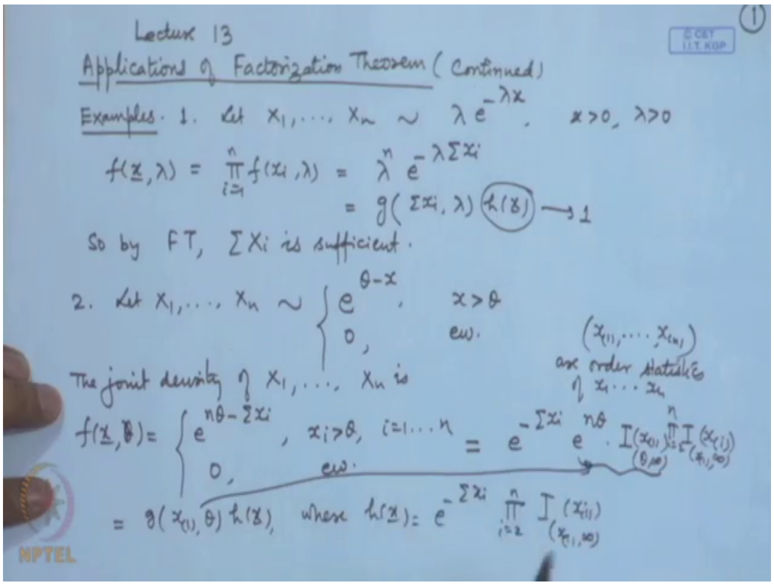
Statistical Inference
Prof. Somesh Kumar
Department of Mathematics
Indian Institute of Technology, Kharagpur

Lecture – 25
Sufficiency and Information – I

In the previous class I have explained the concept of Sufficiency; this concept is the concept which is called the principle of data reduction. So, we have a random sample X 1 X 2 X n, but if we have a sufficient statistic t then that is sufficient that gives the complete information about the parameter which is contained in the sample. So, we did not retain it. We have given one theorem which is called factorization theorem and this is useful for deriving sufficient statistics in various probability models.

Yesterday I have discussed the normal probability model and I have shown you that how if we change the parameter is space; that means, whether we have mu known or sigma square known or both are unknown, in each of the cases the sufficient statistics changes. So, sufficiency is the property of the probability model under consideration; let me explain it through few more examples.

(Refer Slide Time: 01:25)



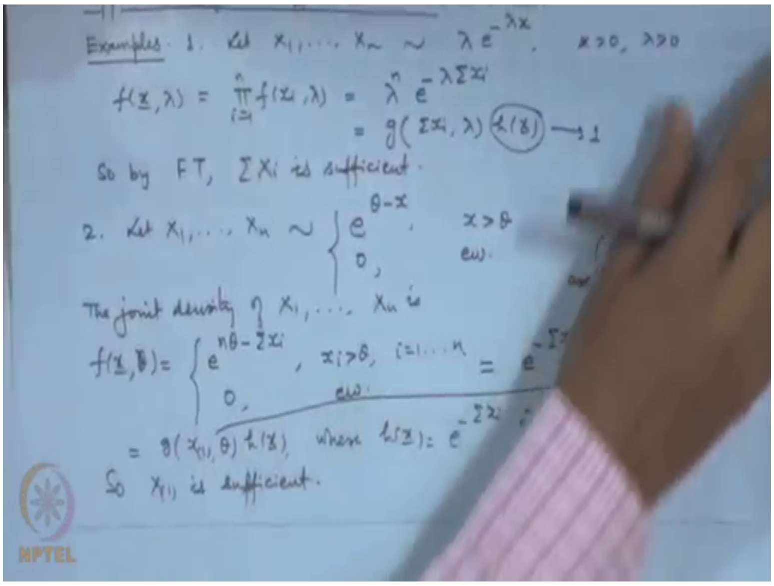
And we will use the concept of this factorization theorem here, let me start with exponential distribution let X 1 X 2 Xn follow exponential distribution say with parameter lambda.

So, in the factorization theorem we need to write down the joint density of X 1 X 2 X n that is equal to lambda to the power n e to the power minus lambda sigma x i. Now this whole thing we can write as a function of sigma x i and lambda and hx; this h x I am taking to be 1 itself the constant. So, you can see by factorization theorem, by factorization theorem sigma x i is sufficient. Let us consider another exponential model in which in place of a scale parameter we will have a location parameter.

So, let us consider say X 1 X 2 Xn following exponential say theta minus x, where x is greater than theta 0 elsewhere. Now in this case the joint density of X 1 X 2 X n is f of x theta that is equal to e to the power n theta minus sigma x i. However, this description of x i greater than theta also plays a role here, now if we want to write it as a product here we will make use of the indicator function. So, we can write it like this e to the power minus sigma xi e to the power n theta indicator function of the set x 1 from theta to infinity and indicator function of other x i's from 2 to n from x 1 to infinity.

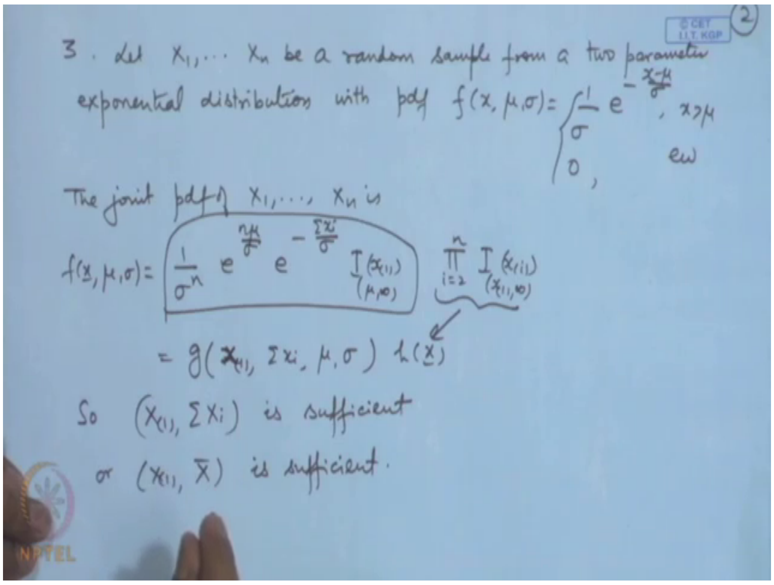
So, what we can consider we can write it as g of x 1 theta into h of x, where h of x I am writing as e to the power minus sigma x i into product i is equal to 2 to n I of x i x 1 to infinity. So, here this x 1 x 2 x n they are denoting the order statistics of x 1 x 2 x n. So, g is this function, this is a function of x 1 and theta. So, we conclude that x 1 is so, x 1 is sufficient.

(Refer Slide Time: 04:49)



Now, note here when we had lambda as the parameter and here we had a scale model the sufficient statistic was sigma x i, although here again we are dealing with exponential distribution, but the nature of the parameter has changed. And therefore, the sufficient statistic is now the minimum of the observations.

(Refer Slide Time: 05:23)



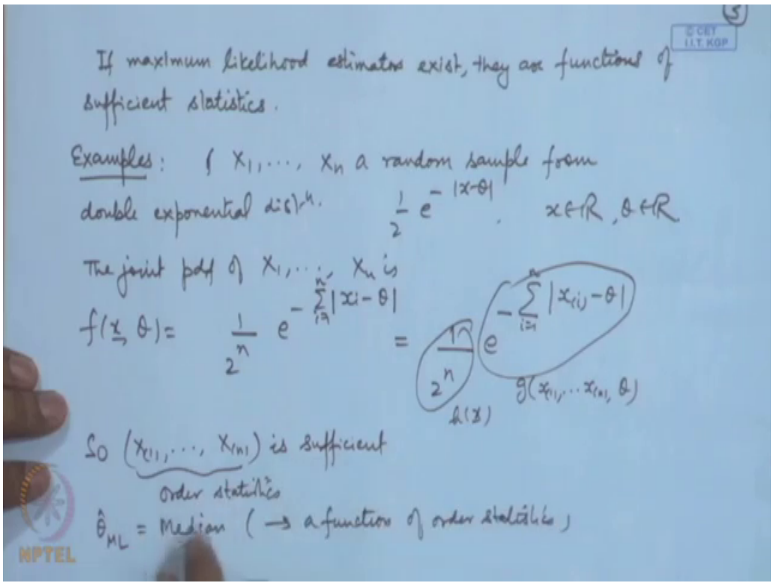
Now in a similar way let us take up the two parameters exponential distribution let us take X 1 X 2 X n be a random sample from a two parameter exponential distribution; say with density function f x mu sigma is equal to 1 by sigma e to the power minus x minus mu by sigma for x greater than mu and it is equal to 0 otherwise. So, once again the joint probability density function of X 1 X 2 X n this is now 1 by sigma to the power n e to the power n mu by sigma e to the power minus sigma xi by sigma. And, once again the condition that each of x i is greater than mu I can expressed in terms of the indicator function like x 1 is from mu to infinity and x i is other x i is they are from x 1 to infinity i is equal to 2 to n.

So, this portion I can write as g of x 1 sigma xi and mu sigma and this part is hx. So, here we conclude that X 1 and sigma Xi is sufficient or we can also say X 1 and X bar because this is a 1 to 1 function of this is sufficient. I also want to mention here we have earlier considered the maximum likelihood estimators; now let us remember our maximum likelihood estimators for each of these problems for example, in this case the maximum likelihood estimator for lambda was 1 by x bar which is the function of sigma xi.

In this particular case the maximum likelihood estimator was x 1 that is a minimum of observations and it is sufficient here. Similarly here you see the maximum likelihood estimator for mu and sigma where X 1 and X bar minus x 1 respectively which is again a 1 to 1 function of X 1X bar that is a sufficient statistics. So, we can observe that the maximum likelihood estimator if it is exists is actually a function of the sufficient statistics.

The reason is obvious because in the factorization theorem we are writing down the density as a function of the parameter and the sufficient statistics into a function which is free from the parameter. Now in the method of maximum likelihood estimator we are maximizing the density function or the mass function with respect to the parameter. Now, the part of the density which contains the parameter contains the variable only through the sufficient statistics. Therefore, the maximization problem will give a solution in terms of the sufficient statistics alone.

(Refer Slide Time: 09:19)



So, we have a general comment here that if maximum likelihood estimators exist they are functions of sufficient statistics. Let us take some more examples here say for example; X 1 X 2 X n a random sample from double exponential distribution half e to the power minus x minus theta where x is any real number theta is any real number. In this case if we considered the sufficiency. So, the joint distribution of X 1 X 2 X n that is equal to 1 by 2 to the power n e to the power minus sigma modulus xi minus theta i is equal to 1 to n. Now here you observe I cannot reduce it further as a function of parameter and another variable here.

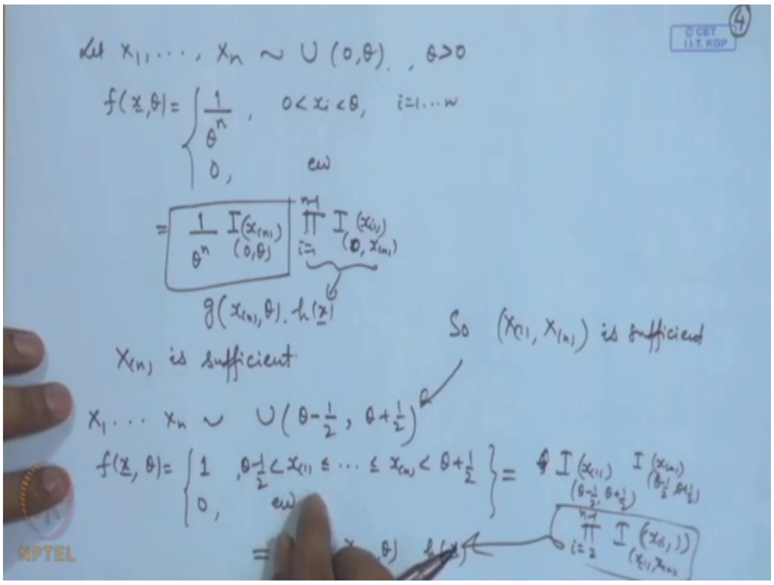
Because each of the x i's are appearing in the modulus sign and therefore, I cannot separate it out. At the most I can consider the reduction as 1 by 2 to the power n e to the power minus sigma modulus of xi order statistics minus theta. So, this function is now a function of the order statistics and theta and this we can call h x. So, we conclude that the order statistics X 1 X 2 Xn is sufficient this order statistic. Now, remember here for this problem what was the maximum likelihood estimator? The maximum likelihood estimator was median of the observations.

And median is a function of this is a function of order statistics because, if we have an odd number of observations say x 2 m plus 1 then x m plus 1 that is the middle of the observation was the median. And, if we have an even number of observations that is x 2 m then any number between xm and x m plus 1 and we can actually consider say the middle of the 2 that is xm plus xm plus 1 by 2 as the maximum likelihood estimator. So, this is a function of order statistics in this case also.

So, this statement is true in general. Another thing which I just now pointed out that many times when we are writing down the density function say in this case we are, we have to incorporate the region of the variable which is dependent upon the parameter as a part of the joint density function. Because, if you do not included it then we cannot decide a sufficient statistic for example, if he had written only this part then, there is no sufficient statistics here because e to the power minus sigma x i can be separately written e to the power n theta can we separately written.

However, this is not a complete description of the density unless we include the region xi greater than theta for all i and this is the way of including this. A similar phenomena is observed in the uniform distributions also like in the uniform distribution the range is dependent upon the range of the variable is dependent upon the parameter.

(Refer Slide Time: 13:43)



So, let us consider say X 1 X 2 X n say a random sample from uniform 0 theta distribution. Now in this case the joint density is equal to 1 by theta to the power n. And once again each of the xi is between 1 to n and it is equal to 0 elsewhere. So, this part we will express as then 1 by theta to the power n I of say x n from 0 to theta and the other xi's are from 0 to x n for i is equal to 1 to n minus 1. So, you can see this portion we can express as g of x n theta and this part we can write as h x.

So, X n is sufficient; however, if we consider say uniform distribution which is on a 2 sided interval here we have taken one side as 0; suppose we consider say from theta minus say 1 by 2 2 theta plus 1 by 2. In this case the joint distribution is simply 1 because, theta plus half minus theta minus half is 1. However, each of the x i's they are between theta plus half and theta minus half. So, this part then we can incorporate as indicator function of x 1 from theta minus half to theta plus half and the indicator function of x n from theta minus half to theta plus half.

And the remaining order statistics line between x 1 and x n i is equal 2 to n minus 1. So, this you can see it is a function of x 1 x n theta into h x this part is h x, this part is a h x and this part is a function of x 1 x n and theta. So, here X 1 X n is sufficient although the parameter remains 1 dimensional here, but the order statistics contains 2 terms. If you remember the maximum likelihood estimator; the maximum likelihood estimator for this problem was any value between x n minus half to x 1 plus half so, which is the function of X 1 X n. So, the statement that the maximum likelihood estimators if they exist they are functions of the sufficient statistics is satisfied here also.

So, these are the topics that I will be covering in the next lecture.