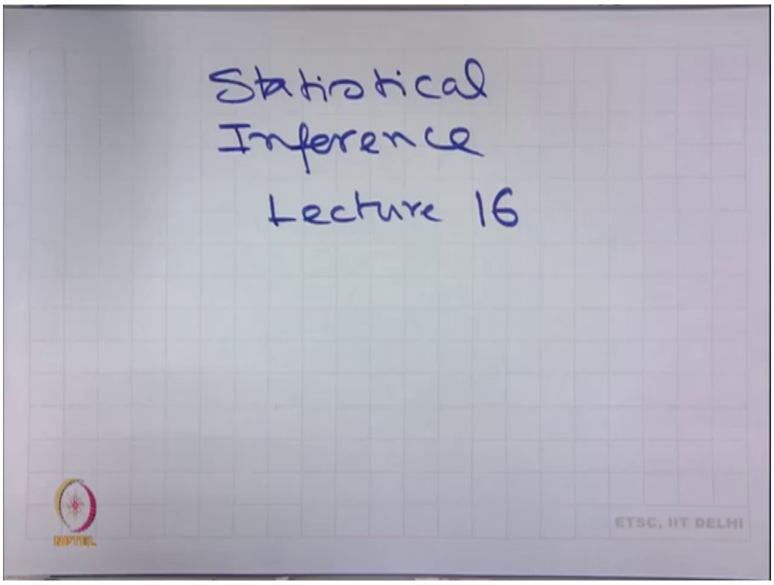
Statistical Inference
Prof. Niladri Chatterjee
Department of Mathematics
Indian Institute of Technology, Delhi

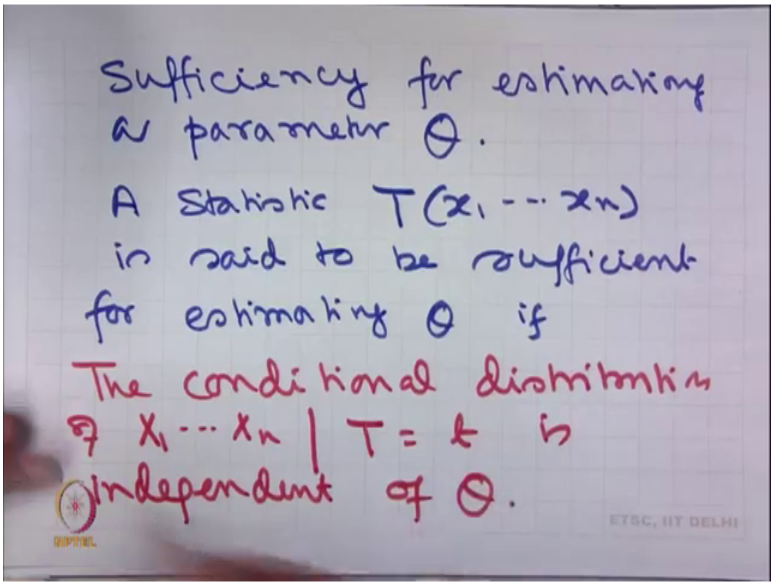
Lecture - 16
Statistical Inference

Welcome students to the MOOC's lecture on Statistical Inference. This is lecture number 16.

(Refer Slide Time: 00:26)

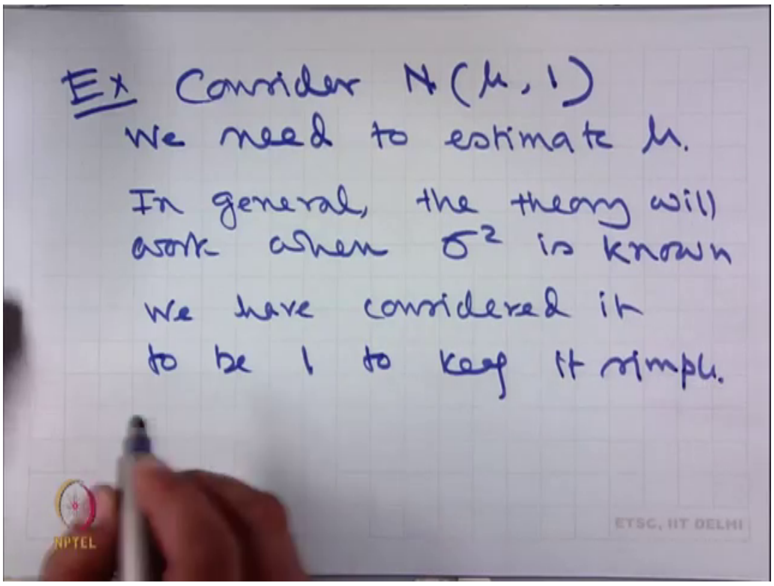


(Refer Slide Time: 00:35)



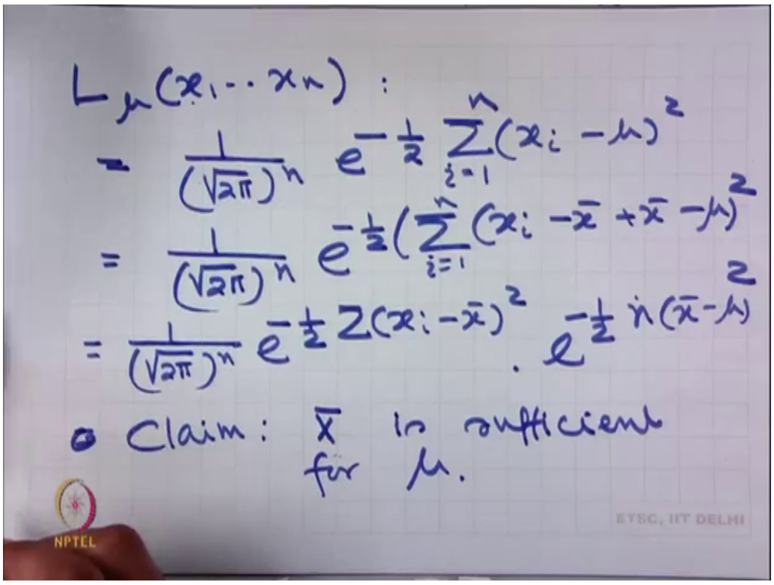
In the last lecture, we have discussed an important property namely sufficiency for estimating a parameter theta and the condition was that a statistic T of x 1 x 2 x n, where x 1 x 2 x n is the sample values is said to be sufficient for estimating theta. If the conditional distribution of X 1 X 2 X n given, T is equal to t is independent of theta. So, let me first give you some more examples of sufficient statistic.

(Refer Slide Time: 02:09)



For example, consider normal mu 1. We need to estimate mu. Note that the variance is already known in this case it is 1. In general the theory will work when sigma square is known. We have considered it 1 to keep it simple, ok.

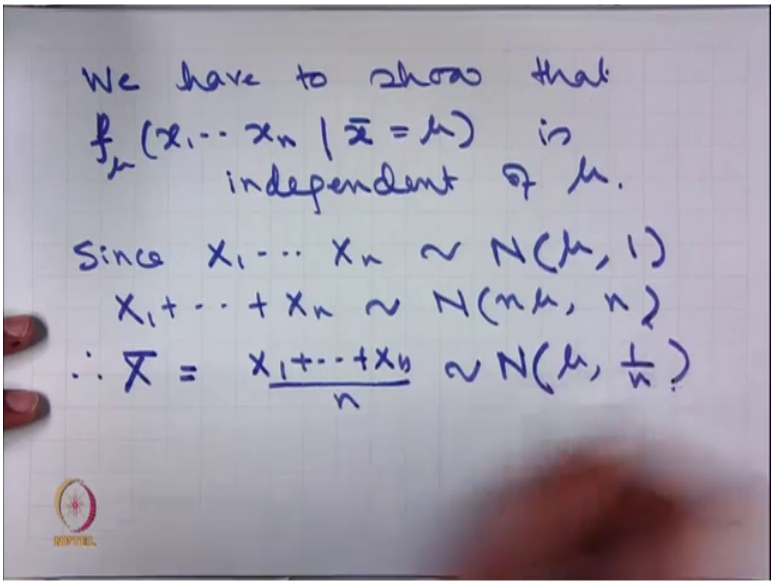
(Refer Slide Time: 03:25)



So, L mu x 1 x 2 x n that is joint density of x 1 x 2 x n under the parameter mu is equal to 1 over root over 2 pi whole to the power n e to the power minus half into sigma x i minus mu whole square i is equal to 1 to n. This is obvious because, the likelihood function here is the product of the individual density and we know that if x is normal mu 1, then the density of x is 1 over root over 2 pi e to the power minus half x minus mu whole square. Now, I am adding it because x 1 x 2 x n there are n observations. So, L becomes the product of their individual densities. Therefore, in the exponent there being added is equal to 1 over root over 2 pi to the power n e to the power minus half sigma i is equal to 1 to n x i minus x bar plus x bar minus mu whole square.

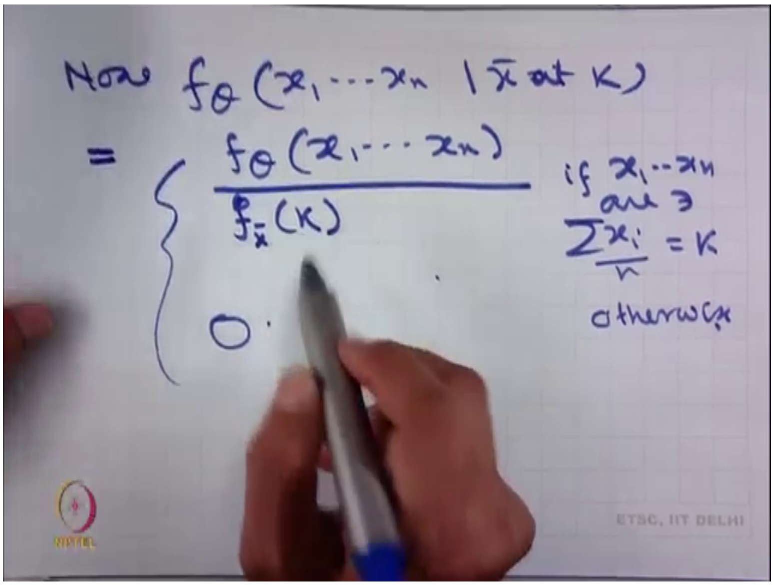
So, what we have done? I have subtracted and added x bar so, this allows us to write it as e to the power minus half sigma x i minus x bar whole square into e to the power minus half n times x bar minus mu whole square. This is because this term is constant and it does not depend upon i. Therefore, as i is equal to 1 to n, the summation gives me n times x bar x bar minus mu whole square and the product term will become 0 because sigma x i minus x bar will become 0. Claim X bar is sufficient for mu. So, what we are claiming that the sample mean is sufficient for mu. So, what we have to show?

(Refer Slide Time: 06:26)



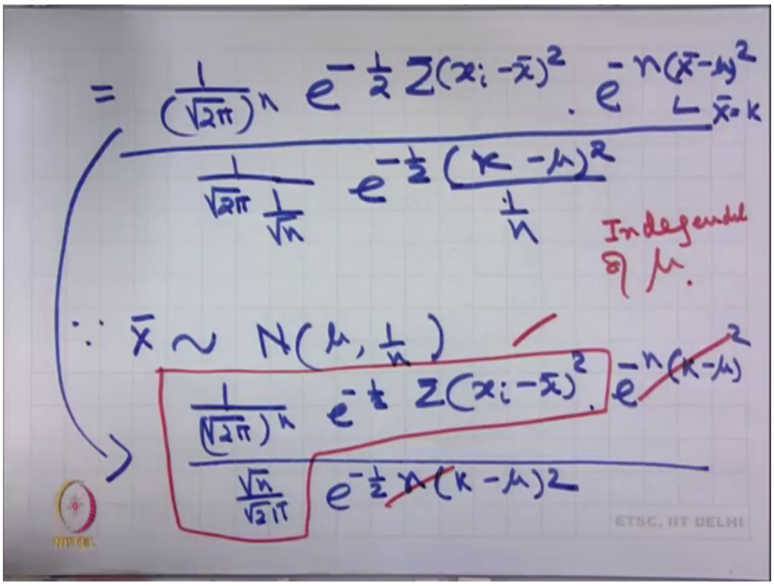
We have to show that joint density of x 1 x 2 x n given x bar is equal to mu is independent of mu. Since X 1 X 2 X n are normal with mean mu and variance 1 X 1 plus X 2 plus X n is distributed as normal with n mu and variance is equal to n. Therefore, X bar is equal to X 1 plus X 2 plus X n by n is distributed as normal with mean mu and variance 1 by n.

(Refer Slide Time: 07:43)



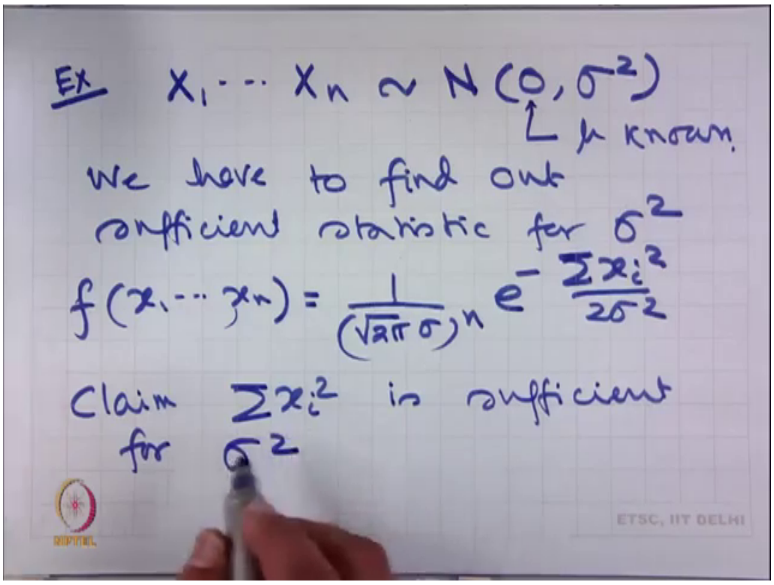
Now, f theta of x 1 x 2 x n, given x bar is equal x bar at k is equal to f theta of x 1 x 2 x n and divided by f x bar at k. If x 1 x 2 x n are such that sigma x i upon n is equal to k or it is 0, otherwise that is if x 1 x 2 x n are such that x bar is not equal to k, then this becomes 0. Otherwise, it is the joint density of x 1 x 2 x n and of course, divided by the density of x bar at k is equal to 1 over root over 2 pi whole to the power n e to the power minus half into sigma x i minus x bar whole square into e to the power minus n into x bar minus mu whole square.

(Refer Slide Time: 08:55)



This we have already seen divided by since, x bar is normal with mean mu and variance is equal to 1 by n, this we can write it as 1 over root over 2 pi into 1 upon root n e to the power minus half into k minus mu whole square divided by 1 upon n and this is x bar is equal to k. Therefore, this is nothing, but 1 over root over 2 pi whole to the power n e to the power minus half sigma x i minus x bar whole square into e to the power minus n k minus mu whole square divided by root n upon root over 2 pi e to the power minus half into n k minus mu whole square. So, now you can see that this cancels with this and whatever remains, this is independent of mu as the term mu does not occur here.

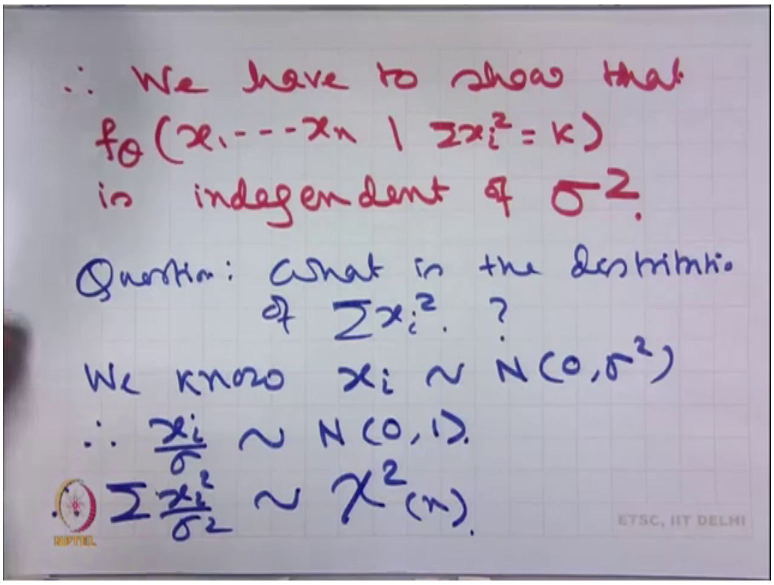
(Refer Slide Time: 11:35)



Therefore, what we can say that x bar is sufficient for estimating mu. Consider again normal population, such that X 1 X 2 X n are from normal 0, sigma square. Here again the mean is fixed. I have kept it at 0 to keep the equation simple. It works with whenever mu is known; we have to find out sufficient statistics for sigma square.

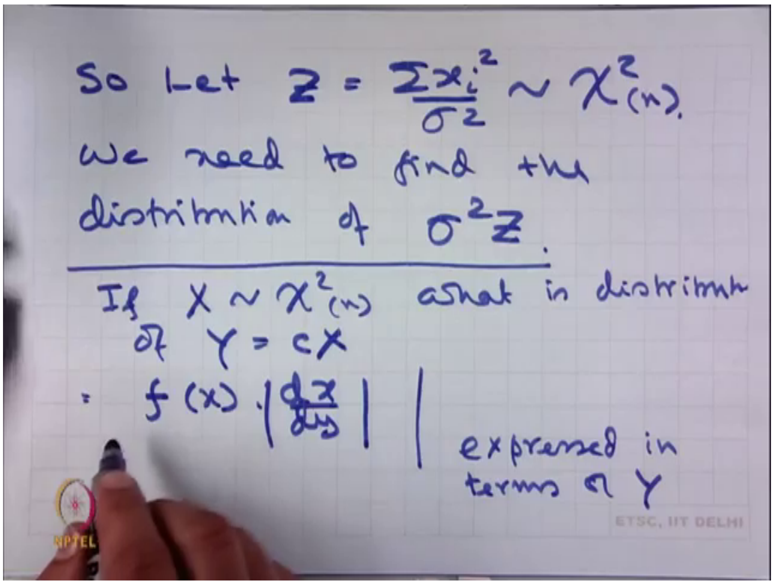
Now, f of x 1 is x n is equal to 1 over root over 2 pi sigma whole to the power n e to the power minus sigma x i square upon 2 sigma square. This is straightforward because again I am multiplying the individual density. Therefore, it is coming out summation as the exponent. So, it is coming as summation of e to the power minus sigma x i square upon 2 sigma square. Claim sigma x i square is sufficient for sigma square.

(Refer Slide Time: 13:37)



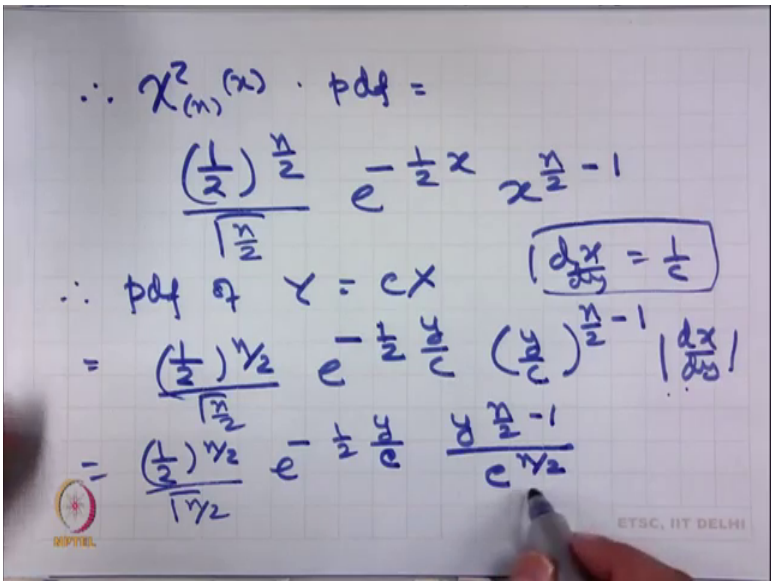
Therefore, we have to show that the joint density of x 1 x 2 x n given sigma x i square is equal to k is independent of sigma square, then only we can show that sigma x i square is sufficient for sigma square. Question what is the distribution of sigma x i square? We know x i is normal with 0 sigma square. Therefore, x i upon sigma is normal with 0 1. Therefore, sigma x i square upon sigma square is distributed as chi square with n degrees of freedom.

(Refer Slide Time: 15:19)



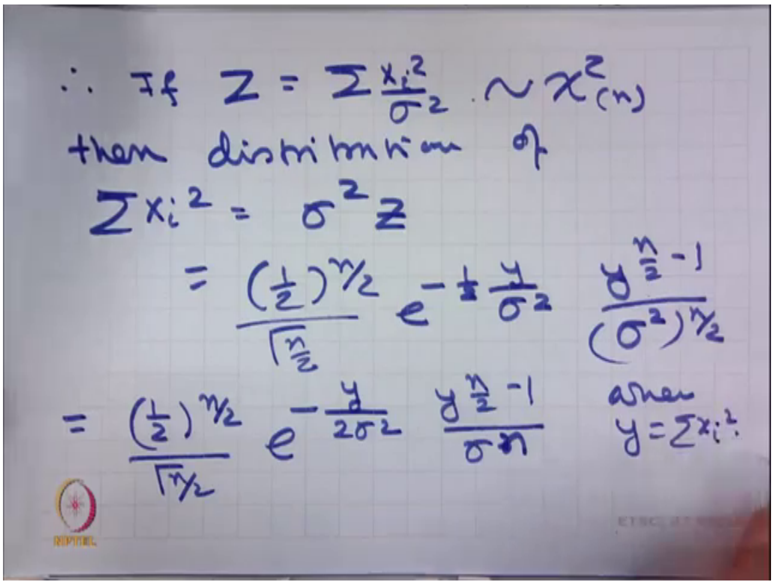
So, let z is equal to sigma x i square upon sigma square which is distributed as chi square with n degrees of freedom. We need to find the distribution of sigma square z. So, if x is chi square with n, what is the distribution of y is equal to c x. We know that this is equal to f at x multiplied by dx dy mod expressed in terms of y. So, this we have seen when we are working on functions of random variables.

(Refer Slide Time: 16:57)



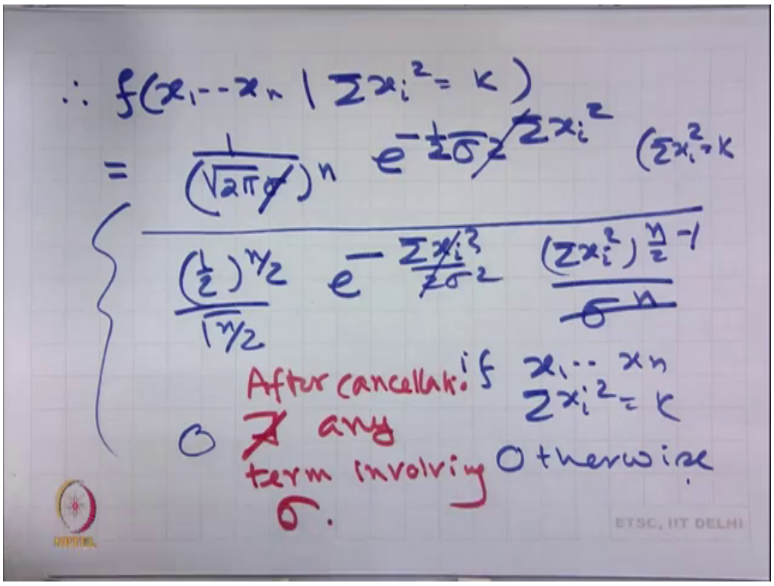
So, this is equal to we know that chi square with n degrees of freedom at x has pdf is equal to half to the power n by 2 upon gamma n by 2 e to the power minus half x, x to the power n by 2 minus 1. Therefore, pdf of Y is equal to CX is half to the power n by 2 gamma n by 2 e to the power minus half y by c, y by c to the power n by 2 minus 1 dx dy and since y is equal to cx, dx dy is equal to 1 by c. Therefore, this is equal to half to the power n by 2 upon gamma n by 2 e to the power minus half y by c y to the power n by 2 minus 1 upon c to the power n by 2 because this is 1 by c. So, this c and this c and that it together we have c to the power n by 2.

(Refer Slide Time: 18:41)



Therefore, if Z is equal to sigma x i square upon sigma square is distributed as chi square with n degrees of freedom, then distribution of sigma X i square is equal to sigma square z is equal to half to the power n by 2 gamma n by 2 e to the power minus half y upon sigma square y to the power n by 2 minus 1 upon sigma square to the power n by 2 is equal to half to the power n by 2 upon gamma n by 2 e to the power minus y 2 sigma square y to the power n by 2 minus 1 upon sigma square sigma to the power n, where y is equal to sigma xi square.

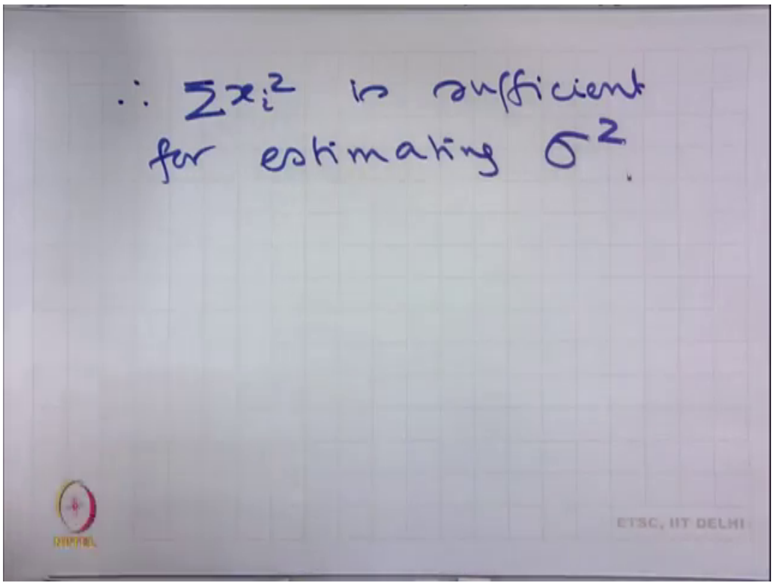
(Refer Slide Time: 20:15)



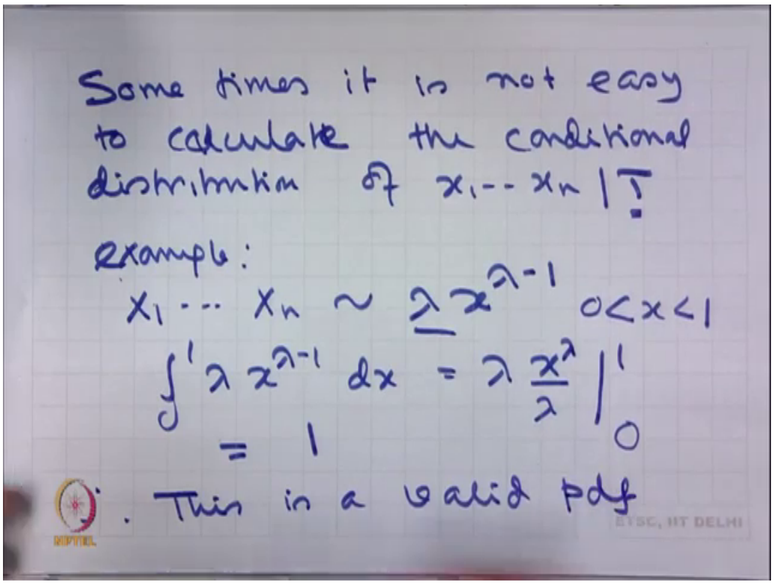
Therefore, f of x 1 x 2 x n given sigma x i square is equal to k is equal to 1 upon root over 2 pi sigma whole to the power n e to the power minus half 2 sigma square sigma x i square sigma. X i square is equal to k divided by half to the power n by 2 gamma n by 2 e to the power minus sigma x i square upon 2 sigma square sigma x i square to the power n by 2 minus 1 upon sigma to the power n if x 1 x 2 x n are such that sigma x i square is equal to k and 0 otherwise.

Now, let us look at this. This sigma to the power n cancels this sigma to the power n and e to the power sigma x i square upon 2 sigma square. That also gets cancelled. So, we can see that after cancellation, there is any term involving sigma. Therefore, what can we say? We can say that the distribution of x 1 x 2 x n given sigma x i square is independent of sigma square and therefore, sigma x i square is sufficient for estimating sigma square.

(Refer Slide Time: 22:40)

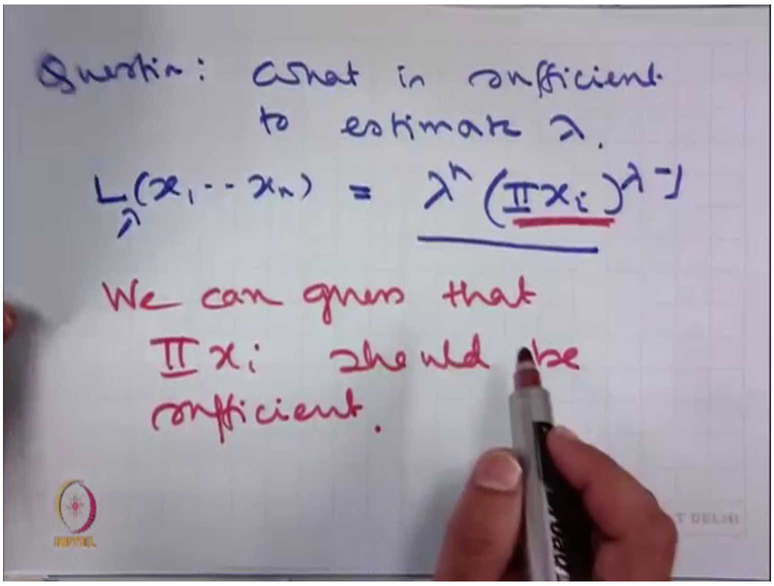


(Refer Slide Time: 22:58)



Sometimes, it is not easy to calculate the conditional distribution of x 1 x 2 x n given T because we need to know the distribution of T. For example, X 1 X 2 X n are from lambda x to the power lambda minus 1 0 less than x less than 1 is a density integration 0 to 1 lambda x to the power lambda minus 1 dx is equal to lambda x upon x to the power lambda upon lambda 1 0 is equal to 1. Therefore, this is a valid pdf.

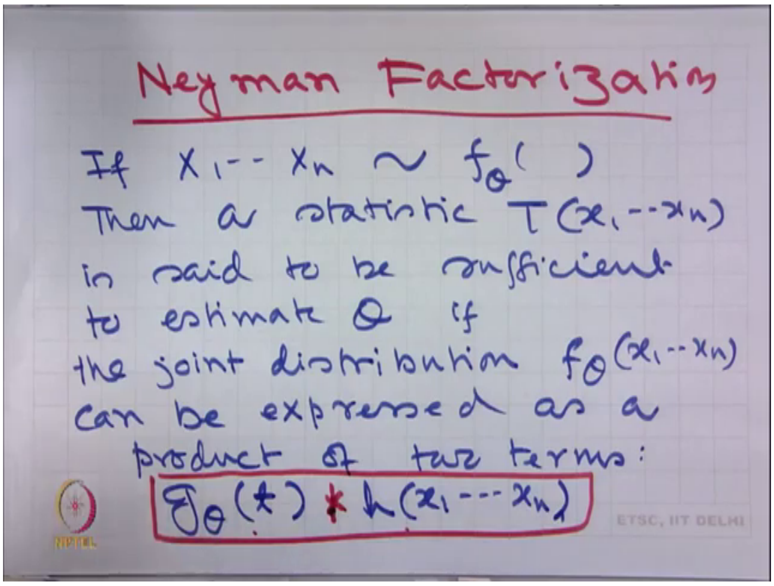
(Refer Slide Time: 24:39)



Question is what is sufficient to estimate lambda? L of x 1 x 2 x n lambda is equal to lambda power n into product of x i to the power lambda minus 1. So, from here we can see that the joint density of x 1 x 2 x n depends upon only the product of x i. So, it appears that to estimate lambda only product of x i is important. We do not need any other information just like that. In Bernoulli, we have observed that the joint density has been a function of sigma x i and therefore, we could guess that sigma x i could be sufficient for p.

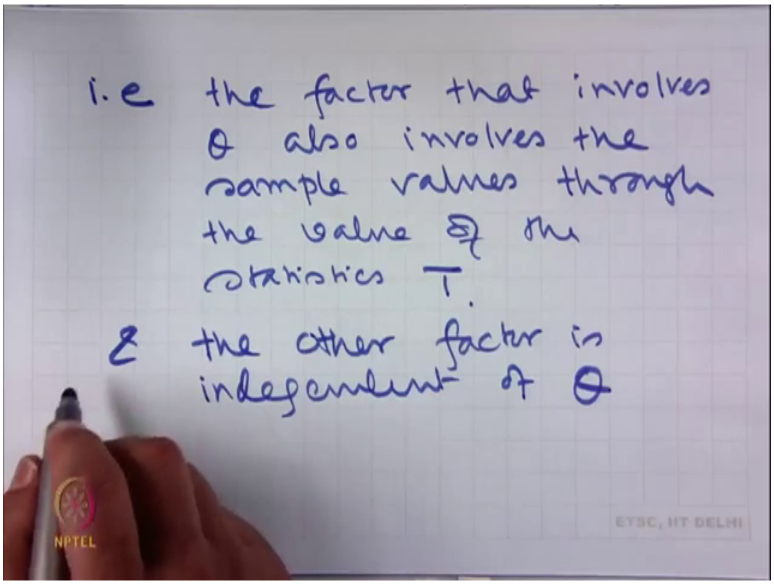
Similarly, in case of normal with 0 sigma square, we could guess that sigma x i square will be sufficient to estimate sigma square. So, we can guess that product of x i should be sufficient, but we cannot easily find that distribution of product of x i. So, we need something else to establish sufficiency and that some something else comes from Neyman Factorization.

(Refer Slide Time: 26:39)



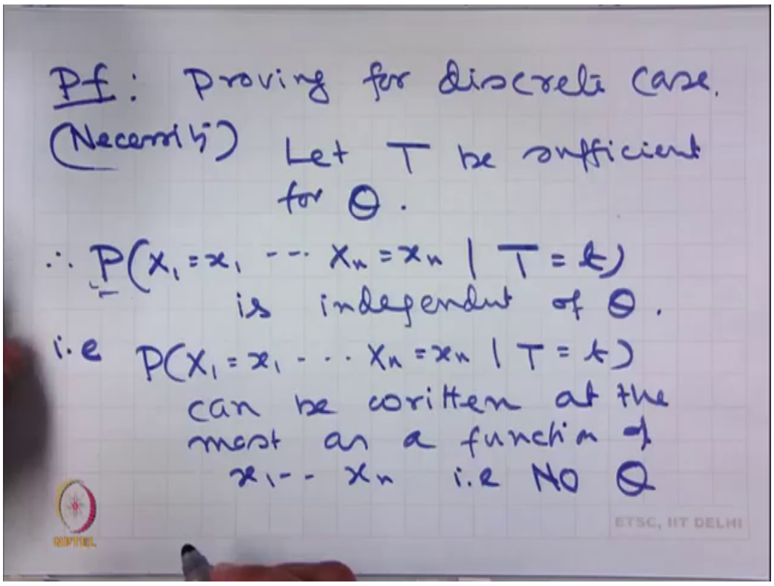
What he says that if X 1 X 2 X n are coming from a distribution with parameter theta, then a statistic T x 1 x 2 x n is said to be sufficient to estimate theta if the joint distribution f theta of x 1 x 2 x n, we can be expressed as a product of two terms g theta of t multiplied by h of x 1 x 2 x n or in other words, let us look at these two terms; the joint distribution can be written as a product of two terms. The first term is involving the statistic t and also it is involving the parameter theta. The product term h of x 1 x 2 x n is a function of the sample values, but it does not involve theta.

(Refer Slide Time: 28:57)



Therefore, what it is saying that is the factor that involves theta, also involves the sample values through the value of the statistics T and the other factor is independent of theta proof.

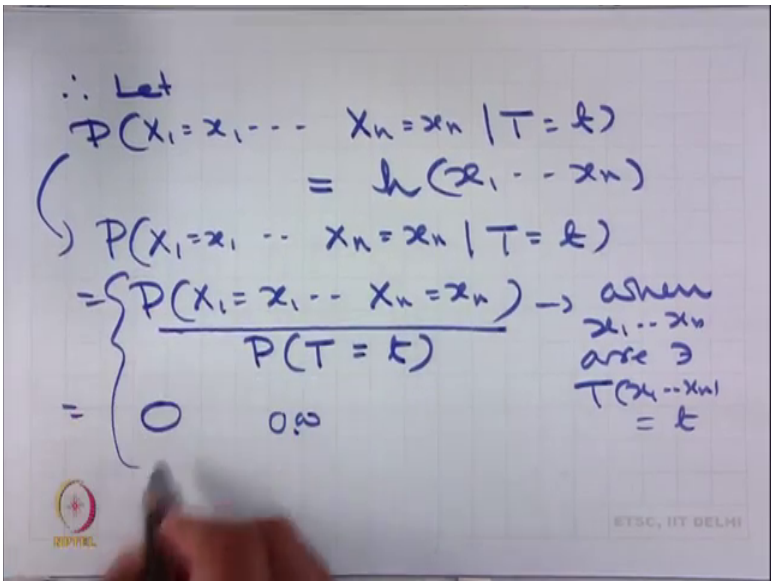
(Refer Slide Time: 29:57)



I am proving it for a discrete case for a continuous case. It can be proved analogously, but since it involves the function T. And therefore, it needs Jacobean that makes it little bit more complicated, but the conceptually it is the same.

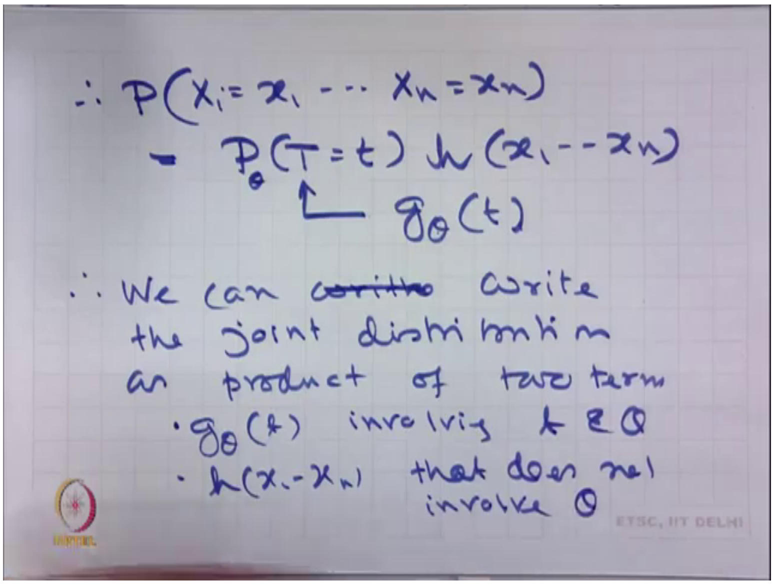
So, the theorem is if and only if we will have to show both the parts, so necessity let T be sufficient for theta. Therefore, probability X 1 is equal to x 1 up to x n is equal to x n given T is equal to t is independent of theta. That is probability. X 1 is equal to x 1 up to x n is equal to x n given T is equal to t can be written at the most as a function of x 1 x 2 x n. That is NO theta.

(Refer Slide Time: 32:15)



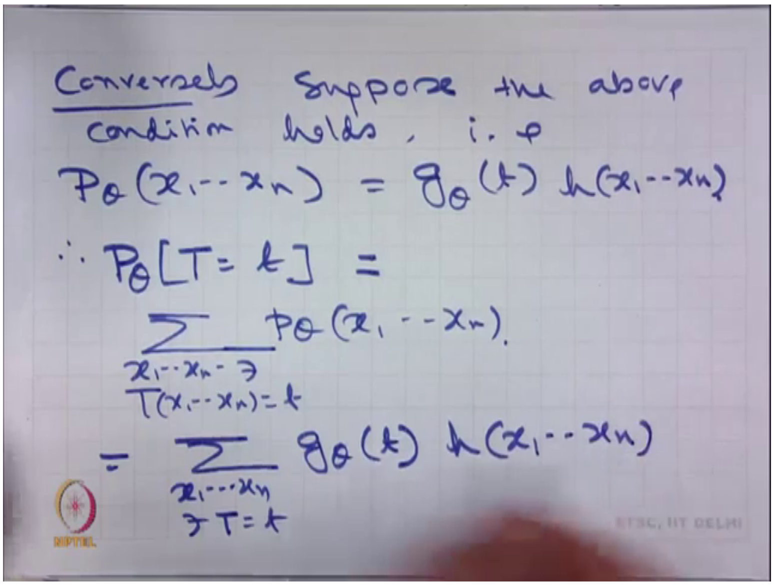
Therefore, let probability X 1 is equal to x 1 up to X n is equal to x n given T is equal to t is equal to say h of x 1 x 2 x n. Now, the left hand side upon probability T is equal to t whereas, before x 1 x 2 x n are such that x 1 x 2 x n is equal to t and it is equal to 0 otherwise. So, we can ignore this part.

(Refer Slide Time: 33:41)



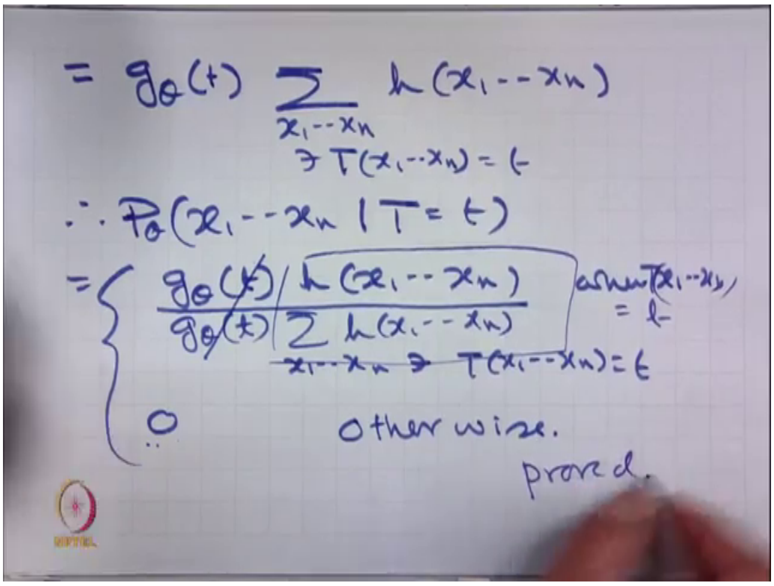
Therefore, probability X 1 is equal to x 1 X n is equal to x n is equal to probability T is equal to t into h of x 1 x 2 x n. Let us call it g theta of t because this probability will automatically involve the parameter theta. Therefore, we can write the joint distribution as product of two terms g theta t involving t and theta and h of x 1 x 2 x n that does not involved theta conversely.

(Refer Slide Time: 35:07)



Suppose the above condition holds that is P theta of x 1 x 2 x n is equal to g theta of t into h x 1 x 2 x n. Therefore, probability theta of T is equal to say t is equal to summation over all those x 1 x 2 x n, such that T x 1 x 2 x n is equal to t of P theta of x 1 x 2 x n. That is all those values of x 1 x 2 x n that generates the value of the statistics T x 1 x 2 x n to be t. I have to sum them up to find the probability that the statistic T is equal to small t is equal to sigma over x 1 x 2 x n, such that T is equal to t g theta of t into h of x 1 x 2 x n is equal to g theta of t sigma x 1 x 2 x n such that T x 1 x 2 x n is equal to T h of x 1 x 2 x n.

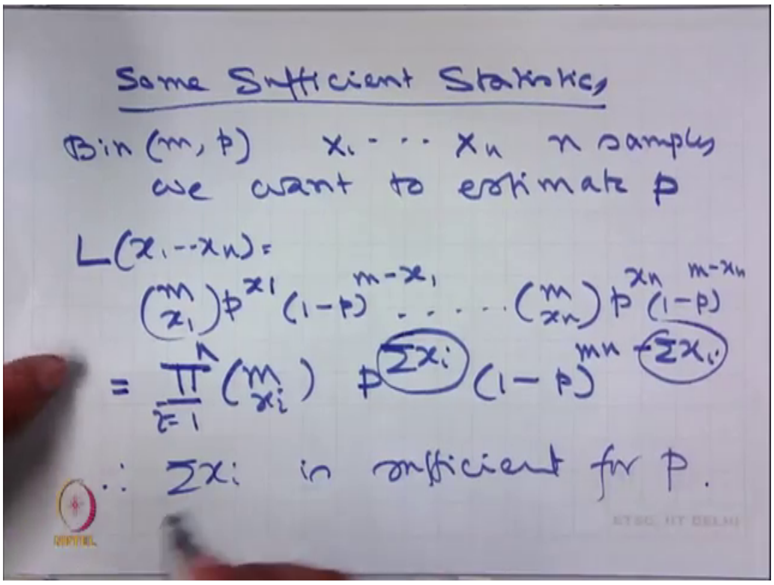
(Refer Slide Time: 36:59)



Therefore, the conditional density P theta of x 1 x 2 x n given T is equal to t is equal to g theta t h of x 1 x2 x n divided by g theta of t sigma h of x 1 x 2 x n, such that x 1 x 2 x n are such that T x 1 x 2 x n is equal to t from here, otherwise it is 0. That is when T of x 1 x 2 x n is equal to t. Now, if I look at this, this is surely independent of theta and if I look at this, this cancels out. Therefore, the term remains is completely independent of theta.

Therefore, it satisfies that the conditional density of x 1 x 2 x n given T is equal to small t is independent of theta. Therefore, that is a necessary and sufficient condition to check if the given t is a sufficient statistic to estimate theta.

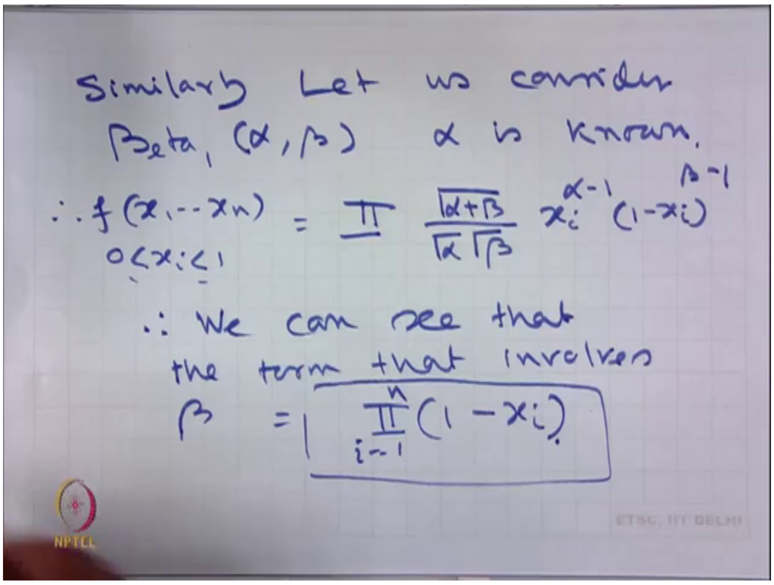
(Refer Slide Time: 39:14)



When you look at the joint density, if we look at the term involving x 1 x 2 x n in which form it is associated with the joint density that gives us a clue of how to obtain a sufficient statistic. For example, binomial with m, p and suppose X 1 X 2 X n are n samples we want to estimate p, then the joint density of L x 1 x 2 x n is equal to mc x 1 p to the power x 1 1 minus p to the power m minus x 1 into up to mc x n p to the power x n 1 minus p whole to the power m minus x n is equal to product of mc x i, i is equal to 1 to n p to the power sigma x i 1 minus p to the power m, n minus sigma x i.

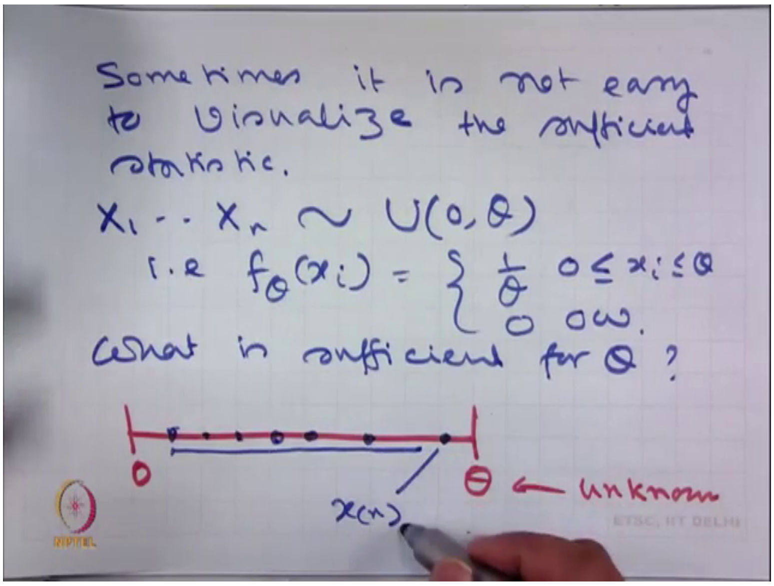
Therefore, if you look at the joint density of x 1 x 2 x n, we find that the sample value are involved here in the form of sigma x i. What does it say? It says that sigma x i is good enough for us to estimate p. Therefore, sigma x i is sufficient for p.

(Refer Slide Time: 41:37)



Similarly, let us consider beta 1 alpha beta where alpha is known. Therefore, f of x 1 x 2 x n 0 less than x i less than 1 is equal to product of gamma alpha plus beta upon gamma alpha gamma beta x i to the power alpha minus 1 1 minus x i to the power beta minus 1. Therefore, we can see that the only term that involves beta is equal to product of 1 minus x i. Therefore, we can say that this is sufficient to estimate beta.

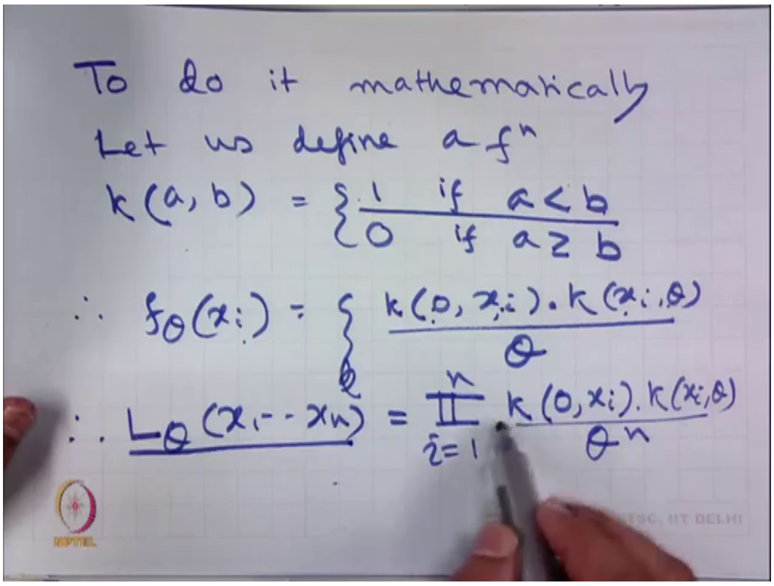
(Refer Slide Time: 43:19)



Still sometimes it is not easy to visualize the sufficient statistic. For example, X 1 X 2 X n are from uniform 0 theta that is f theta of x i is equal to 1 by theta 0 less than equal to x i less than equal to theta 0, otherwise what is sufficient for theta suppose I ask you the question. Now, 0 to theta this theta is unknown to us.

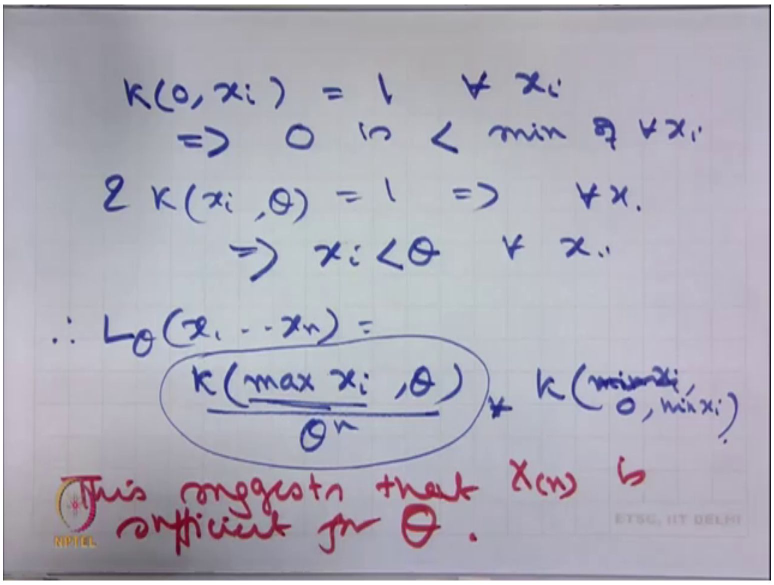
Now, I have taken n samples from here, what we can say that this is the highest order statistic. Only thing that we can say is theta is greater than equal to this value and since theta is greater than equal to this value, we know none of these observations have any influence on that decision making or in other words, the best estimate that we can do for theta is true the nth order statistic or x n. Therefore, we can say that x n is the sufficient statistic to estimate theta how to do it mathematically.

(Refer Slide Time: 45:45)



Let us define a function kappa of a, b which is 1 if a less than b and 0 if a is greater than equal to b. Therefore, f theta of x i is equal to kappa of 0, x i times kappa of x i, theta divided by theta and it will be 1 only if x i is greater than 0 or greater than equal to 0 and less than theta. So, only in these cases when a less than b kappa is 1, so when 0 is less than x i and x i is less than theta, then this will be 1 and therefore, this is going to be let theta of x i for each x i. Therefore, L theta of x 1 x 2 x n is equal to product of i is equal to 1 to n kappa 0 of x i kappa x i of theta upon theta to the power n or in other words, this is saying that the joint density function is a product of these terms.

(Refer Slide Time: 48:17)



Now, kappa 0 x i is equal to 1. For all x i implies 0 is less than minimum of all x i and kappa x i, theta is equal to 1 implies for all x i implies x i is less than theta for all x i. Therefore, L theta of x 1 x 2 x n can be written as kappa of max of x i theta upon theta to the power n multiplied by kappa of minimum of x i 0 minimum of x i. Therefore, the term that involves theta is maximum over x i say for these suggest that x n is sufficient for theta, ok.

Students with that I stop here today. So, over the last 3 lectures, we have studied different properties of the estimators namely unbiasedness, consistency, efficiency and today we have seen examples of sufficient statistic. In the next class, I shall be dealing with how do we actually estimate a parameter from the observations x 1 x 2 x n or in other words, methods of estimation.

Thank you so much.