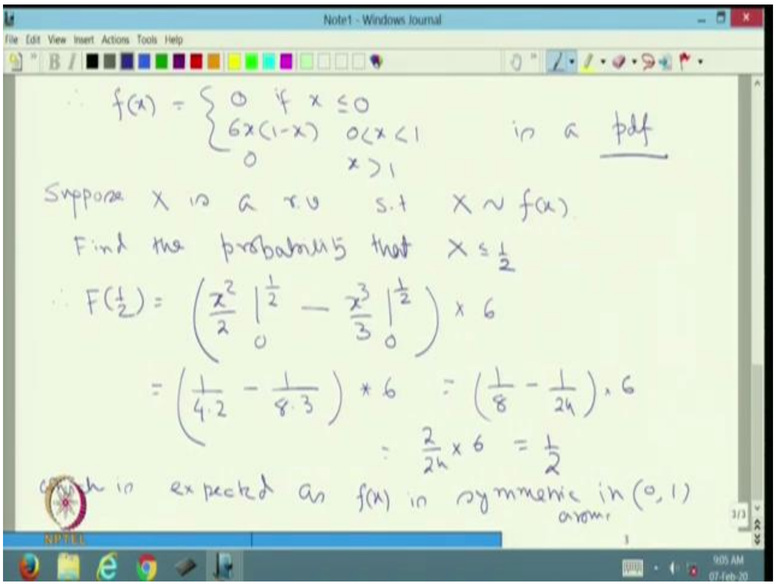
If you remember with respect to two binomial random variables x and y, when x is binomial in P and y is binomial in P, what is the probability that X plus Y takes a value z. We have summed over all possible combinations of X and why, such that their sum is equal to Z. The same thing we have done here and this technique is called convolution. We shall see some examples of convolution in subsequent part of this lecture or in the next lecture.

(Refer Slide Time: 30:30)



With that, let us study some simple continuous distributions. So, as you can understand, there can be any number of distributions, we shall study them as different examples. So, example 1, consider the curve x into 1 minus x when 0 less than x less than 1. Suppose, we want to use it as a pdf, therefore, we integrate x into 1 minus x dx in the range 0 to 1. So, this is equal to 0 to 1 x dx minus 0 to 1 x square dx is equal to x square by 2, 0 to 1, minus x cube by 3, 0 to 1 is equal to half minus 0 minus one third minus 0 is equal to 1 by 6. Therefore, in order to make it a pdf, we need a constant multiplier 6.

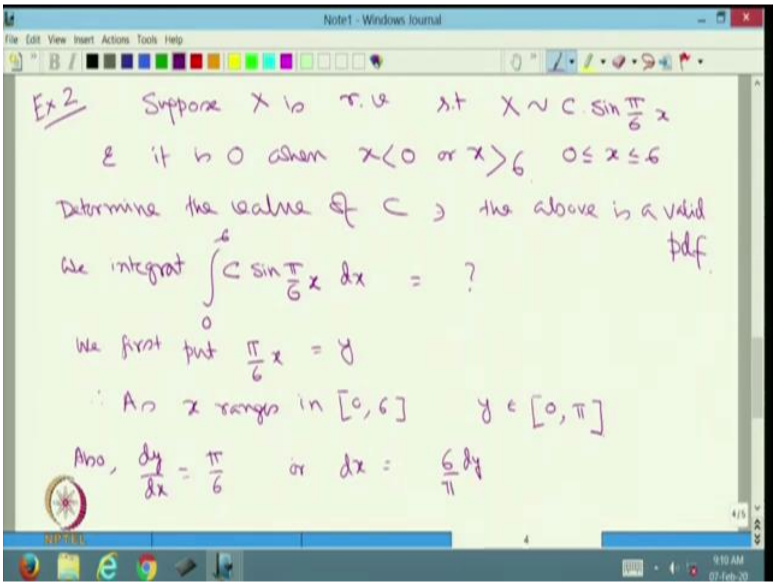
(Refer Slide Time: 32:46)



Therefore, fx is equal to 0, if x is less than equal to 0, 6x into 1 minus x, if 0 less than x less than 1 and 0 if x greater than 1, is a pdf or probability density function. Suppose, X is a random variable, such that X is distributed as the above fx, find the probability that X less than equal to half. Therefore, F of half is equal to x square by 2 in the range 0 to half minus x cube by 3, in the range 0 to half multiplied by 6.

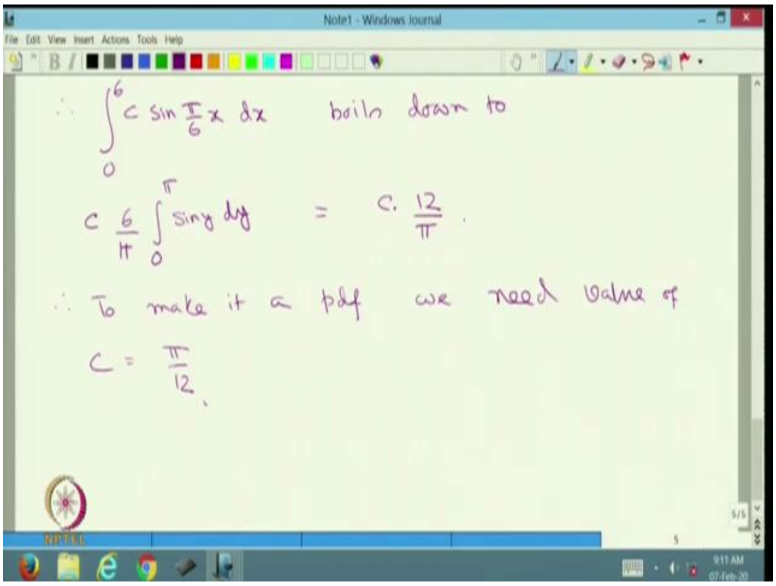
This is equal to, 1 by 4 into 2 minus 1 by 8 into 3 multiplied by 6, which is equal to 1 by 8 minus 1 by 24 multiplied by 6, is equal to 2 by 24 into 6, is equal to half. Which is expected as fx is symmetric in 0 to 1 around half.

(Refer Slide Time: 35:18)



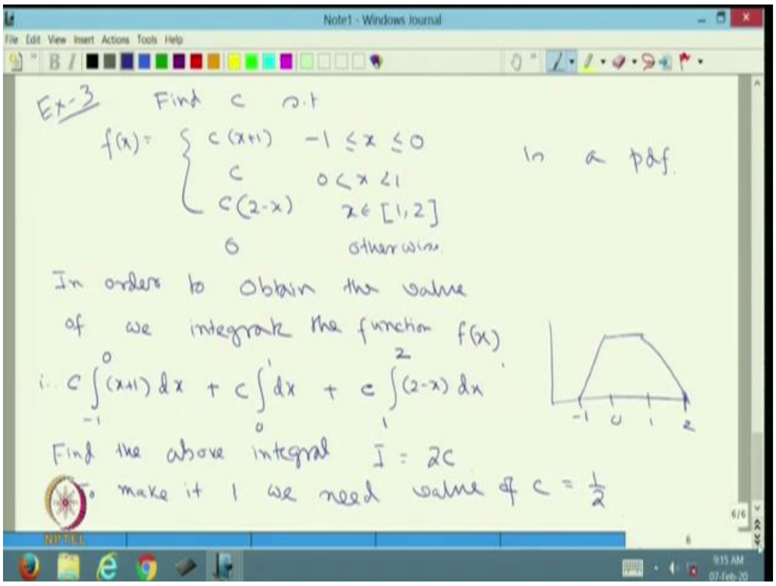
Example 2, suppose X is a random variable, such that X follows the following density function, C sin pi by 6x for 0 less than equal to x less than equal to 6, outside that range it is 0 when x is less than 0 or x is greater than 6. Determine C, such that the above is a valid pdf. So, what we do? We integrate C sin pi by 6x dx in the range 0 to 6 is equal to. So we first put pi by 6 x is equal to y. Therefore, as x ranges in 0 to 6, y belongs to 0 to pi. Also dy dx is equal to, therefore dy dx is equal to pi by 6 or dx is equal to 6 dy upon pi.

(Refer Slide Time: 37:39)



Therefore, integration 0 to 6 C sin pi by 6x dx boils down to C 6 by pi integration 0 to pi sin y dy which is, is equal to if you simplify, coming out to be C into 12 upon pi. Therefore, to make it a pdf we need value of C is equal to pi by 12.

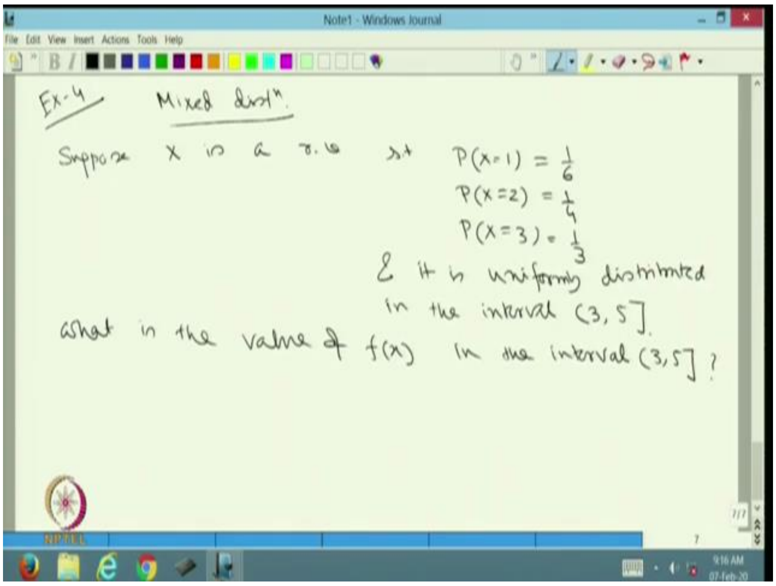
(Refer Slide Time: 38:48)



Example 3, find C such that fx is equal to C into x plus 1 for minus 1 less than equal to x less than equal to 0, C for 0 less than x less than 1 and C into 2 minus x for x belonging to 1, 2 and it is 0 otherwise. This function fx is a pdf, you can easily draw the function which should look like, something like, it should be linear here, it should be constant here and it will again linearly fall down to 2.

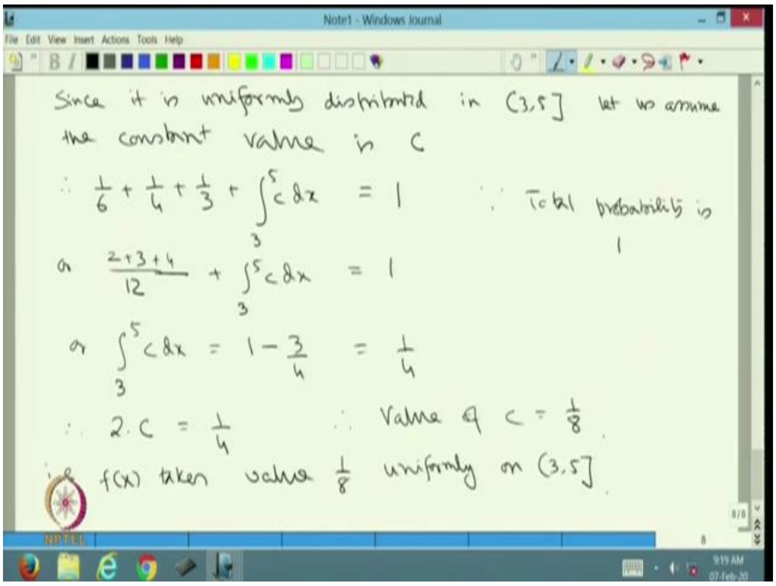
In order to obtain the value of C, we integrate the function fx, that is C into integration minus 1 to 0, x plus 1 dx plus C into 0 to 1 dx plus C into 2 minus x dx, for x belonging to 1 to 2. Find the above integral and we will get this is equal to, if I call it I 2C. Therefore, to make it 1, we need value of C is equal to half.

(Refer Slide Time: 41:44)



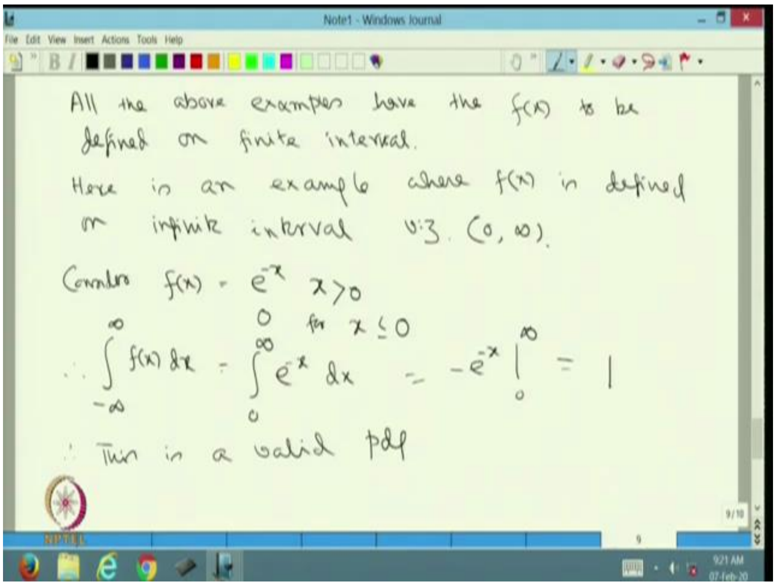
Example 4. Mixed distribution suppose X is a random variable, such that probability X is equal to 1 is equal to 1 by 6, probability X is equal to 2 is equal to 1 by 4, probability X is equal to 3 is equal to 1 by 3, respectively. And it is uniformly distributed in the interval 3 to 5, what is the value of fx in the interval 3 to 5? That is the question.

(Refer Slide Time: 43:11)



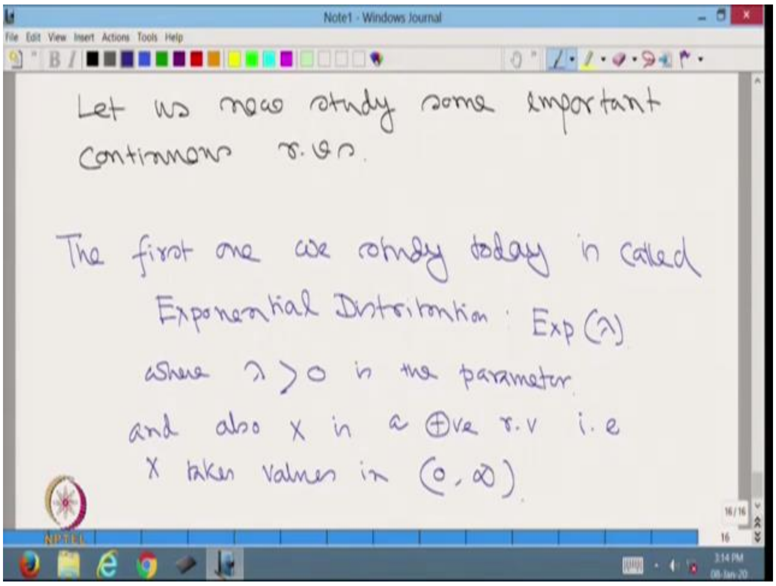
Since, it is uniformly distributed in 3 to 5, let us assume the constant value that it takes is C, therefore, 1 by 6 plus 1 by 4 plus 1 by 3 plus integration 3 to 5 C dx is equal to 1, since total probability is 1 or 2 plus 3 plus 4 upon 12 plus integration 3 to 5 C dx is equal to 1 or integration 3 to 5, C dx is equal to 1 minus 3 by 4 is equal to 1 by 4. Therefore, 2 C is equal to 1 by 4. Therefore, value of C is equal to 1 by 8, that is fx takes value 1 by 8 uniformly on 3 to 5.

(Refer Slide Time: 45:12)



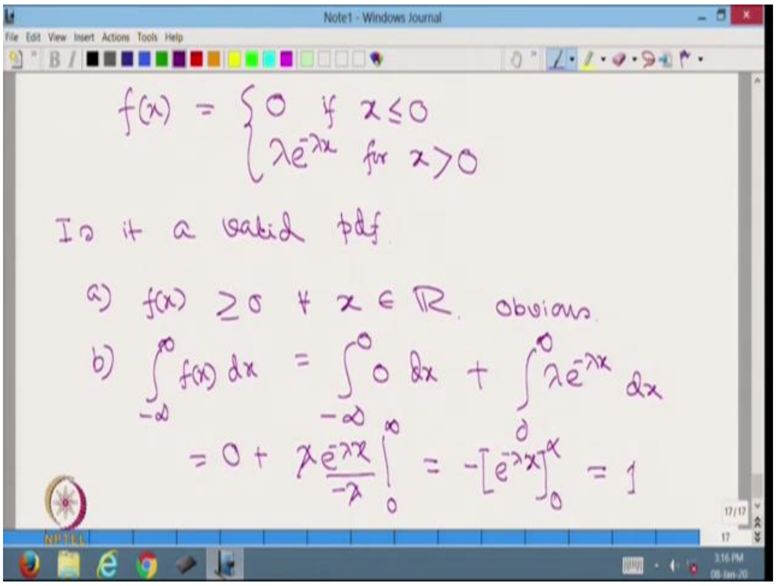
Now, all the above examples have the fx defined on finite intervals. Here is an example, where fx is defined on infinite interval, namely 0 to infinity. So, consider fx is equal to e to the power minus x when x greater than 0 and 0 for x less than equal to 0. Therefore, integration minus infinity to infinity fx dx is equal to 0 to infinity e to the power minus x dx is equal to minus e to the power minus x, 0 to infinity is equal to 1. Therefore, this is a valid pdf.

(Refer Slide Time: 47:06)



With this background, let us now study some important continuous random variables. The first one we study today is called exponential distribution or denoted as Exp lambda, where lambda greater than 0 is the parameter and also X is a positive random variable. That is X takes value in 0 to infinity.

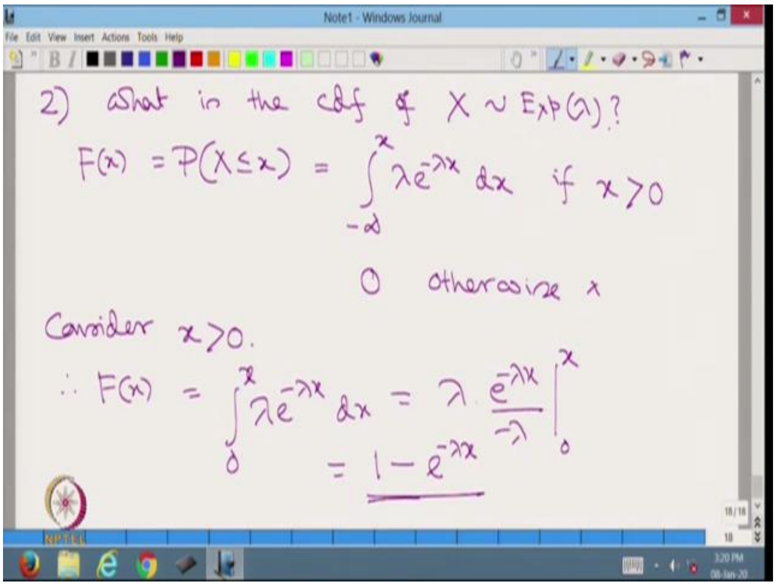
(Refer Slide Time: 48:46)



f of x is equal to 0, if x less than equal to 0 and it is lambda e to the power minus lambda x for x greater than 0. Is it a valid pdf? A fx is greater than equal to 0 for all x belonging to R. This is obvious, because when x is greater than 0 and lambda is positive, lambda into e to the power minus lambda x will always remain positive. Next we need to check whether this integrates to 1 or not. Therefore, minus infinity to infinity fx dx is equal to minus infinity to 0, 0 dx plus 0 to infinity lambda e to the power minus lambda x dx.

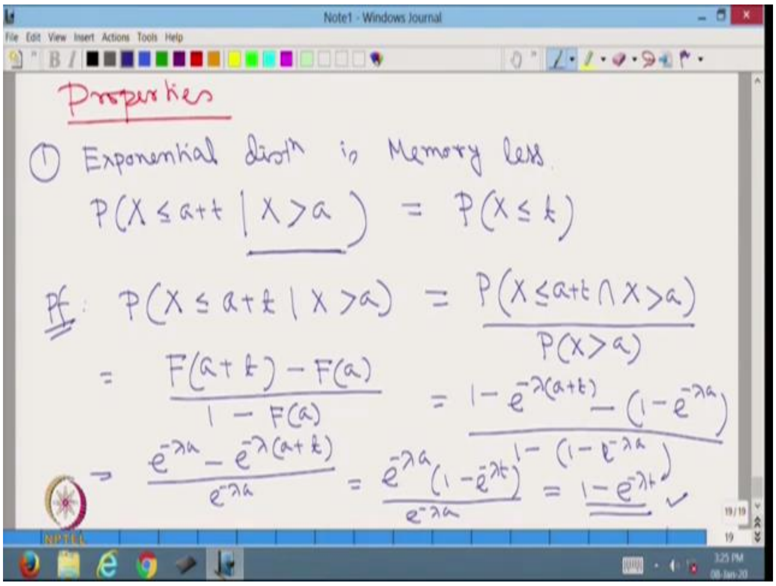
Which is is equal to 0 plus lambda e to the power minus lambda x divided by minus lambda, this is going from 0 to infinity, which is equal to lambda cancels with minus lambda remains the minus, e to the power minus lambda at plus infinity and on this side it is 0. So, at infinity this is going to be 0, at 0 it is going to be 1, therefore, result is equal to 1. Hence, the above function is a valid probability density function.

(Refer Slide Time: 51:08)



Two, what is the cdf when x is distributed as exponential lambda. So, f of x is equal to probability x less than equal to x is equal to integration minus infinity to x lambda e to the power minus lambda x dx if x is greater than 0 or it is 0 otherwise. Therefore, this is of no interest, therefore, we look at x greater than 0, therefore, fx is equal to 0 to x lambda e to the power minus lambda x dx is equal to lambda e to the power minus lambda x upon minus lambda in the range 0 to x, which is coming out to be 1 minus e to the power minus lambda x. So, that is going to be the cumulative distribution function for an exponential random variable.

(Refer Slide Time: 52:40)



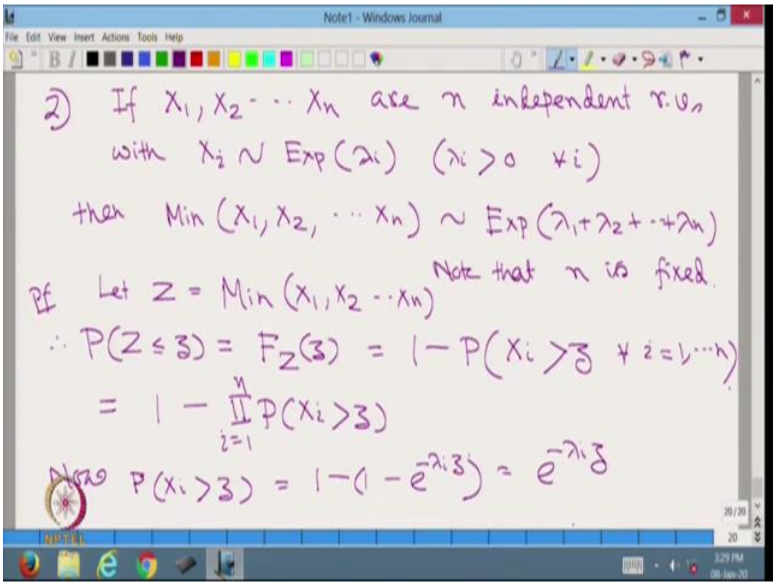
Properties, one, exponential distribution is memoryless. What does it mean? It means that probability X less than equal to a plus t given that X greater than a is equal to probability X less than equal to t. That means that if we know that x has value greater than a, then the probability that exists within a distance t from a that is same as probability that x is less than equal to t. That means, this knowledge that X is already greater than a has no effect on the subsequent value that X may take.

So proof, probability x less than equal to a plus t, given x greater than a, we know that from the knowledge of conditional probability, it is probability X less than equal to a plus t and x is greater than a divided by probability X is greater than a. Now, the numerator is we are saying that x is greater than a, but the X is less than equal to a plus t. Therefore, we can write it as F of a plus t minus F at a divided by 1 minus F a, because probability X greater than a is equal to 1minus- the cdf at a, that is 1 minus F a.

And this we know that is equal to 1 minus e to the power minus lambda a plus t minus 1 minus e to the power minus lambda a divided by 1 minus 1 minus e to the power minus lambda a, is equal to e to the power minus lambda a minus e to the power minus lambda a plus t upon e to the power minus lambda a, which is equal to, if we take e to the power minus lambda a to be common, 1 minus e to the power minus lambda t upon e to the power minus lambda a is equal to 1 minus e to the power minus lambda t.

This we know that is the probability that x is less than equal to t. So, we see that probability X less than equal to a plus t given that X is greater than a is same as probability X less than equal to t. Therefore, exponential distribution is called memoryless. So, it is very similar to what we have seen with respect to geometric distribution, when we were looking at discrete random variables.

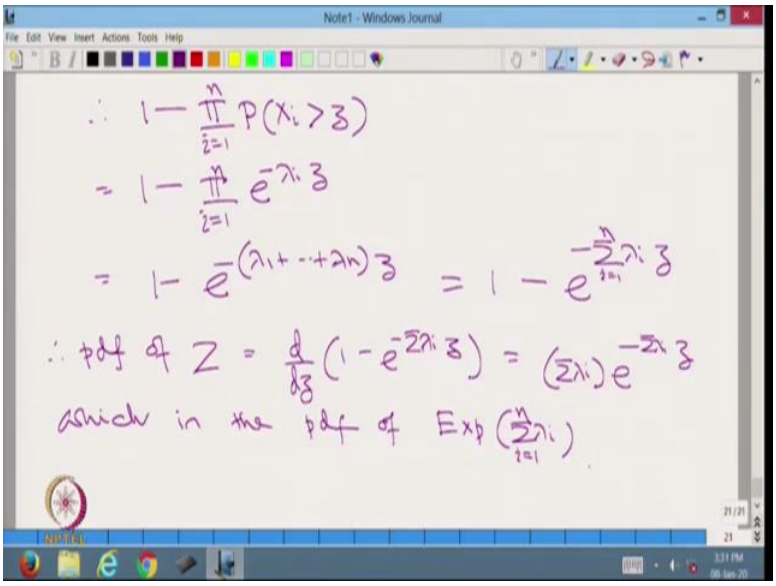
(Refer Slide Time: 57:16)



Property 2. If X1, X2, Xn are n independent random variables with Xi, distributed as exponential with lambda i, lambda i greater than 0 for all i, then minimum of X1, X2, Xn is distributed as exponential with lambda 1 plus lambda 2 plus lambda n. Note that n is fixed. So, proof let Z equal to minimum of X1, X2, Xn. Therefore, probability Z less than equal to z is equal to F Z at z there is the cumulative distribution function for z is equal to 1 minus probability Xi greater than z for all i is equal to 1 to n.

Which is equal to 1 minus product of probability Xi greater than z, i is equal to 1 to n. Now, probability Xi greater than z is equal to 1 minus 1 minus e to the power minus lambda i z is equal to e to the power minus lambda i Z.

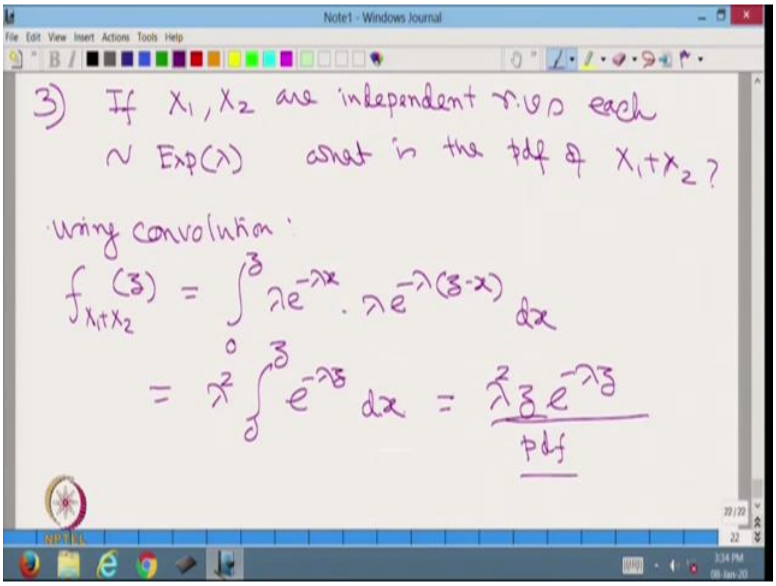
(Refer Slide Time: 59:58)



Therefore, 1 minus product i is equal to 1 to n probability Xi greater than z is equal to 1 minus product of i is equal to 1 to n, e to the power minus lambda i z is equal to 1 minus e to the power minus lambda 1 plus lambda 2 up to lambda n Z is equal to 1 minus e to the power minus i is equal to 1 to n sigma lambda i times z. Therefore, pdf of Z is equal to d dz of 1 minus e to the power minus sigma lambda i z is equal to sigma lambda i into e to the power minus sigma lambda i times z, which is the pdf of the exponential with parameter sigma lambda i, i is equal to 1 n.

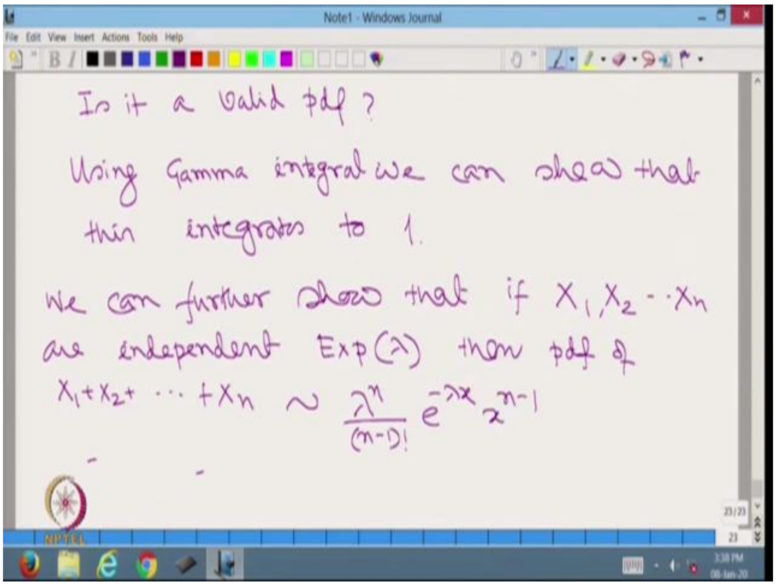
Thus if X1, X2, Xn are independent exponential random variables with the parameters lambda 1, lambda 2, lambda n respectively, then the minimum of them will have the distribution exponential sigma lambda i.

(Refer Slide Time: 61:52)



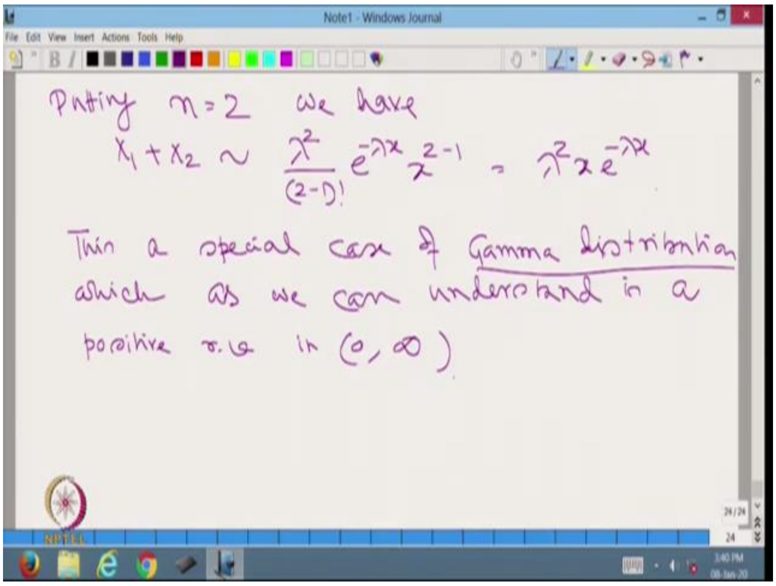
Point 3, if X1 and X2 are independent random variables, each distributed as exponential with lambda, what is the pdf of X 1 plus X 2? So, this is the problem I mentioned some time back, so I will do it using convolution. F of X 1 plus X 2 at a point z is equal to integration 0 to z lambda e to the power minus lambda x multiplied by lambda, e to the power minus lambda z minus x dx, is equal to lambda square 0 to z e to the power minus lambda z dx is equal to lambda square z e to the power minus lambda z. Thus we get a new type of density function whose pdf is lambda square z into e to the power minus lambda z.

(Refer Slide Time: 63:54)



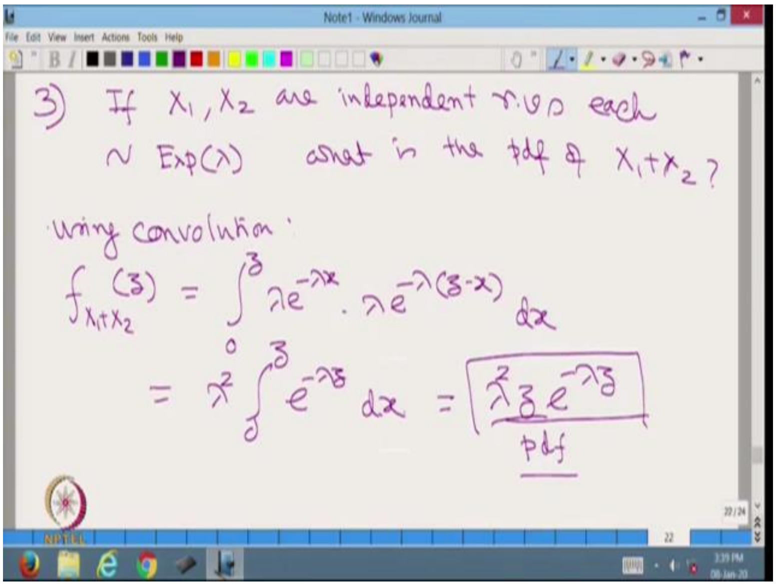
Is it a valid pdf function? Using Gamma integral, we can show that this integrates to 1. Not only that, we can further show that if X1, X2, Xn are independent exponential lambda then pdf of X1 plus X2 plus Xn will be distributed as lambda to the power n upon n minus 1 factorial e to the power minus lambda x, x to the power n minus 1.

(Refer Slide Time: 65:28)



Putting n is equal to 2, we have x 1 plus x 2 is distributed as lambda to the power 2 upon 2 minus 1 factorial e to the power minus lambda x, x to the power 2 minus 1, which is equal to lambda square x into e to the power minus lambda x.

(Refer Slide Time: 67:11)



If we look at it, we see that that is the pdf that we have got by using convolution. So, this is a special case of gamma distribution which as you can understand is a positive random variable in 0 to infinity. In the next class, I shall start with gamma distribution in a very general way. Till then, thank you so much.