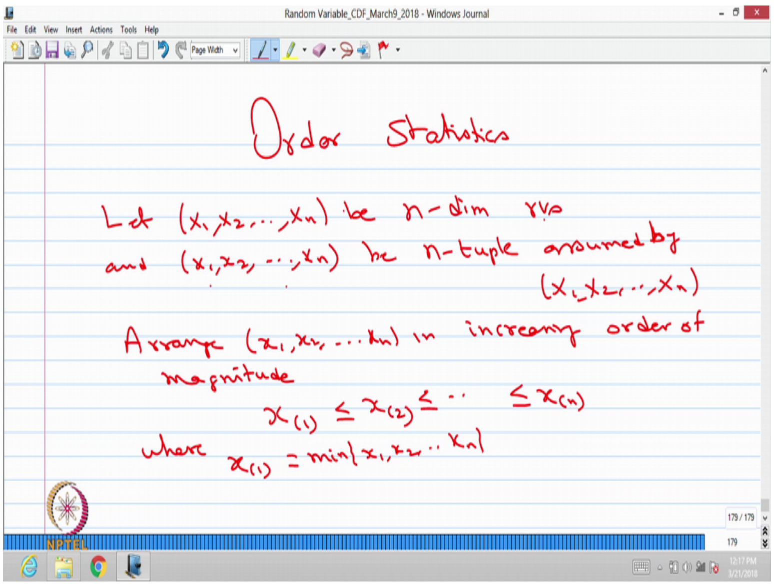
Introduction to Probability Theory and Stochastic Processes
Prof. S. Dharmaraja
Department of Mathematics
Indian Institute of Technology, Delhi

Lecture – 32

In the last class, we have discuss a distributions of functions of several variables. In that we have started with the discrete type random variables. If you have n dimensional discrete type random variables and one can find the distributions of a the new set of n dimensional random variable of the discrete types.

Then later, we have discussed the; when you have a continuous type of random variable of n dimensional and you have another set of a new continuous type n dimensional random variable using nice theorem, one can get the joint probability density function of a new set of n dimensional random variable. By applying the theorem, one can get the joint distribution, then you can get the margin.

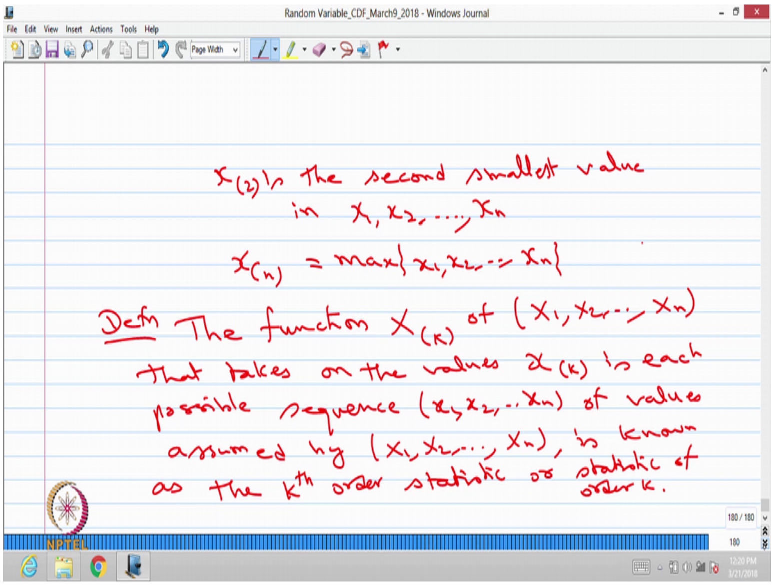
(Refer Slide Time: 01:13)



Now, in this lecture, we are going to discuss one particular type of a functions of several random variables that is called Order Statistics. Why the name is called the order statistics? Why the order statistics as to be in this lecture as the functions of several random variables? You will understand. So, for that I am going to give first, what is the meaning of a order statistics from the scratch.

Let X 1, X 2 X n be a n dimensional random variables and small x 1, small x 2, small x n be a n triple be a n triple assumed by the random variables X 1, X 2 X n; that means, the possible variables of a the n dimensional random variables are small x 1, x 2, x n's. Arrange this possible values in increasing order of magnitude. So, that I can make out that is, make it as a x bracket 1, x bracket 2 and so on till x bracket n where the x bracket 1 is nothing, but a minimum of a this xi values.

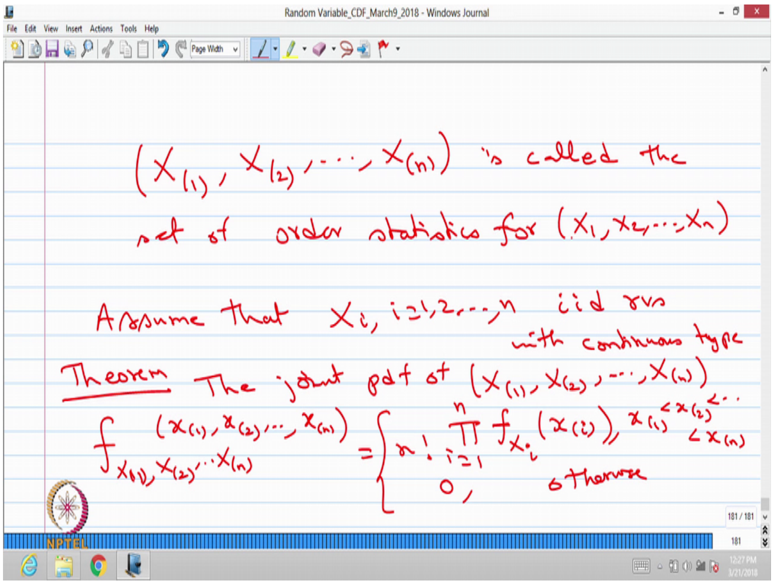
(Refer Slide Time: 03:44)



Similarly x bracket 2 is the second smallest value in x 1, x 2 so on x n. Similarly the x n, x bracket n that is nothing, but a maximum of x 1, x 2 and so, on till x that is a different between x i's and x bracket i's fine.

Now I am going to define, the order statistics. That is the function X bracket K of X 1, V 2 so on X n, that takes on the values x small k a small x within bracket k in each possible sequence x 1 x 2 and so on x n of values assumed by assumed by the n dimensional random variables X 1, X 2, X n. That function, the function X bracket K in that capital X bracket K that is known as the K'th order statistics or statistics of K'th order statistics of order K.

(Refer Slide Time: 06:54)



Now, the capital X bracket 1, capital X bracket 2 and so, on till capital X bracket n that is called the set of order statistics for the random vector X 1 X 2 so on X n. So, from the given n dimension random vector, we are getting another n dimension random vector X bracket 1 and X bracket 2 and so on. In which each X bracket case is the function of X i's in which the X bracket 1 is a minimum of all those random variables, X bracket n the maximum of n such random variables. This set of n random variable is called the order statistics.

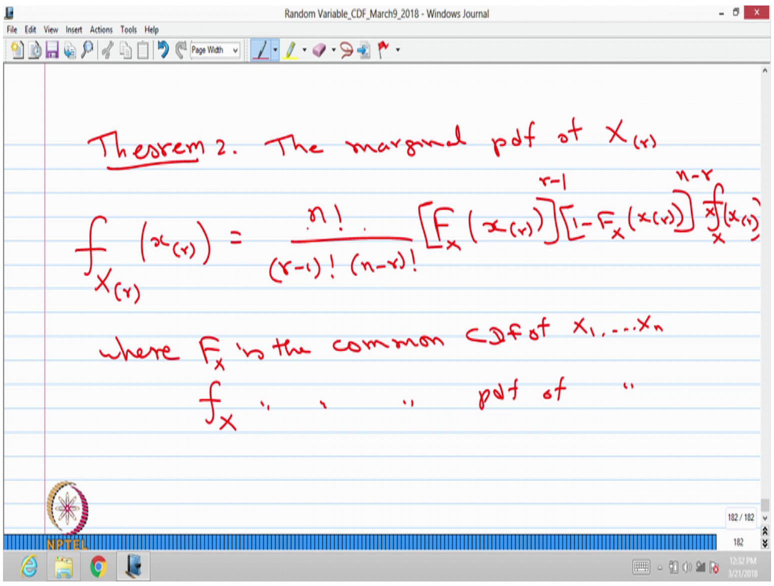
We will go for some important results on order statistics. For that I am going to make the assumption. Assume that assume that the random variable X i's, i is equal to 1 to n or i i d random variables. We have already given the definition of i i d; that means, each random variable having the same distribution and all the random variables are mutual independent. All the random variables are the mutual independent and having identical distribution. Therefore, they are called i i d random variables with continuous type; that means, all the random variables are of the continuous type as well as they are i i d random variables, going to give the result as the following theorem.

One can find the joint probability density function of the order statistics as the joint probability density function with a possible values, in terms of the probability density function of a X i's that is n factorial. Product of i is equal to 1 to n, the probability density function of the i'th random variable by substituting a x i by x bracket i. The product of this probability density function multiplied by the n factorial that is going to be the joint probability density function within the interval when x bracket 1 is lesser then x bracket 2 and so on till x bracket n, otherwise it is going to be 0.

So, this results also comes from the theorem. One can find the joint probability density function of order statistics, in terms of the joint probability density function of X 1, X 2 X n, but since this random variables are i i d random variables. Therefore, in the right hand side, instead of joint probability density function of X i's; we have a probability density function of a x i itself, the n factorial product of a i is equal to 1 to n and probability density function of x i by substituting the value x i by x of bracket i between the interval x of 1 less then x of 2 and so on till less then x of n. It is a non zero joint probability density function, otherwise it is 0.

One can verify, the multidimensional integration with respect to d of x bracket 1 d of x bracket 2 d of x bracket n that is going to be n. So, here we are finding the joint probability density function of a order statistics.

(Refer Slide Time: 12:42)

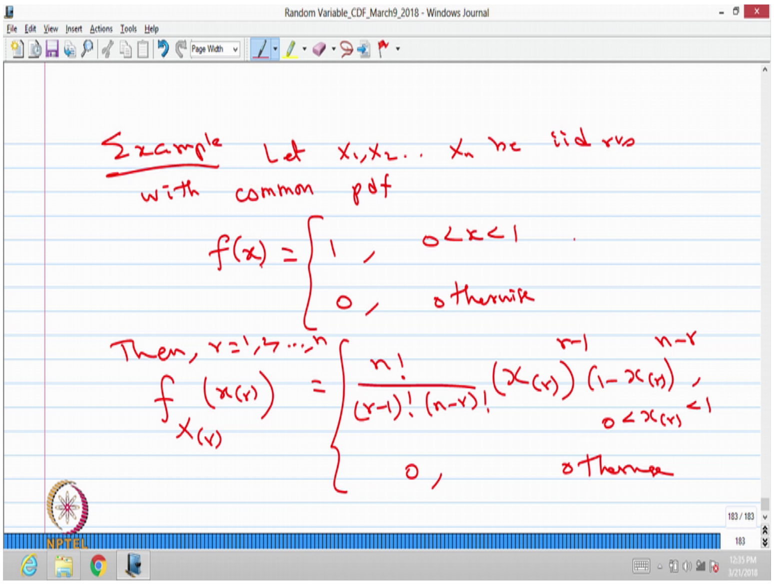


The next result, one can get the marginal distribution of a order statistics that has second theorem. The theorem 2, there marginal probability density function of any r'th order statistics is given by f suffix x bracket r with the value of x of r that is going to be, since you know the joint probability density function of order statistics; one can easily, not one can easily, one can get the marginal probability density function of any one r'th order statistics from the n dimensional joint probability density function of order statistics. So, this is going to be n factorial divided by r minus 1 factorial n minus r factorial multiplied by the CDF of since all the random variables are i i d random variables, you do not need to mention or you can just mention x or any x i of evaluated at x bracket r.

This CDF raised to the power r minus 1 multiplied by the 1 minus CDF of any one random variable. All are identical random variable. CDF evaluated at x bracket r raised to the power n minus r multiplied by the probability density function of any one random variable evaluated at x of r [vocalize-. This is going to be the probability density function of r'th order statistics by using the CDF of any one random variable and the probability density function of any one random variable. So, instead of writing x i, I am writing x where F is the common; since they are identical, it is a common CDF of the random variables X 1, X 2 to X n.

Similarly, the small f x is the common probability density function of the random variable X 1, X 2, X n. So, one can get the probability density function of order statistics with the help of the probability density function and the CDF of the random variables common CDF and the common PDF of the random variable X 1, X 2 to X n as a example.

(Refer Slide Time: 16:24)



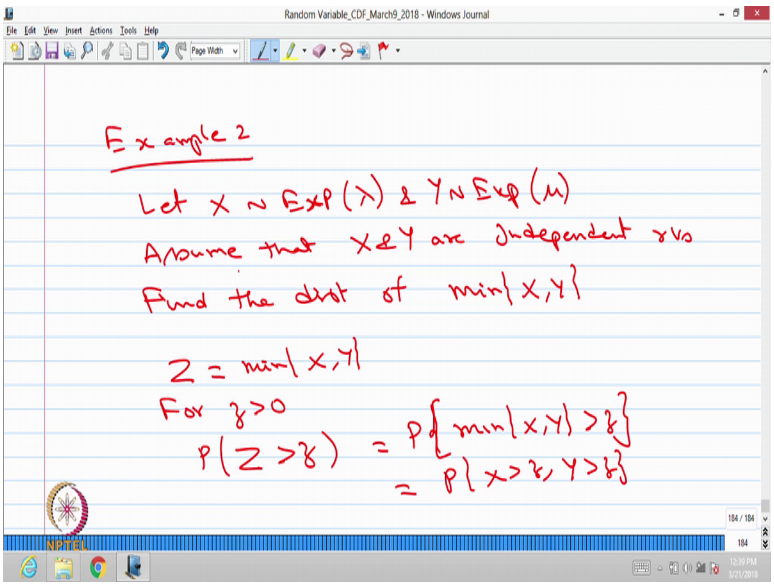
We are going to make, let X 1, X 2, X n be i i d random variables with common with common probability density function; that means, I am going for all the random variables of the continuous type. So, the common probability density function that is 1, when x is lies between 0 to 1; otherwise it is 0. All are i i d random variable as well as all are continuous type random variable with the common probability density function f of x is this.

By using the previous result, one can get the probability density function of any r'th order statistics. That is the probability density function of any r'th order statistics; that is going to be n factorial divided by r minus 1 factorial multiplied by n minus r factorial and then you have to substitute. See the previous result, you have to substitute the CDF as well as the probability density function.

And if you see this example; you can make out, the distribution of the X i's is a uniform distribution between the interval 0 to 1. If the random variable is uniform distributor between the interval 0 to 1, you can get the CDF easily. So, you substitute the CDF; that is the x suffix within bracket r that power r minus 1 and the 1 minus CDF that is a 1 minus x suffix r that is power n minus r and multiplied by the probability density function that is 1.

So, this is valid whenever the r'th order statistics is going to be 0 to 1 and this is true for r is equal to 1 2 and so on till n, 0 otherwise. So, this result that the probability density function is non-zero, when x bracket r is lies between 0 to 1, otherwise it is 0 and this way you can find out the probability density function of r is equal to 1 to n.

(Refer Slide Time: 19:45)



We will go for second example. The second example is let X follows exponential distribution with the parameter lambda and Y follows exponential distribution with the parameter mu. Assume that both the random variables are independent. We have two exponential distributed random variables with the parameters lambda and my respectively. Our interest is to find the distribution of, find the distribution of minimum of X comma Y.

Since it is a only 2 random variables, you will have a minimum of 2 random variables as well as we will have a maximum of 2 random variables. Both together that is a set of order statistics. In that, I am interested to find out the distribution of a minimum of 2 random variables. I can use the previous theorem, I can get the joint distribution; then I can get the marginal distribution by applying the theorem, but I am not going to apply the theorem.

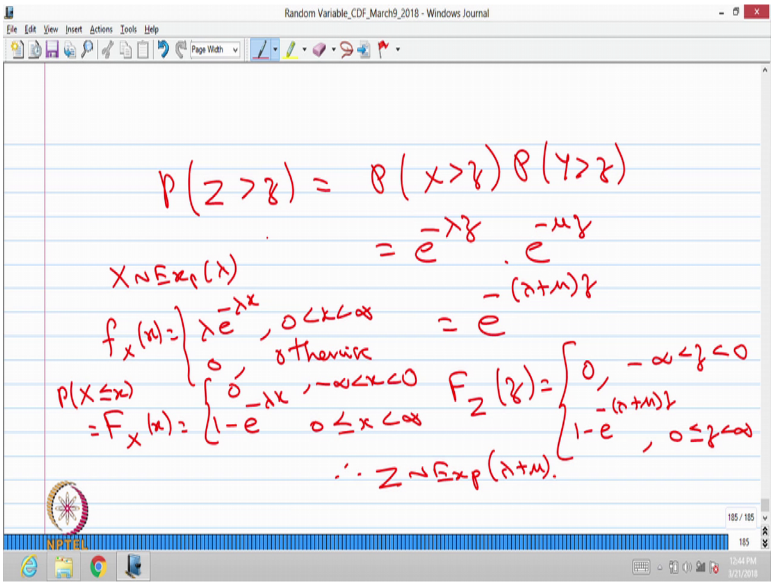
Instead of that I can easily able to find out the minimum of 2 random variables without using the previous theorem; that is let me make new random variable Z is minimum of X comma Y. Since X is a continuous type, Y is the continuous type minimum of X comma Y that is also going to be a continuous type random variable. Therefore, I have to find the CDF of the random variable Z or probability density function of Z.

By using the previous theorem, I can get directly the probability density function of Z, but I am not going to do that. I am going to find out the distribution of Z, in the form of CDF. Since it is a minimum, I will go for compliment CDF that is for Z is greater than 0. The probability of capital Z is greater than small z. I am going for finding out the compliment CDF of the random variable Z, that is same as the probability of Z is nothing, but minimum of X comma Y; that is greater than small Z.

Since minimum of X comma Y is going to be greater than z; that means, each random variable is also going to be greater than 0. So, that is same P of X is greater than z as well as Y is greater than z. If minimum of X comma Y is greater than z, that means; both X and Y greater than z.

If I know the joint distribution, I can find out the probability of this. But I made the assumption; both the random variables are independent. Therefore, the joint distribution is same as the product of individual distributions. Therefore, the probability of Z is greater than z is same as probability of X is greater than z as multiplied by probability of Y is greater than z.

(Refer Slide Time: 23:43)



I have already made the assumption X follows exponential distribution with the parameter lambda. Therefore I should know, what is the probability density function? Similarly I should know, what is the CDF of exponential distribution? That is 1 minus e power minus lambda x. It is 0 between minus infinity to 0 from 0 onwards, it is going to be 1 minus e power minus lambda x.

So, I am going to use this result in the probability of X is greater than z. The CDF is nothing, but probability of X is less than or equal to small x. So, what I want is probability of X is greater than z. So, that is same as e power minus lambda z. Similarly Y is also exponential distributor with the parameter mu. The similar derivation makes; that is e power minus mu times z. So, the conclusion is, it is e power minus lambda plus mu times z, when z is going to be greater than 0.

So, I can write down things correctly that is CDF of z that is going to be 0 when z is between minus infinity to 0 and 1 minus e power minus lambda plus mu times z, when z is starting from 0 to infinity. So, this is the distribution of z. Distribution means, here it is a CDF.

In this page itself, you can compare; when X is exponential distribution the CDF is 0 from minus infinity to 0, then 0 to infinity it is 1 minus e power minus lambda x. It is in the same form. Therefore, one set two different variable having the same CDF, then you can conclude both the random variables are of the same distribution. Therefore I can conclude, Z is also same distribution of exponential distribution. The parameter is, here it is minus lambda, here it is minus lambda plus mu and this is exponential distribution with the parameter lambda. Therefore, Z is going to be exponential distribution with the parameter lambda plus mu.

When two different random variables having the distribution of the similar form, then you can conclude both the random variables are having the similar distributions. So, here Z is going to be exponential distribution with the parameter lambda plus mu. So, the observation I,s whenever you have a two independent exponential distributor random variable random variables, then the minimum of a independent exponential distributed random variables is again exponential distribution.

This concept can be extended for n dimensional random variable, n dimensional random variables; that means, if you have a mutually independent exponentially distributed n dimensional random variables, then the minimum of a those mutually independent exponentially distributed random variables is again exponential distribution with the parameter is some of the parameters of individual distribution. This is a easy way of finding the minimum of two independent exponential distribution.