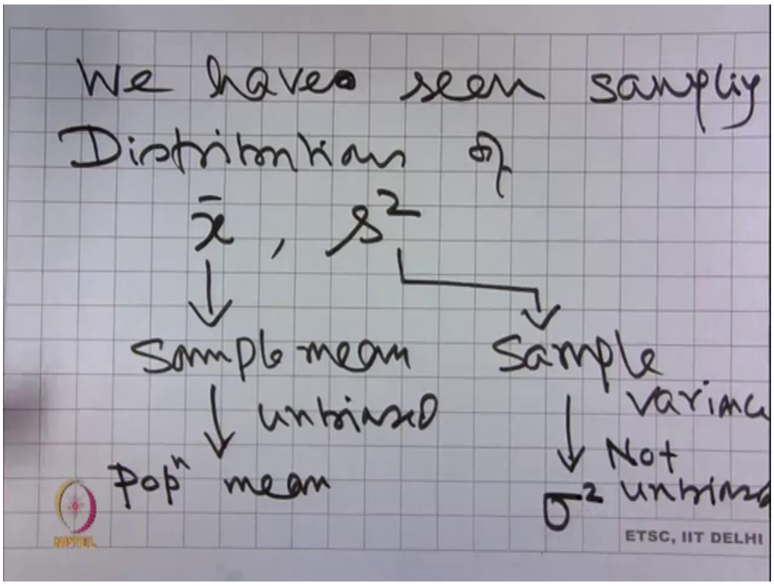
Statistical Inference
Prof. Niladri Chatterjee
Department of Mathematics
Indian Institute of Technology, Delhi

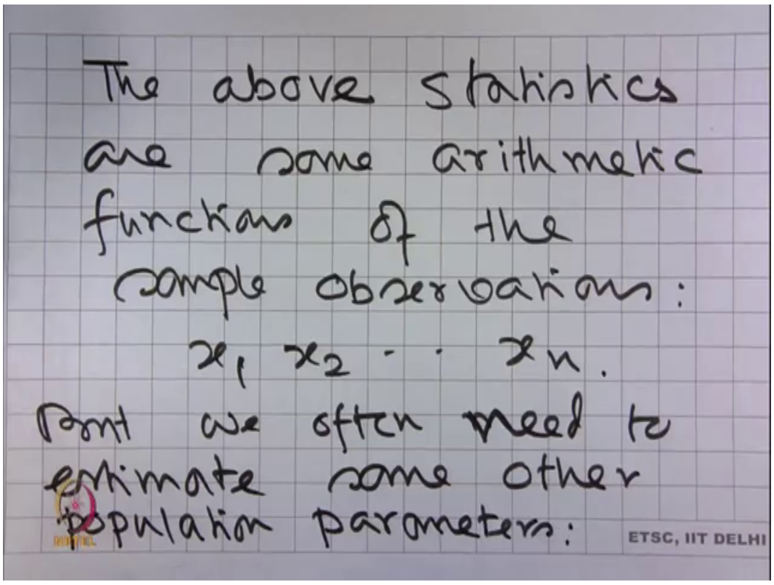
Lecture – 09
Statistical Inference

(Refer Slide Time: 00:38)



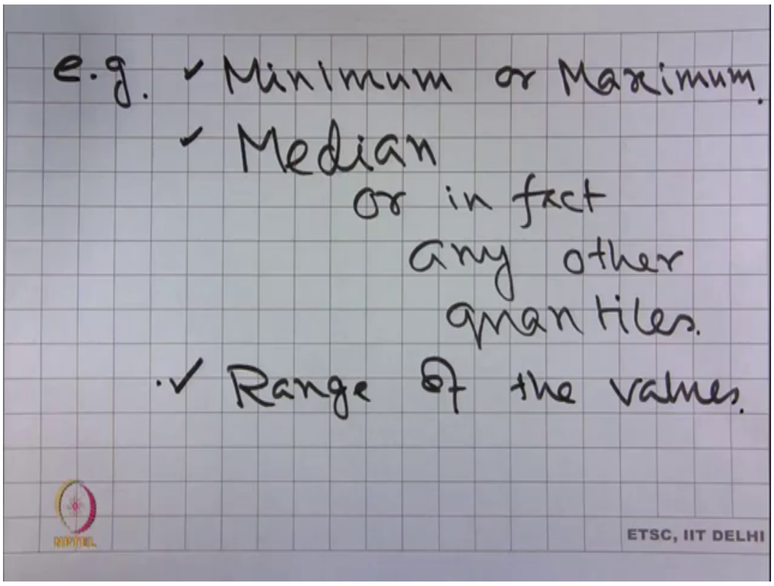
Welcome students to the MOOCs series of lectures on Statistical Inference. And this is the 9th lecture of the series. If you remember in the previous classes, we have seen sampling distributions of say x bar s square, this is the sample mean, this is sample variance. And we have seen that this is unbiased for population mean. This is not unbiased for sigma square, this is not unbiased for sigma square, which is the population variance, but a constant multiplier of sample variance will give an unbiased estimator for sigma square.

(Refer Slide Time: 01:55)



So, if you observe, we see that the above statistic are some arithmetic functions of the sample observations; x 1, x 2 up to x n. But, we often need to estimate some other population parameters.

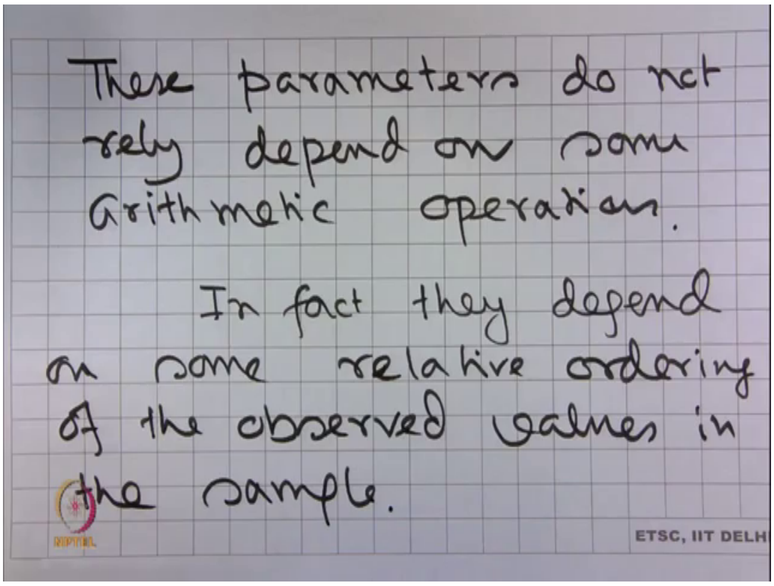
(Refer Slide Time: 03:07)



For example, say minimum or maximum. So, if we take a sample based on that can we estimate the minimum value of the attribute that we are interested in the population or say the maximum value of the attribute in the population or say median or in fact any other quantiles? For example, what is the 10 percentile of the population or what is the

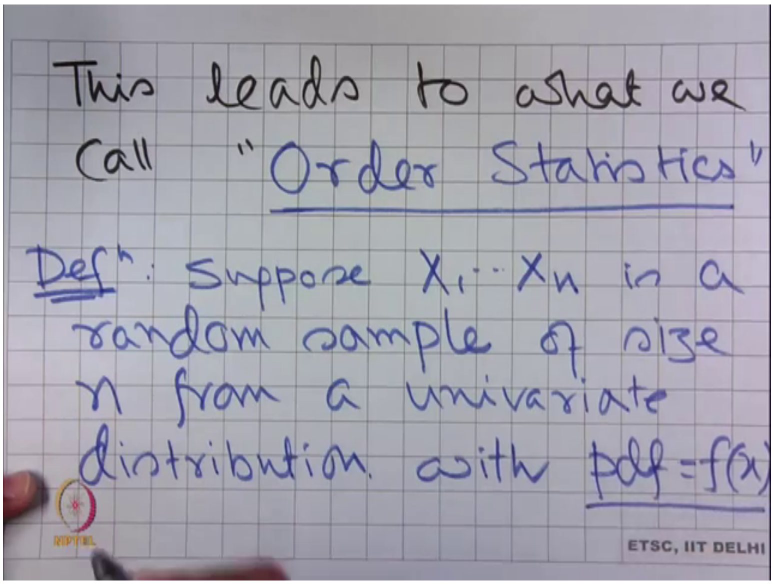
25 percentile, which is the first quartile of the population. So, we may like to estimate things like that range of the values say, what is the expected range of the attribute in the population that is the maximum minus the minimum.

(Refer Slide Time: 04:43)



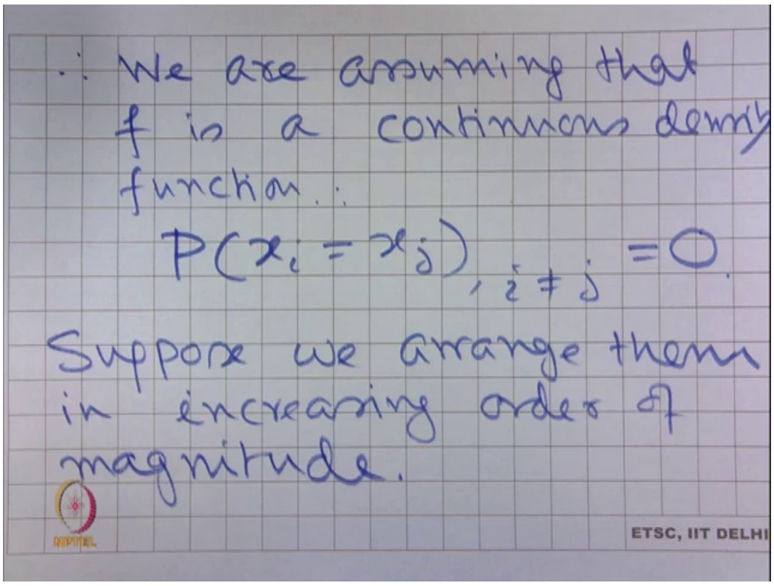
So, these parameters do not really depend on some arithmetic operations. In fact, they depend on some relative ordering of the values in the sample. So, basically you have observed n samples x 1, x 2, x n. And if we find a relative ordering among those values to some extent, we can feel that they may give us clue about what can possibly be the expected value of these parameters that we have just mentioned.

(Refer Slide Time: 06:22)



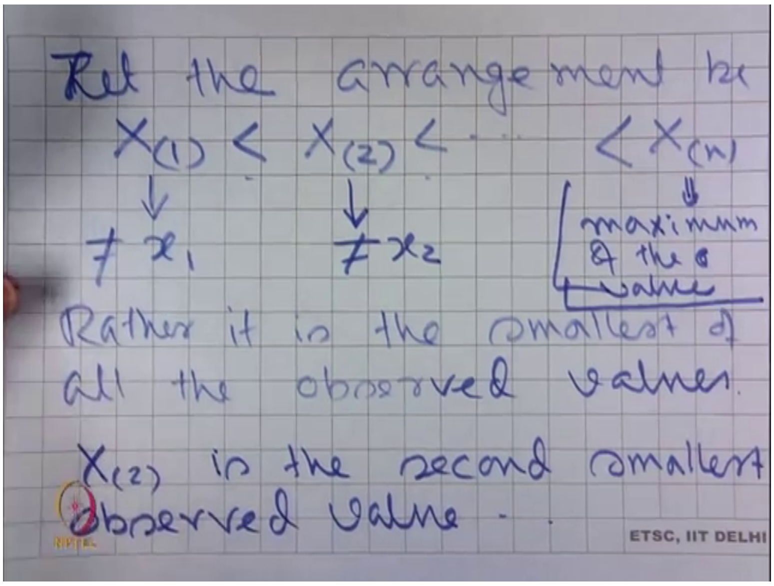
So, this leads to what is called order statistic. As the name suggests, it is not only one statistic, but there are different statistics involved. And they are based on the relative ordering of the values in the sample. So, definition suppose x 1, x 2, x n is a random sample of size n from a univariate distribution with pdf is equal to f x.

(Refer Slide Time: 08:13)



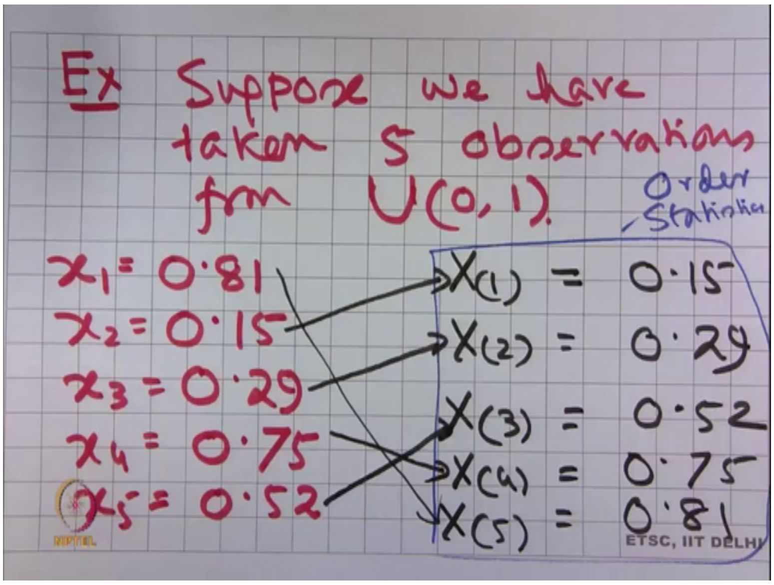
So, since we are talking about pdf, we can we are assuming that f is a continuous distribution continuous density function. What does it mean? It means that probability x i is equal to x j; i not equal to j is equal to 0. So, we have x 1, x 2, x n, n observation from a continuous density function f x. Suppose, we arrange them in increasing order, so we are arranging them in increasing order of magnitude.

(Refer Slide Time: 09:37)



Let the arrangement be X 1, I am using strictly less than, because we assumed that two sample values cannot be equal that probability is 0. Therefore, we can get a relative order of the sample. So, what is X 1? You remember this parenthesis, this is not equal to x 1, the first observed sample rather it is the smallest of all the observed values. X 2 is similarly not necessarily equal to x 2; X 2 is the second smallest observed value etcetera. And X n is the maximum of the of the values ok.

(Refer Slide Time: 11:39)

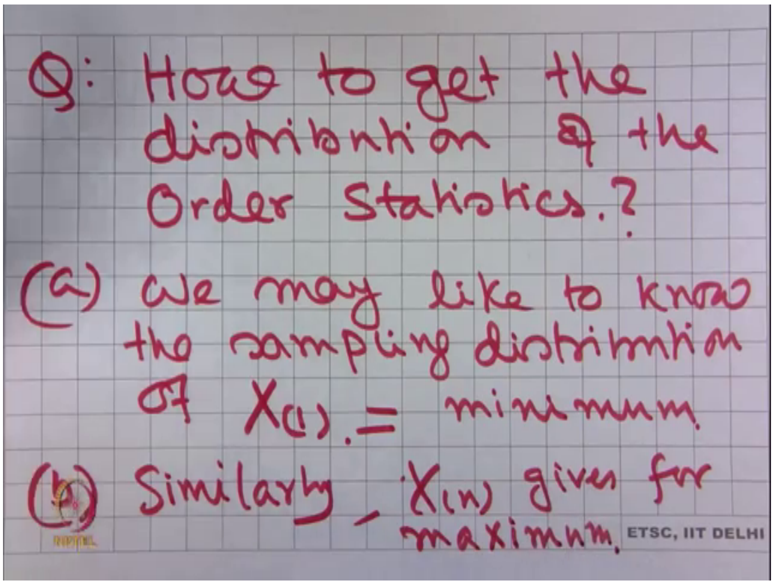


So, let me give you an illustration. Suppose, we have taken 5 observations from Uniform 0, 1; that means, from the real line interval 0 to 1, we have taken 5 values. Let x 1 be 0.81, x 2 is equal to 0.15, x 3 is equal to say 0.29, x 4 is equal to 0.75 and x 5 is equal to say 0.52, as you can see that they are not in any sorted order.

So, from this sample, we can get X 1 is equal to 0.15, X 2 is equal to 0.29, X 3 is equal to 0.52, X 4 is equal to 0.75 and X 5 is equal to 0.81. Therefore, you can understand that once we sort these values, we get a particular arrangement of the observed values. So, this is the order statistic generated from this sample. In fact, this is not the only sample that generates this.

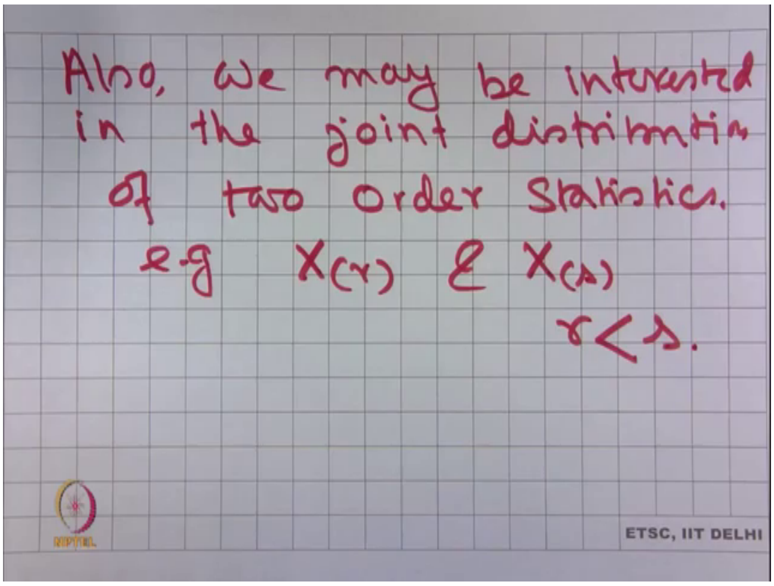
In fact, since there are five observations, we can have factorial five many different orderings of the observation each of which will give rise to the same order statistic. Therefore, what is an order statistic? Given a sample of n observations from a distribution function say f x, when we arrange the observed values in increasing order, where x subscript bracket i is the ith minimum in the arranged order. Then this sequence X 1, X 2, X n is the order statistic generated from that sample.

(Refer Slide Time: 15:20)



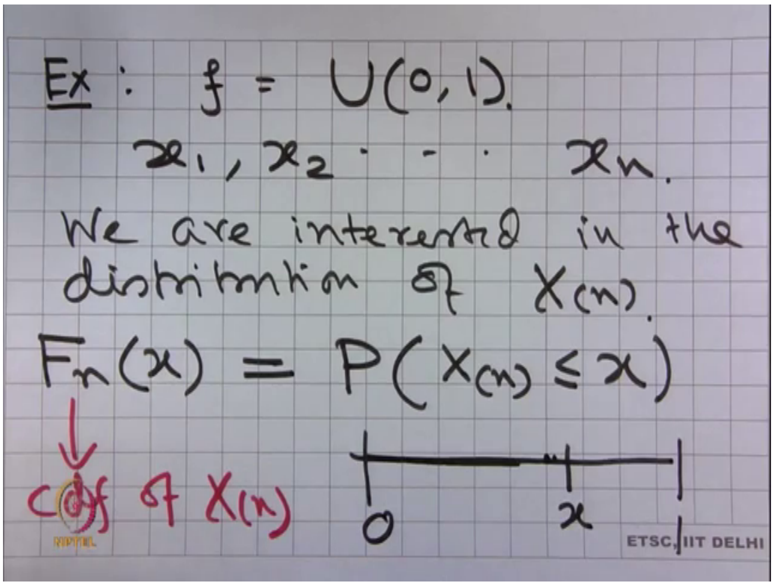
Question: How to get the distribution of the order statistics? That is the question. We may like to know the sampling distribution of X 1 or that is the minimum. Therefore, if we get the distribution of X 1 the first order statistic, then we can know or we can infer about the distribution of the minimum of the sample. Similarly, X n gives for maximum.

(Refer Slide Time: 17:05)



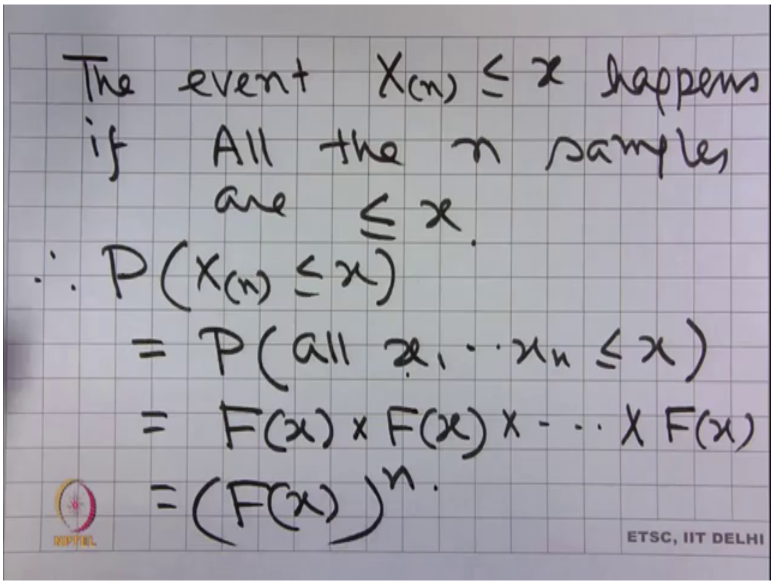
Also, We may be interested in the joint distribution of two order statistics e.g X r and X s where r less than s. Say for example, how the 3rd order statistic and the 8th order statistic in a sample of size say 10 are jointly distributed? We will see the applications of such distributions, but let me first illustrate with some simpler examples.

(Refer Slide Time: 18:16)



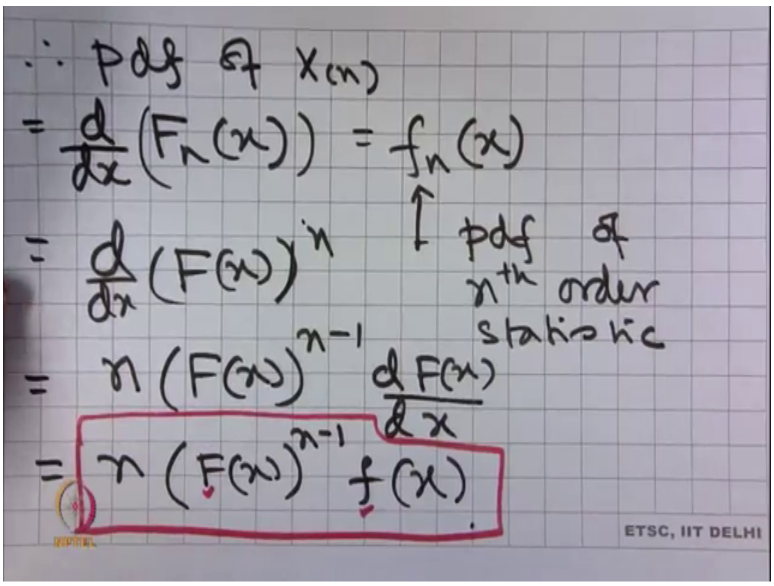
So, let us consider uniform 0, 1 distribution and we have observed n samples from this distribution. Suppose, we are interested in the distribution of X n, the nth order statistic. So, what we will do? We will first calculate the cumulative distribution function of the nth ordered statistic. So, this is the cdf of X n, we are denoting with F n ok. So, this is probability that the nth order statistic is less than equal to x. Suppose, this is 0, 1 and this is x and we have taken n samples and we want that the maximum of this; suppose this is the maximum, which is less than equal to x. What is the probability?

(Refer Slide Time: 20:16)



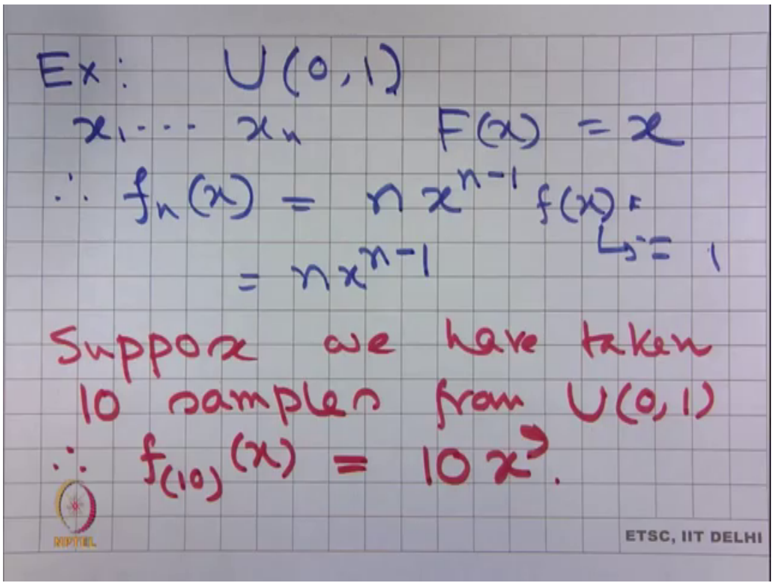
We know that the event X n less than equal to x happens. If when it will happen? That the maximum is less than equal to x that means that all the n samples are actually less than equal to x right. Therefore, probability therefore probability X n less than equal to x is equal to probability all x 1, x 2, x n less than equal to x. And what is the probability that any one of them is less than equal to x? That is F x. Similarly, what is x 2 less than equal to x? That is also F x into F x is equal to F x to the power n right. Therefore, the maximum will be less than equal to x if all the observations are less than equal to x and that probabilities F x, where f is the parent distribution whole to the power n.

(Refer Slide Time: 22:09)



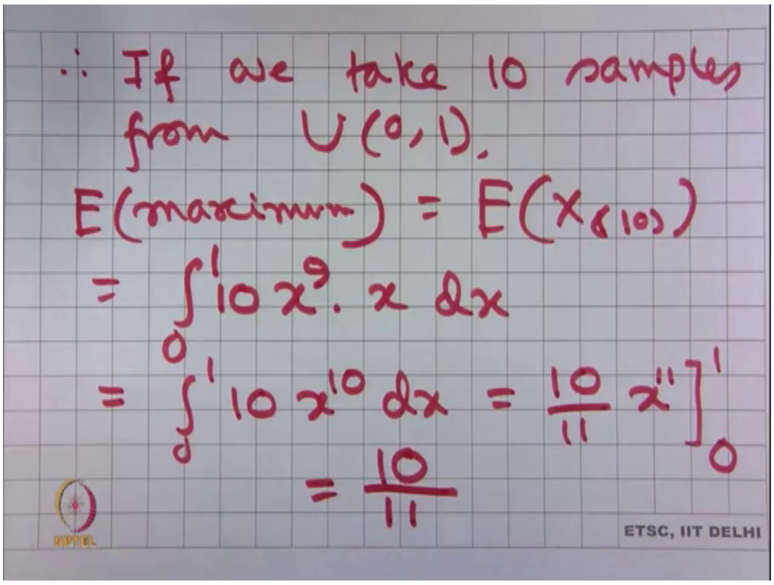
Therefore, if we differentiate the cumulative distribution with respect to x, we will get the F x or we denote it as F n x that is that pdf of nth order statistic, which is equal to d dx of F x power n, which is equal to n times F x to the power n minus 1 d F x dx, which is equal to n into F x to the power n minus 1 to f x. So, once we know the parent distribution at the parent density function and the corresponding cdf. Then if we take n samples, then we can get that pdf like this.

(Refer Slide Time: 23:56)



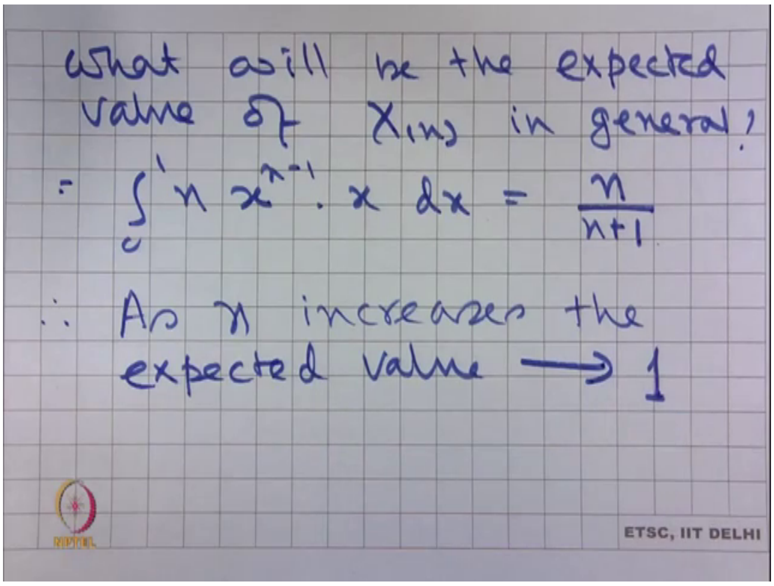
So, example as I said, I will take uniform 0, 1 n samples, F x is equal to x. Therefore, f n x is equal to n times x to the power n minus 1 into f x, which is equal to 1 is equal to n into x to the power n minus 1. So, suppose we have taken 10 samples from uniform 0, 1 one. Therefore, f 10 at x is equal to 10 into x to the power 9 right.

(Refer Slide Time: 25:26)



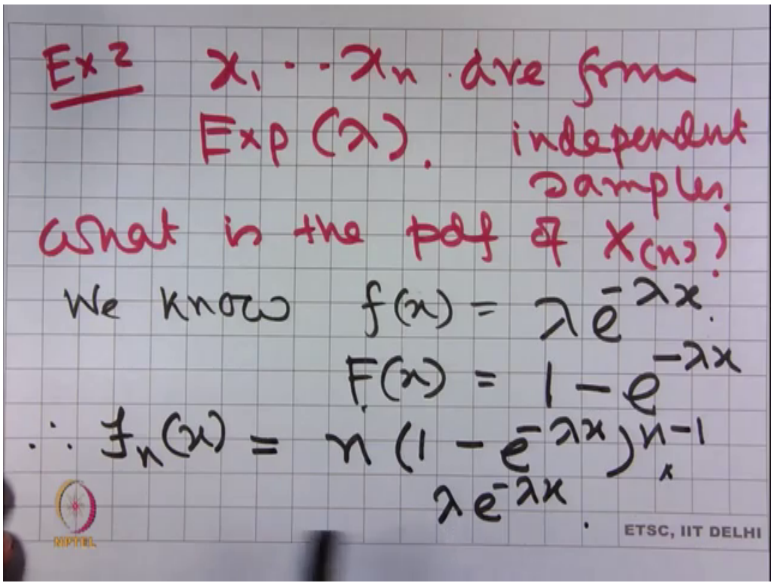
Or in other words, 10 samples from uniform 0, 1, then expected value of the maximum is equal to expected value of X 10 is equal to 0 to 1 10 x to the power 9 multiplied by x dx is equal to 0 to 1 10 x to the power 10 dx is equal to 10 upon 11 x to the power 11 1, 0 is equal to 10 upon 11.

(Refer Slide Time: 26:35)



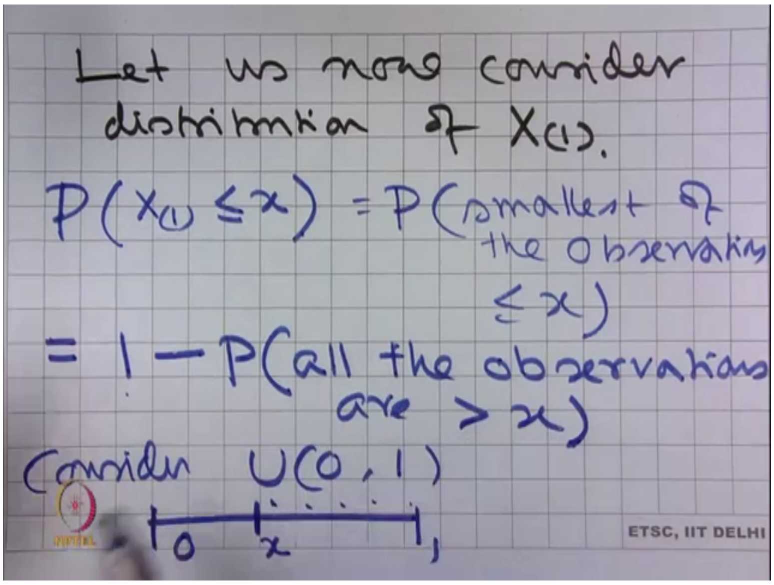
Can you from here guess, what will be the expected value of X n in general? Obviously, it is 0 to 1 n x to the power n minus 1 times x dx, and it is n upon n plus 1. Therefore, what does it say, therefore as n increases the expected value converges to 1, because n upon n plus 1. If you take n to be large, it comes close and close to 1.

(Refer Slide Time: 27:48)



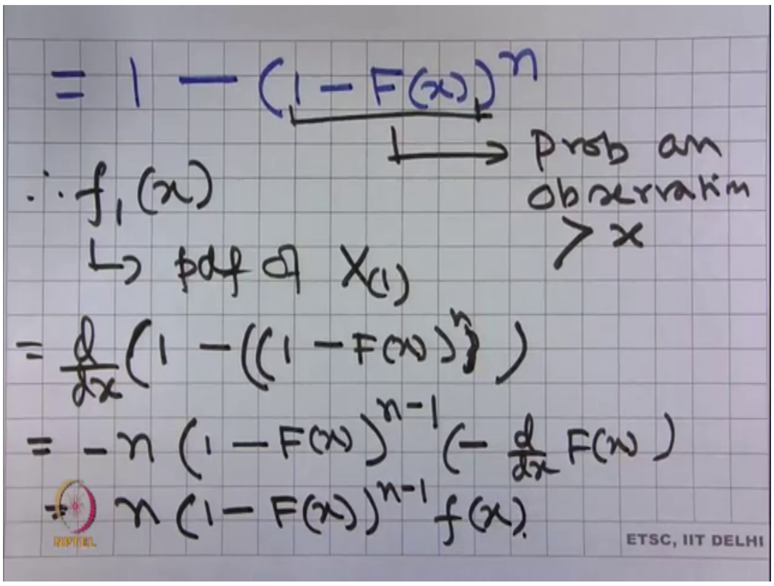
Another example x 1, x 2, x n are from exponential with lambda. Obviously, they are independent with from exponential lambda. So, what is the pdf of X n, we know that f of x is equal to lambda e to the power minus lambda x. And what is F of x capital F of x 1 minus e to the power minus lambda x. Therefore, f n of x is equal to n into 1 minus e to the power minus lambda x whole to the power n minus 1 into lambda e to the power minus lambda x. So, like that we can get the distribution of the maximum of the n samples from a exponential random variable.

(Refer Slide Time: 29:49)



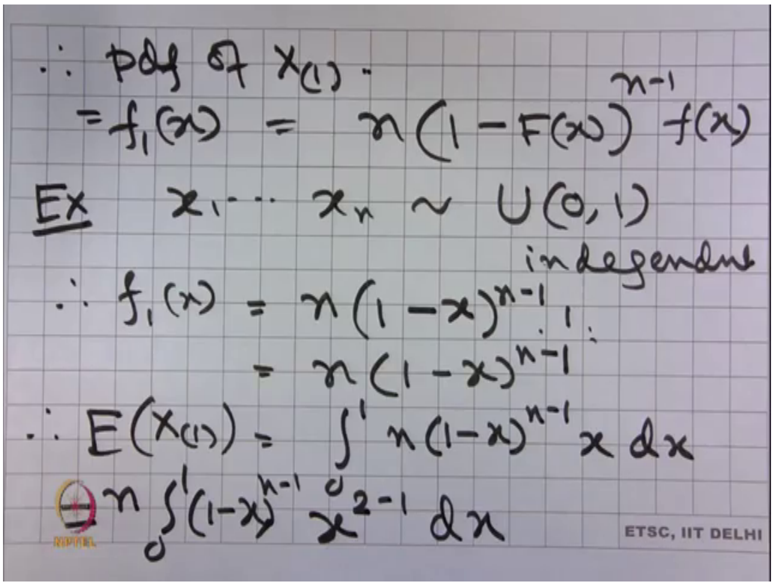
Now, let us consider distribution of X 1. So, what is the probability that the smallest of the observation is less than equal to x. Now, the smallest will be less than equal to x, it is the compliment of right. So, we consider uniform 0, 1, this is x. Therefore, if any one of them is less than x, then the minimum has to be less than equal to x. But if two of them are less than x, still the minimum is less than equal to x. And if all n are less than x, then also the minimum is less than equal to x. Therefore, the probability that minimum less than equal to x is complement of all the observations are here. Therefore, 1 minus probability all the observations are greater than x.

(Refer Slide Time: 32:13)



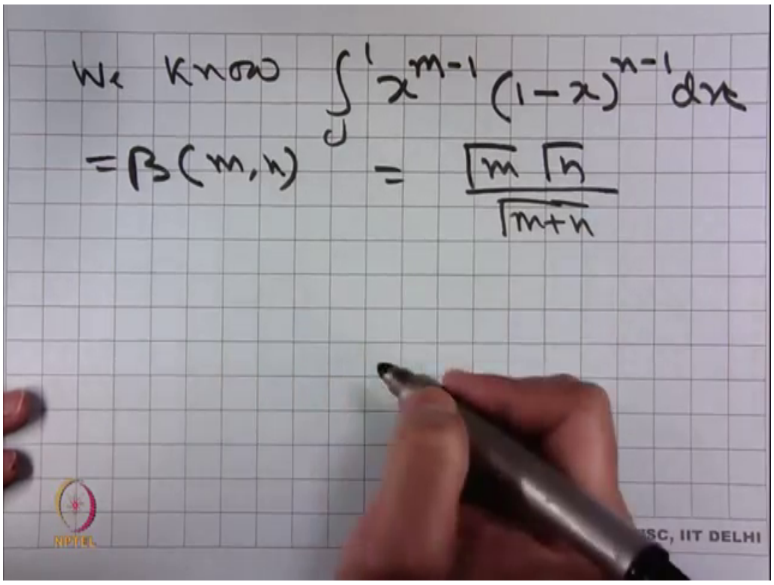
And therefore, this is 1 minus 1 minus F x whole to the power n, because this is the probability and observation is greater than x. Therefore, what is f 1 x pdf of X 1 is equal to d dx of 1 minus 1 minus F x whole to whole to the power n is equal to n into 1 minus F x whole to the power n minus 1. This minus will give you a minus into minus of d dx of F x is equal to n into 1 minus F x whole to the power n minus 1 into f x.

(Refer Slide Time: 33:57)



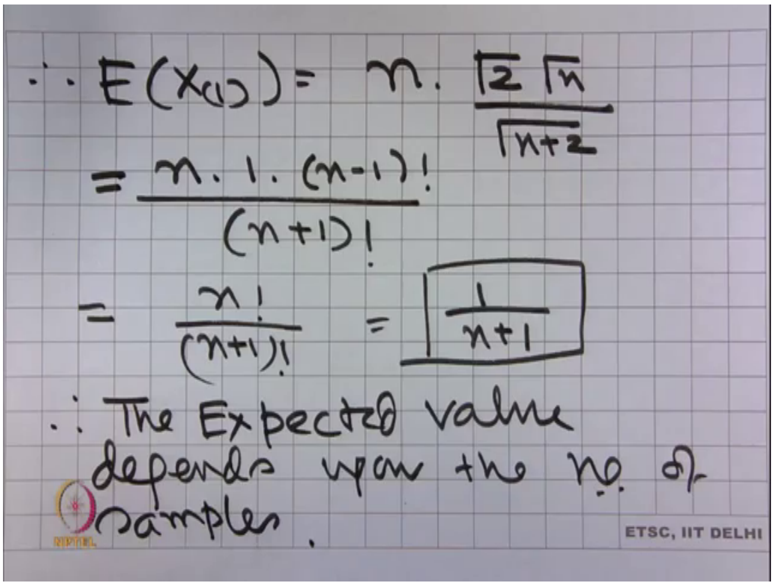
Therefore, example x 1, x 2, x n are from uniform 0, 1 independent. Therefore, f 1 x is equal to n times, what is F x, F x is x for uniform 0, 1, it is 1 minus x whole to the power n minus 1 into f x is equal to 1. Therefore, this is n into 1 minus x whole to the power n minus 1. Therefore, expected value of X 1 is equal to integration 0 to 1 n into 1 minus x whole to the power n minus 1 into x dx. How do you integrate that? This is n into 0 to 1 minus x whole to the power n minus 1 x to the power 2 minus 1 dx. And this comes under our familiar beta integral right.

(Refer Slide Time: 35:41)



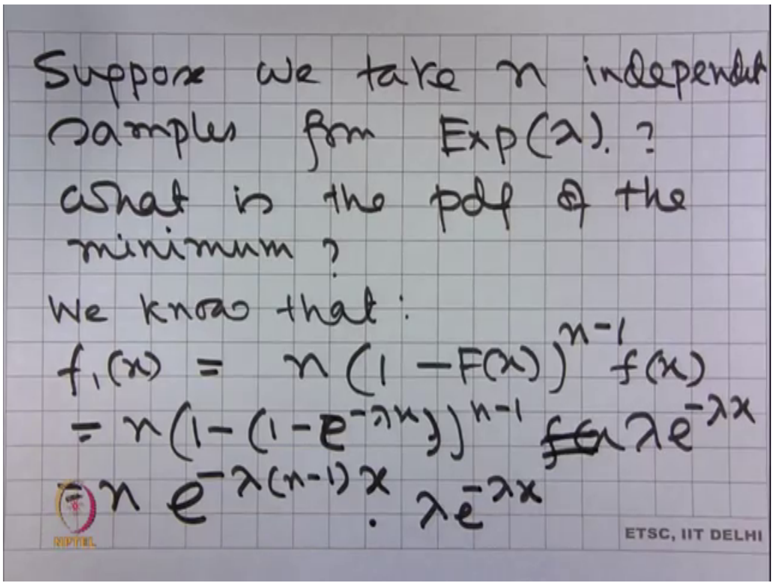
We know integration 0 to 1 x to the power m minus 1 into 1 minus x whole to the power n minus 1 dx is equal to beta m comma n is equal to gamma m gamma n upon gamma m plus n.

(Refer Slide Time: 36:11)



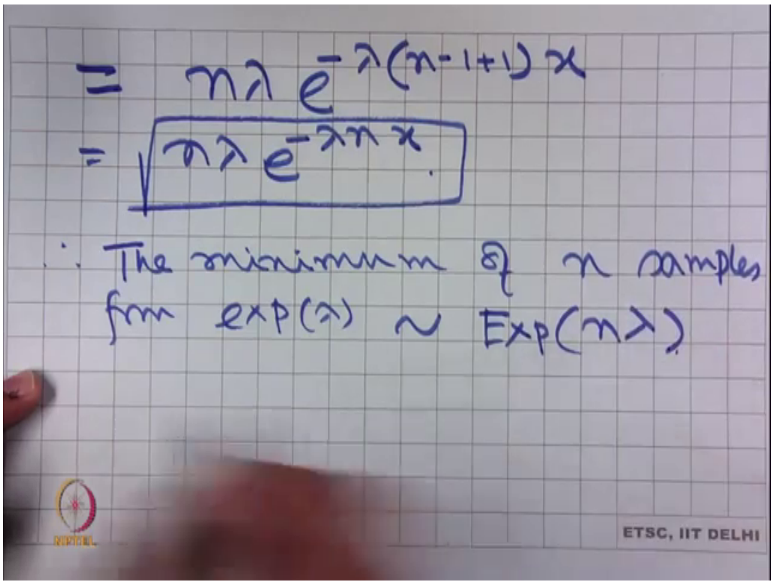
Therefore, the expected value of X 1 is equal to n times gamma 2 gamma n upon gamma n plus 2 is equal to n into 1 into factorial n minus 1 upon factorial n plus 1 is equal to n factorials upon n plus 1 factorial is equal to 1 upon n plus 1. Therefore, expected value of depends upon the number of samples. And as n increases this quantity 1 upon n plus 1 converges to 0, which is expected Because, if you are sampling from uniform 0, 1, then the minimum is expected to go to 0. And as we have observed, the maximum is expected to converge to 1.

(Refer Slide Time: 37:52)



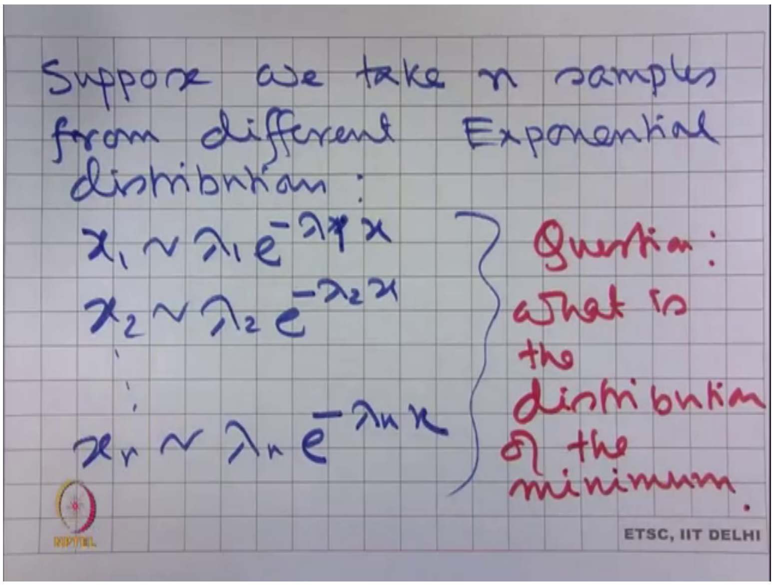
Now, suppose we take n independent samples from exponential lambda. What is the pdf of the minimum? We know that f 1 x is equal to n into 1 minus F x whole to the power n minus 1 into f x. So, in this case, it is going to be n into 1 minus 1 minus e to the power minus lambda x whole to the power n minus 1 times lambda e to the power minus lambda x is equal to n into 1, 1 cancels e to the power minus lambda into n minus 1 x into lambda e to the power minus lambda x.

(Refer Slide Time: 39:35)



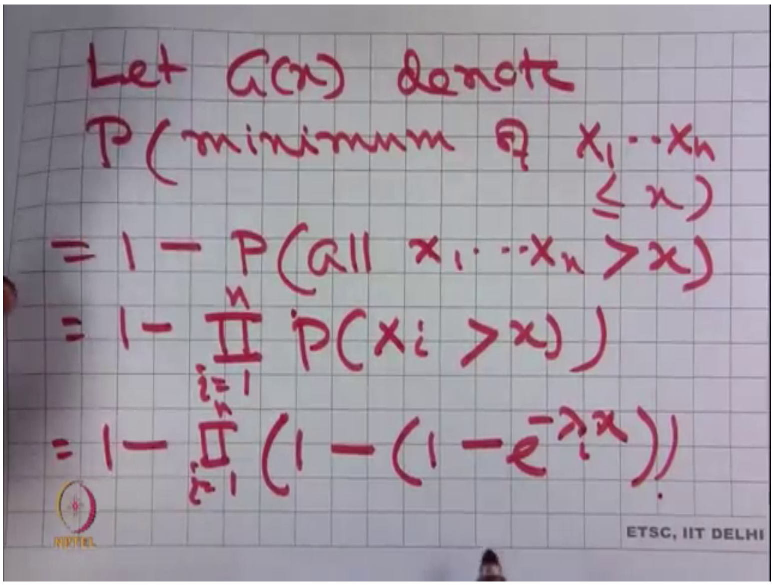
This is equal to n lambda e to the power minus lambda into n minus 1 plus 1 x is equal to n lambda e to the power minus. Therefore, the minimum of n samples from exponential lambda is distributed as exponential with n lambda.

(Refer Slide Time: 40:33)



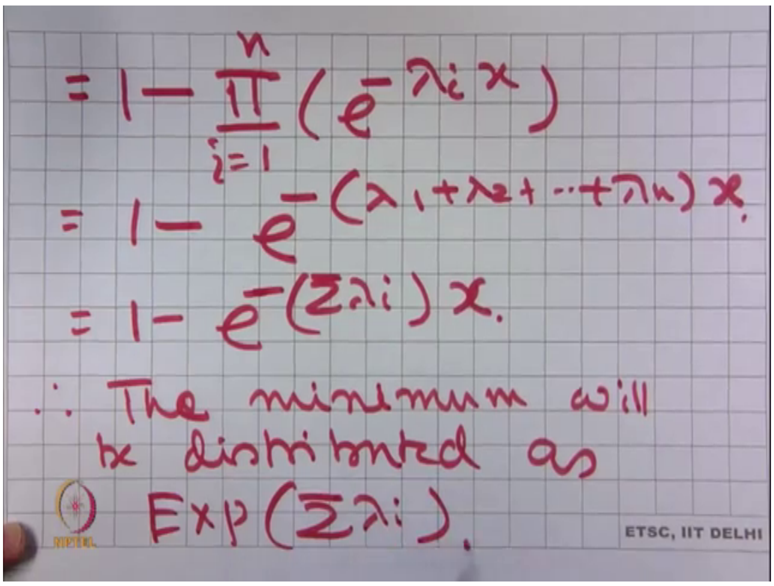
Suppose, now we take n samples from different exponential distributions. Say x 1 is from lambda 1 e to the power minus lambda 1 x. x 2 is from lambda 2 e to the power minus lambda 2 x. x n is from lambda n e to the power minus lambda n x. The question is what is the distribution of the minimum? This is not exactly coming under order statistic, but the idea that I gave will tell you.

(Refer Slide Time: 41:59)



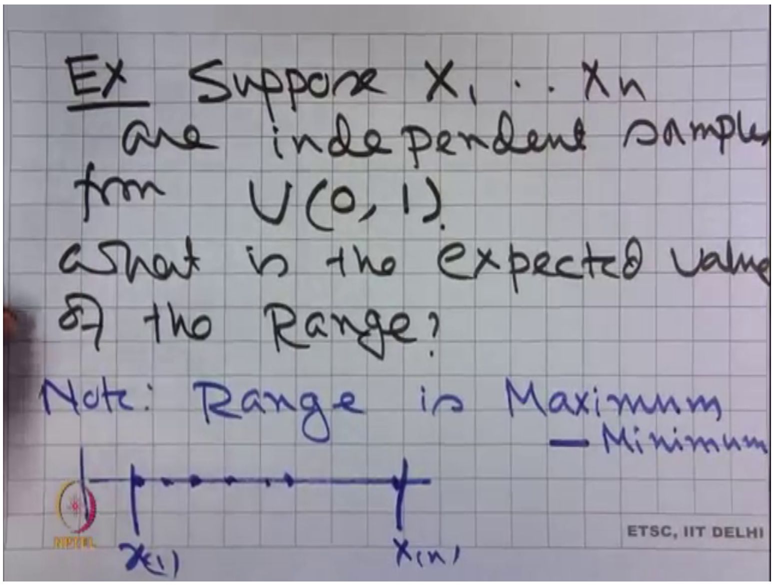
That let G x denote probability minimum of X 1, X 2, X n less than equal to x is equal to 1 minus probability all X 1, X 2, X n greater than x is equal to 1 minus product of probability X i greater than x i is equal to 1 to n is equal to 1 minus product of i is equal to 1 to n 1 minus 1 minus e to the power minus lambda i x.

(Refer Slide Time: 43:11)



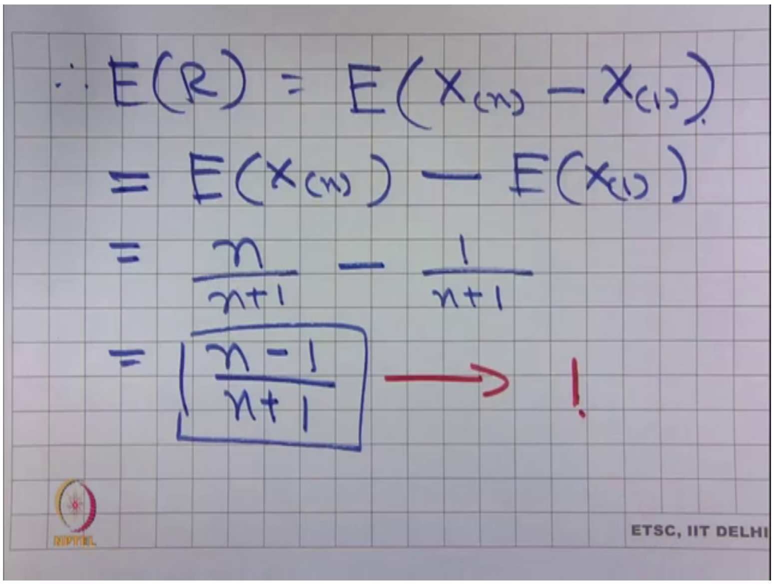
Is equal to 1 minus product i is equal to 1 to n of e to the power minus lambda i x is equal to 1 minus e to the power minus lambda 1 plus lambda 2 plus lambda n whole to the power x is equal to 1 minus e to the power minus sigma lambda i to the power in to x. Therefore, we see that the actual result is more generalized. Even if lambda 1 lambda 2 and lambda n are all different, then the cdf is going to be 1 minus e to the power minus sigma lambda i into x. In order statistic we found that if all the lambda is are similar, then this came out to be n lambda. Therefore, in this case the minimum will be distributed as exponential with sigma lambda i.

(Refer Slide Time: 44:54)



Now, let me give an interesting example suppose X 1, X 2, X n are independent samples from uniform 0, 1, what is the expected value of the range. What is the range? It is the difference of the maximum minus minimum. We know that if these are the samples, then the range is x n minus x 1.

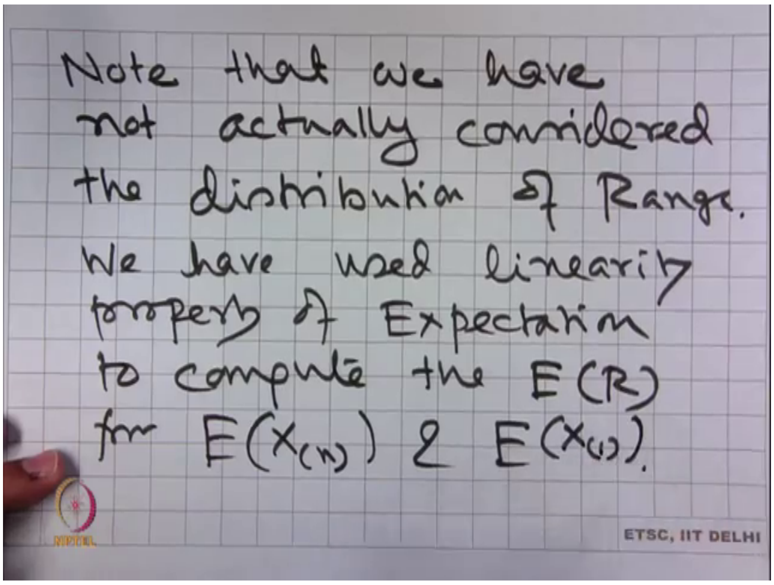
(Refer Slide Time: 46:21)



Suppose, you want to find out the expected value of the range, by linearity of expectation, we can say it is the expected value of X n minus expected value of X 1 is equal to n upon n plus 1 minus 1 upon n plus 1 is equal to n minus 1 upon n plus 1. Therefore, if we take n samples, the expected value of ranges n minus 1 upon n plus 1 which is obviously less than 1 because we are taking samples from uniform 0, 1.

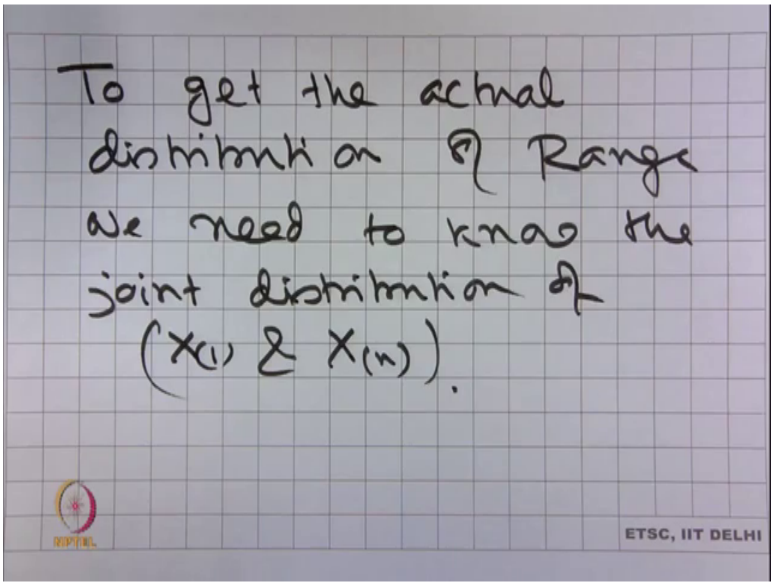
But what is going to happen? As n increases these value converges to 1 right. And this is expected because we are talking about samples from 0 to 1. Therefore, as more and more samples will be taken we expect that the entire range will be covered the entire span from 0 to 1 will be covered. Therefore, the expected value of the range is going to be very very close to 1 as n increases.

(Refer Slide Time: 48:03)



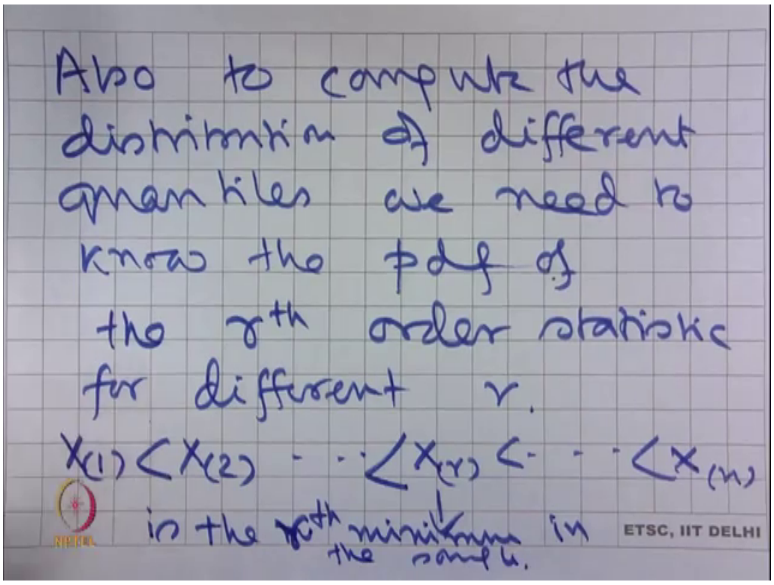
Note that we have not actually considered the distribution of range. In fact, we have used linearity property of expectation to compute the expected value of range from expected value of the maximum and expected value of the minimum.

(Refer Slide Time: 49:12)



To get the actual distribution of range we need to know the joint distribution of X 1 and X n. From there we can find out the expectation of the range. In fact, from there we find out the distribution of the range and from there we shall calculate the expected value of range.

(Refer Slide Time: 50:10)



Also to compute the distribution of different quantiles, we need to know the pdf of the rth order statistic for different r. What is rth order statistic? We know that X 1 less than X 2. So, rth order statistic is the minimum the rth minimum in the sample.

In the next class, I shall look at the distribution of the rth order statistic for different r from 1 to n. And I shall also look at the joint distribution of rth and sth order statistic in the next class.

Thank you.