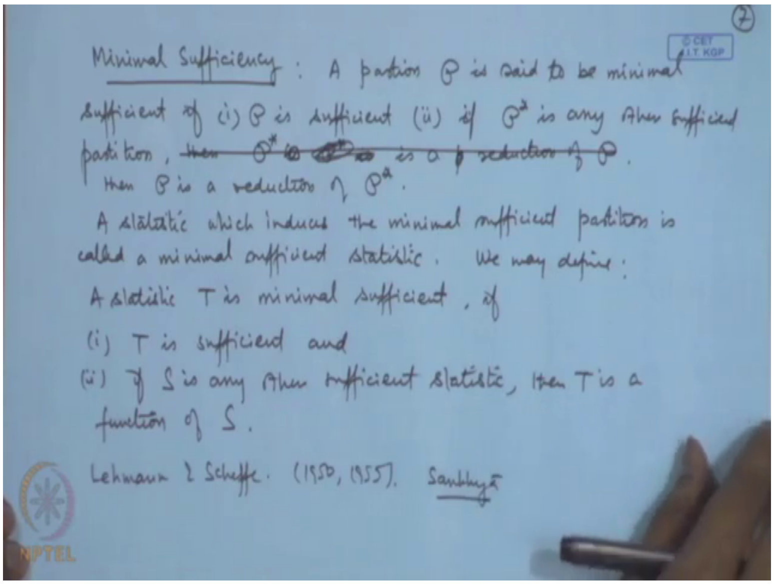
Statistical Inference

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

Lecture – 28
Minimal Sufficiency, Completeness – II

Now, we define the concept of Minimal Sufficient partition and minimal sufficient a statistics.

(Refer Slide Time: 00:31)

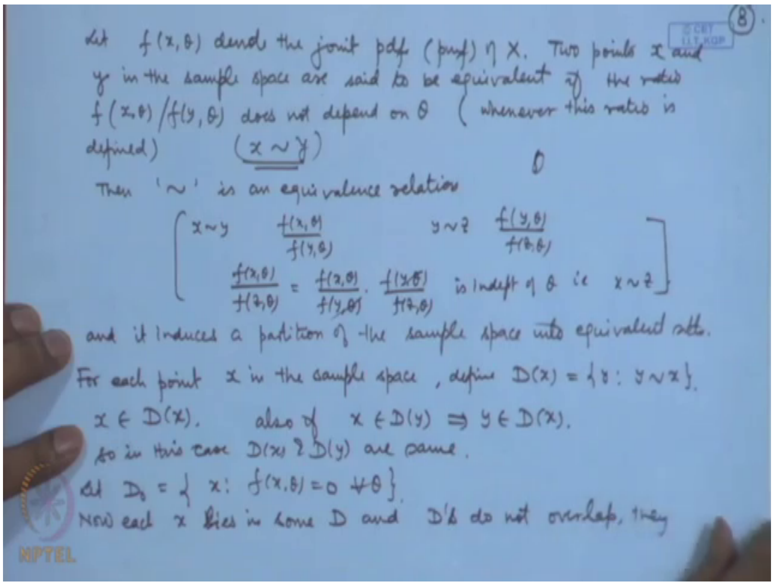


So, a partition say P is said to be minimal sufficient if this is sufficient and second if P star is any other sufficient partition, then P is then P star is sorry; then P star is a reduction of P. No, I am sorry this is written wrongly. If P star is any other sufficient partition, then P is a reduction of P star. So, let me explain we will call it minimal sufficient partition if first of all this should be sufficient partition.

And if there is any other sufficient partition then this should be a reduction of that. That is why this is the maximal reduction or we say that it is a minimal sufficient partition. So, a statistic which induces the minimal sufficient partition is called a minimal sufficient statistic. So, we can say that, a statistic T is minimal sufficient, if it is sufficient and if S is any other sufficient statistic, then T is a function of S.

So, that is how it is the minimal sufficient that is the maximal reduction of the data. Now, the question is that how to determine a minimal sufficient statistic or a minimal sufficient partition in a given problem. This problem is settled by Lehman and Scheffe in 1950 and 55 in papers in Sankhya. We consider here the case when the distribution is either discrete or continuous.

(Refer Slide Time: 04:27)



So, let us consider f x theta let f x theta denote the joint probability density function or probability mass function of X; that means, we have observations X 1 X 2 X n which we are calling as X here. Now two points x and y in the sample space are said to be equivalent if the ratio f x theta by f y theta does not depend on theta. Of course, when we write the ratio of the densities at two different variable points and there is a possibility that either the numerator or the denominator may be 0 or both may be 0.

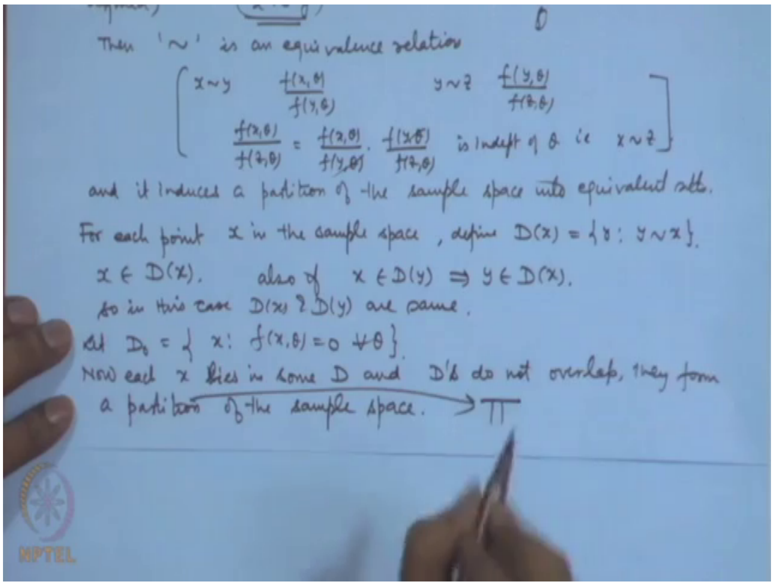
So, in that case we qualify this statement by saying whenever this ratio is defined. So, this we say that x and y are equivalent and we use the notation x is equivalent to y. Then, this relation is an equivalence relation because it is reflexive if I consider f x theta by f x theta that is going to be 1 which is free from the parameter. If f x theta by f y theta is free from theta then f y theta by f x theta is also free from the parameter. Therefore, x related to y is equivalent to saying y is equivalent to x so the relation is symmetric.

If I say x is equivalent to y that is f x theta by f y theta is independent of theta and if I say y is related to z or y is equivalent to z then f y theta by f z theta is independent of theta.

So, if I consider f x theta by f z theta then that is equal to product of these two terms. So, that is also free from so this is independent of theta that is we can say x is equivalent to z. So, the relation is also transitive. So, this is an equivalence relation and it induces a partition of the sample space into equivalent sets. That means, if I consider one set in this partition class then within that class all the points will be equivalent and if I take two different partition sets then the points in that will not be equivalent.

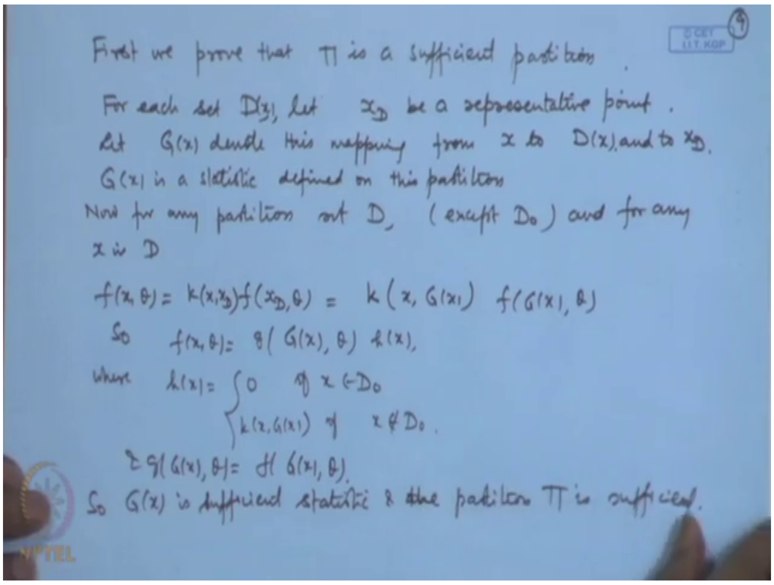
So, now let us consider for each point x in the sample space, define D x as the set of all the y's such that y is equivalent to x. That is for every point, whatever be the equivalent points I will put them in the set D x. Then x belongs to D x and also if x belongs to D y then y will belong to D x. So, in this case D x and D y are same. And also there will be place where the density will take value 0 that is density or the probability mass function we put in another set. Let D naught be the set of all those points for which f x theta is equal to 0 for all theta.

(Refer Slide Time: 10:03)



So, now each x lies in some D and D's do not overlap, they form a partition of the sample space. Let us call this partition pi this partition I will name as pi.

(Refer Slide Time: 10:29)

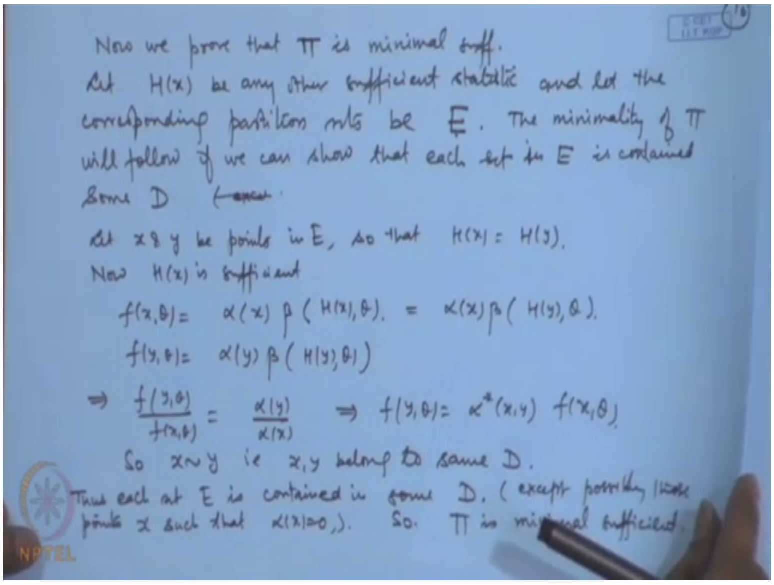


First we prove that pi is a sufficient partition. Now, let us consider for each set D, let x D be a representative point ok. Now, let G x denote this association; that means, from D to x D we are having a mapping. So, let G x denote this mapping from x to D x and to x D. So, for a given point x we have the point D x and then I am choosing a representative point D x of that set. That means, in this set D x all the points are equivalent to each other and I choose I specify one point x D there.

So, G x is a statistic defined on this partition. Now for any partition set D, of course, I am not considering D naught where f x theta is 0 and for any x in D let us write f x theta. Now x belongs to D x and x D also belongs to this. So, f x theta divided by x D theta is free from the parameter; that means, this is a multiple of f x D theta by a term which we can say it is free from theta it is a function of x and x D. So, we can call it a function of x and G x and f G x theta, this x D I am writing as G x which we can write as f x theta is equal to g of G x theta m to h of x. Where, H of x is actually 0 if x belongs to D naught and it is equal to k of x G x if x does not belong to D naught. And g of G x theta is nothing, but f of G x theta.

Now, if you see this carefully this is nothing, but the factorization here. So, we conclude that G x is in a sufficient statistic and the partition pi is sufficient partition because, that is induced by g.

(Refer Slide Time: 14:27)

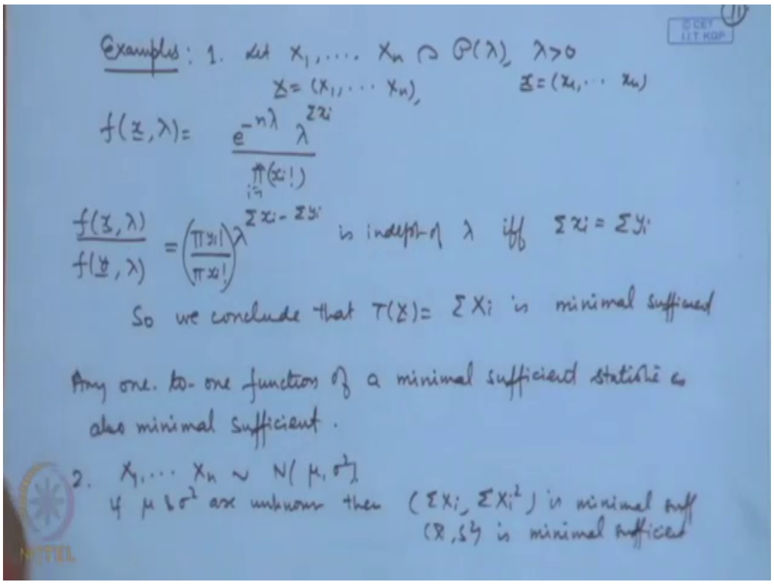


Now let us consider. Now, we prove that pi is minimal sufficient. For that let us consider another say H x that H x be any other sufficient statistic and let the corresponding partition sets be let me call them E. The partition sets induced by pi were D and the partition sets induced by E let me call by H let maybe call it to be E.

Now, if we can show that the minimality of pi will follow if we can show that each set in E is contained in some D. Except of course, the points where the probability is 0. So, let us consider x and y be points in E, so that say H x is equal to H y. Now H is sufficient so we can write f x theta is equal to say alpha x into beta of H x theta. Now, that we can write as alpha x beta of H y theta and f y theta we can write as alpha y beta H y theta. So, if I take the ratio here we get f y theta by f x theta is equal to alpha y by alpha x; that means, we can say f y theta is equal to a function of say x y into f x theta.

So, x is equivalent to y that is x y belong to same D. Thus, each set E is contained in some D. Of course, except possibly those points x such that alpha x is equal to 0. So, pi is minimal sufficient. Because, pi is a reduction of this partition that we have introduced second partition. So, this gives us a method of determining the minimal sufficient statistics. What we consider that you take ratio f x theta divided by f y theta and this should be free from the parameter. So, what is the partition that will induce this condition and the corresponding sufficient statistic corresponding statistic then if you find out that will be minimal sufficient. So, let me explain through some examples.

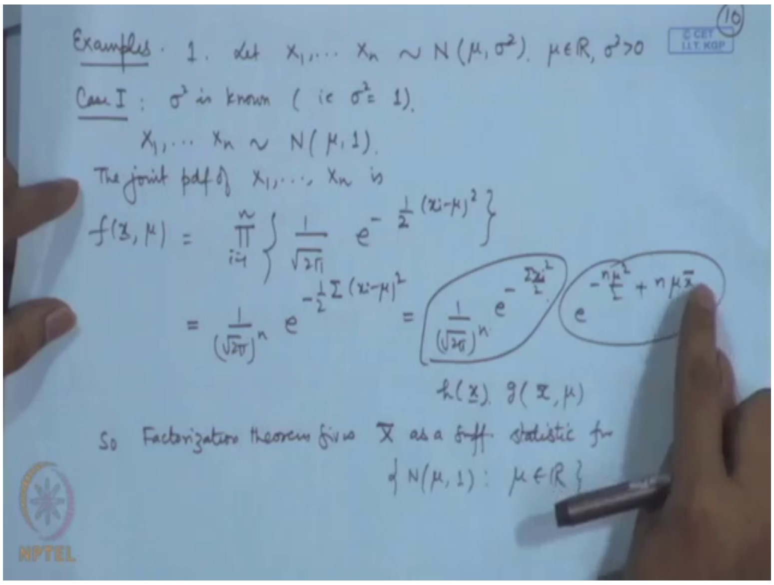
(Refer Slide Time: 19:23)



Let us consider the cases of the standard estimations suppose I consider Poisson lambda distribution. And we denote by X the X 1 X 2 X n and by small x we denote the points. So, consider the f x lambda here that is e to the power minus n lambda lambda to the power sigma x i divided by product of x i factorial i is equal to 1 to n. So, let me consider the ratio f x lambda divided by f y lambda then that is equal to e to the power minus n lambda will cancel out we will get lambda to the power sigma x i minus sigma y i and product of y i factorial divided by product of x i factorial.

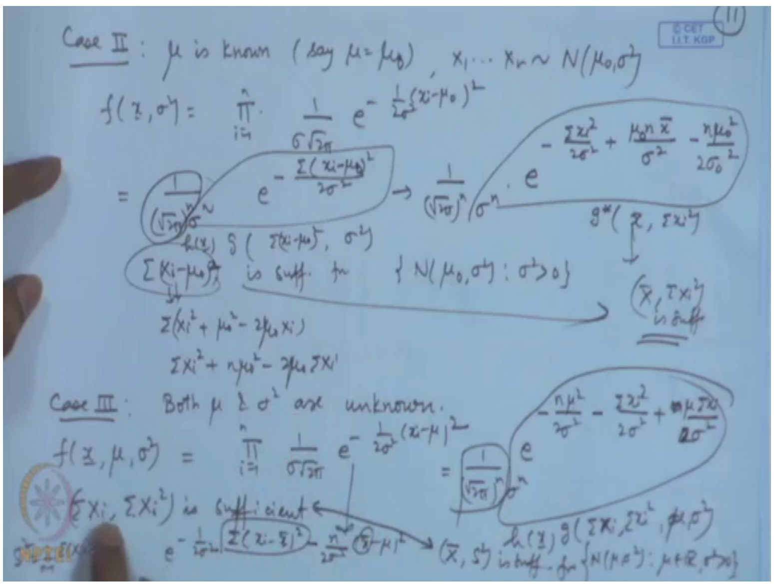
Now, this term is dependent upon parameter through this and we can easily see that this is independent of lambda if and only if sigma x i is equal to sigma y i. So, whether previous result that we have proved of laid by Lehman and Scheffe we conclude that T X is equal to sigma X i is minimal sufficient. Of course, we can say that any one to one function of a minimal sufficient statistic is also minimal sufficient. Let me just take up the cases of the sufficient statistics that we worked out in the previous classes. We had seen like binomial distribution, normal distribution, exponential distribution etcetera let us look at each of those cases and see what were the sufficient statistics.

(Refer Slide Time: 22:11)



Consider this case, X 1 X 2 X n is a random sample from normal mu sigma square and sigma square is known. Now, in this case the joint distribution that we wrote was of the form 1 by root 2 pi to the power n e to the power minus sigma x i square by 2 e to the power minus n mu square plus n mu x bar. Now in this case if I consider the ratio by taking f x mu divided by f y mu this term will become free from the variable free from the parameter e to the power minus n mu square will also cancel out we will be left with e to the power n mu x bar minus y bar. Now that will be free from mu if and only if x bar is equal to y bar and therefore, X bar is the minimal sufficient statistics.

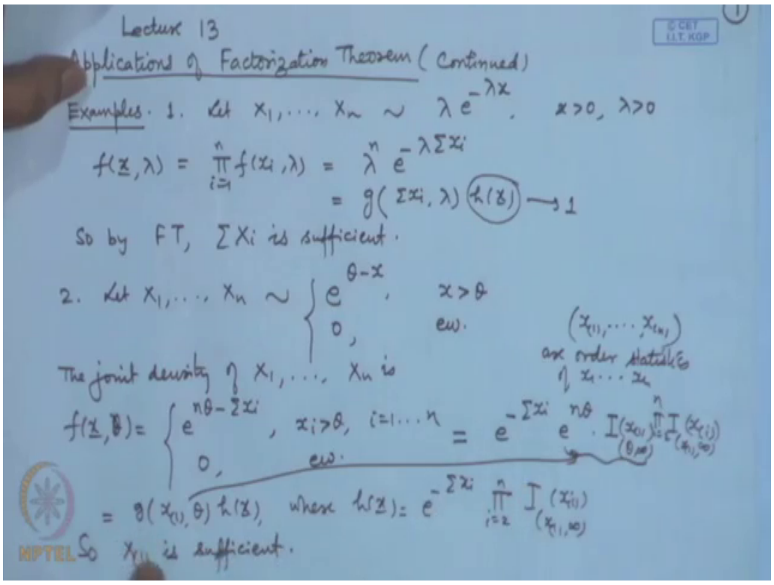
(Refer Slide Time: 23:07)



So, like that if we consider various problems like in the second case we have taken mu naught is known. And in this case we figured out that sigma X i minus mu naught whole square is sufficient so this will also become minimal sufficient. When both mu and sigma square are unknown then sigma X i and sigma X i square will become minimal sufficient. So, in most of the problems where we have applied factorization theorem we actually have a factorization. So, if we write down the ratio then the term which is consisting of the parameter theta there then it is related to g of T X theta divided by g of T Y theta.

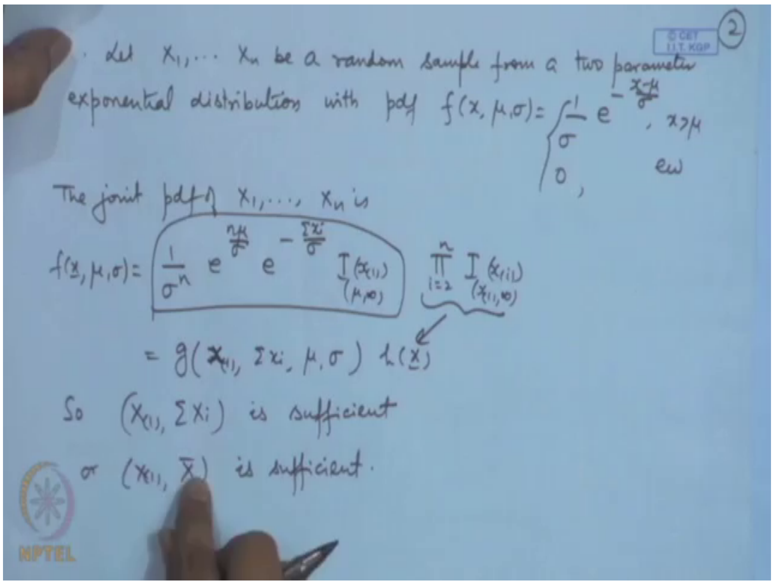
So, this ratio if you consider and obtain the condition when it is going to be free from the parameter, that will give the minimal sufficient statistics. So, like that if I just mention X 1 X 2 X n follow normal mu sigma square. So, if mu and sigma square are unknown then sigma X i and sigma X i square is minimal sufficient. Of course, you can say X bar and S square is minimal sufficient. And we can answer various other questions let me just tell few of this here.

(Refer Slide Time: 24:57)



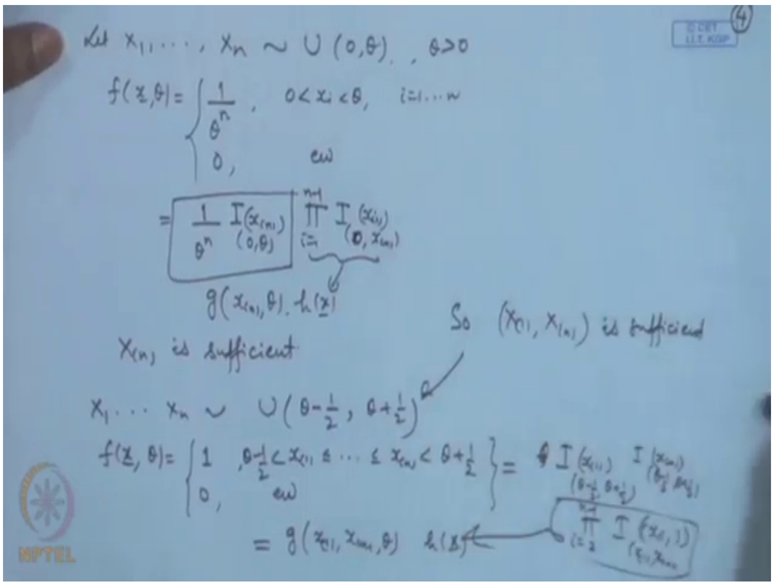
Let us consider say exponential distribution with parameter lambda. Here if I write down the ratio we will get sigma X i as the minimal sufficient.

(Refer Slide Time: 25:17)



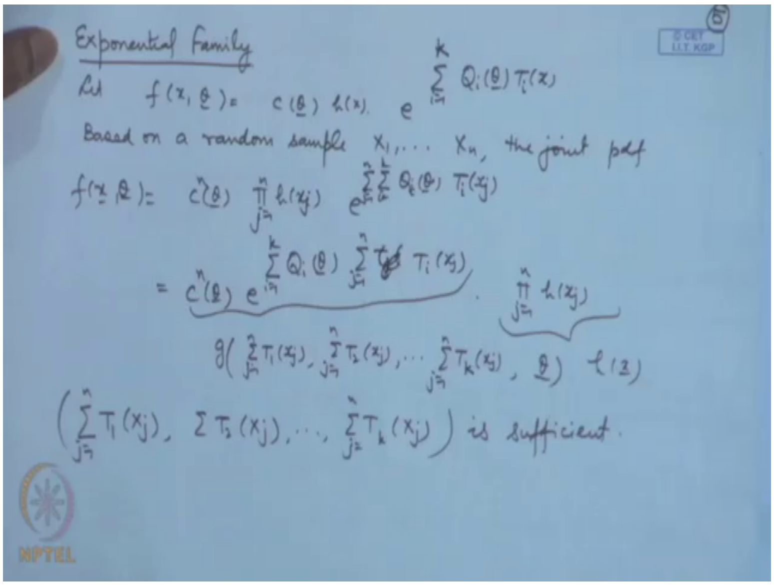
If we consider exponential distribution with location parameter then X 1 will be turning out to be minimal sufficient. If we consider say two parameter exponential distribution with parameter mu and sigma here then X 1 and X bar or X 1 and sigma X i will be minimal sufficient. If we consider say a double exponential distribution in that case the full sample which is written in by a reduced to the order statistics that will be minimal sufficient.

(Refer Slide Time: 25:43)

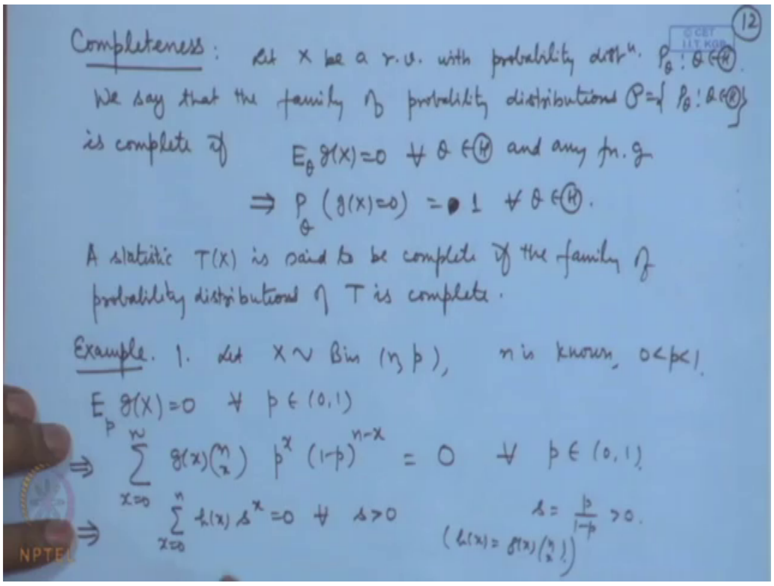


If we consider uniform distribution on the interval 0 to theta then X n will be minimal sufficient. If we are considering exponential family then this statistic that we have written this will be minimal sufficient.

(Refer Slide Time: 25:51)



(Refer Slide Time: 26:23)

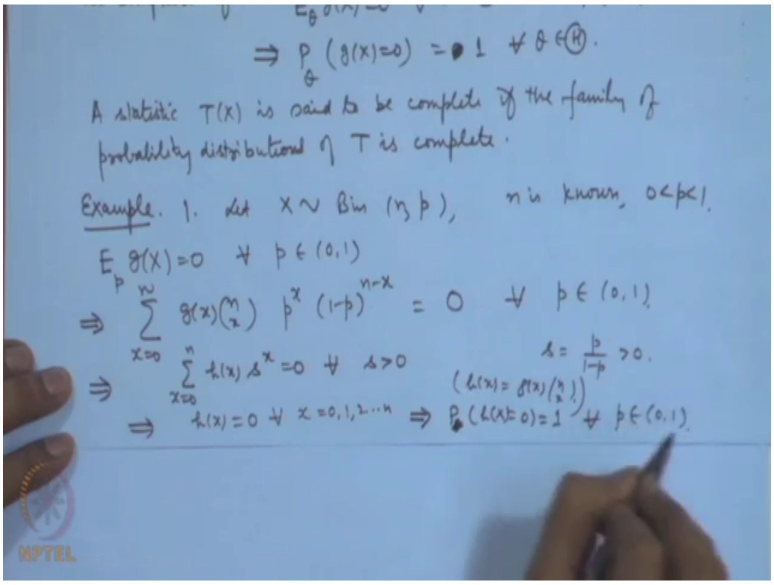


Let me introduce another concept that is completeness. Let X be a random variable with probability distribution P theta; theta belonging to theta. So, we say that the family of probability distributions P that is equal to P theta; theta belonging to theta is complete if, expectation theta g x is equal to 0 for all theta belonging to theta and any function g implies that probability that g x is equal to 0 is 0 is 1 for all theta belonging to theta. Then a statistic T is said to be complete if the family of probability distributions of T is complete.

Let me give an example here. Let X follow say binomial n p distribution where n is known and parameter p lies between 0 to 1. Let us consider expectation of g x is equal to 0 for all p in the interval 0 to 1. Now, this statement is equivalent to g x n x p to the power x 1 minus p to the power n minus x x is equal to 0 to n that is equal to 0 for all p belonging to 0 to 1.

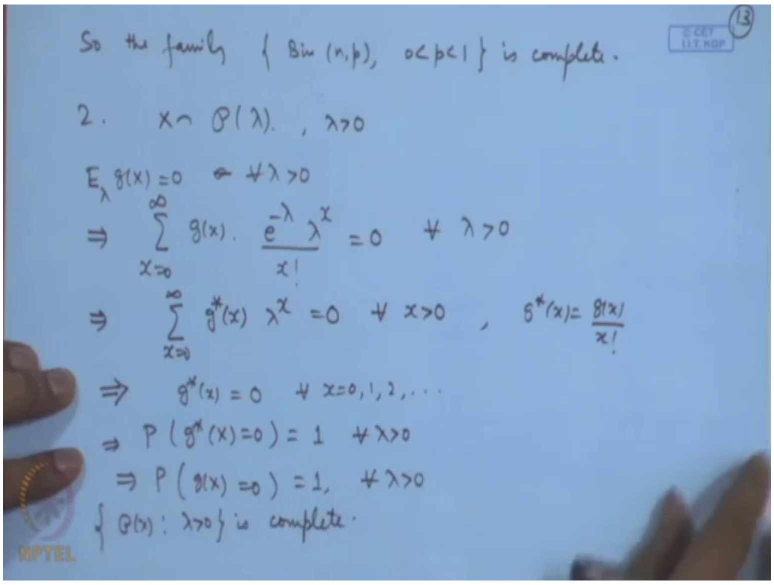
Now this we can also write as see 1 minus p to the power n we can cancel out on both the sides and let us write say let me write say s is equal to p divided by 1 minus p. So, this will be any positive term. So, we can say h x into s to the power x x equal to 0 to n is equal to 0 for all s greater than 0 where h x is nothing, but the function g x into n x ok. Now, if you see this left hand side this is a polynomial of degree n in s and I am saying it is vanishing identically over an interval.

(Refer Slide Time: 29:49)



This implies that h x must be 0 for all x; now, here for all x means because x can take values 0 1 to n. This means that probability that h x is equal to 0 is 1 for all p. So, the family of binomial distributions is complete.

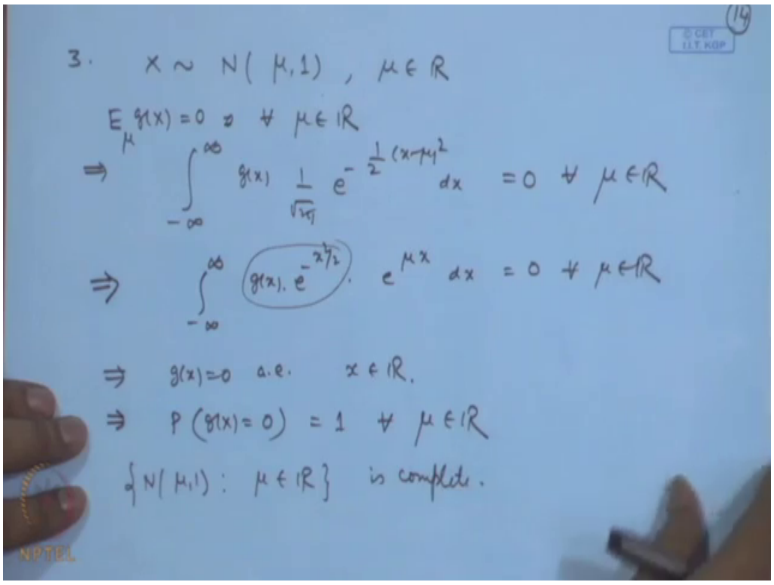
(Refer Slide Time: 30:15)



So, the family of binomial distributions n p where p lies between 0 to 1 is complete or we can say here x is a complete statistic. Let us take another example say X follows Poisson lambda. Then lambda is a positive parameter here, let us write down the statement expectation of g x is equal to 0 for all lambda. Now, this is equivalent to sigma g x e to the power minus lambda lambda to the power x by x factorial x is equal to 0 to infinity that is 0 for all lambda greater than 0. Now e to the power minus lambda is a positive term so we can multiply by e to the power lambda on both the sides. This statement becomes equivalent to say g star x in to lambda to the power x where this g star is nothing, but g x by x factorial.

Once again if you notice on the left hand side I have a power series in lambda which is vanishing identically over the positive half of the real line. So, if a power series vanishes identically over an interval all the coefficients must vanish. So that means, this is equal to 0 for all x equal to 0 1 2 and so on. Therefore, we can say that probability that g star X is equal to 0 is 1 for all lambda. Now g star is nothing but g x by x factorial; that means, g X itself is 0 with probability 1. So, this family of probability distributions of Poisson lambda is complete.

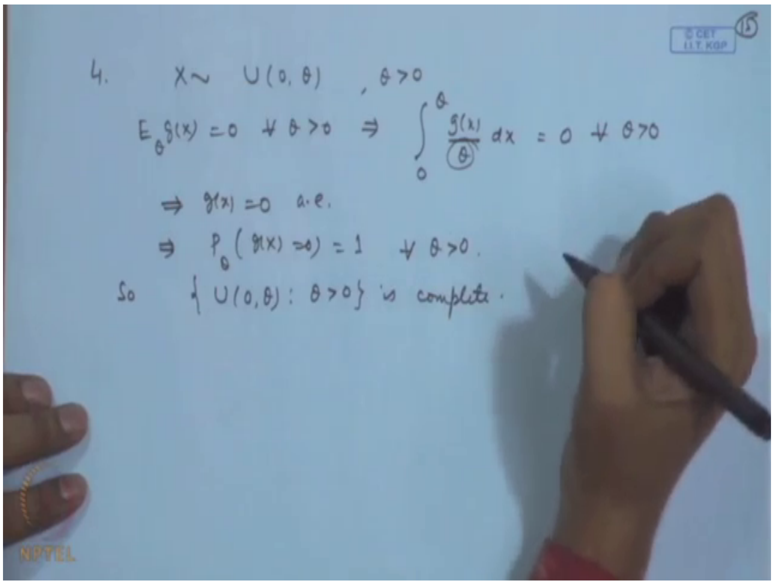
(Refer Slide Time: 32:23)



Let us consider say X follows normal mu 1, here mu is any real number let us write down expectation of g X is equal to 0 for all. Now, this is equivalent to saying g x 1 by root 2 pi e to the power minus 1 by 2 x minus mu square dx. This we may write as g x into e to the power minus x square by 2 e to the power mu x dx is equal to 0 for all mu belonging to R.

Now this is nothing, but the bilateral or bi-variate Laplace transform of this function and we are saying this vanishes identically and therefore, the function itself should vanish. That means, we should have g x is equal to 0 almost everywhere on x real line. This means that probability that g X is equal to 0 is 1 for all mu. So, the family of the normal distributions is complete family.

(Refer Slide Time: 33:59)



Let us consider X following uniform 0 theta, expectation of g X is equal to 0 for all theta this is equivalent to the statement g x by theta dx 0 to theta is equal to 0. Now this term I can adjust here. So, what we are saying is that the integral of g x is 0 for all values for over all the intervals of the form 0 to theta. Therefore, we can using the Lebesgue integration theory we can say that the function gx itself is 0 almost everywhere; that means, probability that g x is equal to 0 must be 1 for all theta.

So, the family of uniform distributions is also complete. So, what in the next lecture I will give a general framework for the completeness. We will also define bounded completeness and the consequence of the sufficiency and completeness is that we can easily derive uniformly minimum variance unbiased estimators. So, we will give these applications in the next lecture.