Statistical Inference

Prof. Somesh Kumar

Department of Mathematics

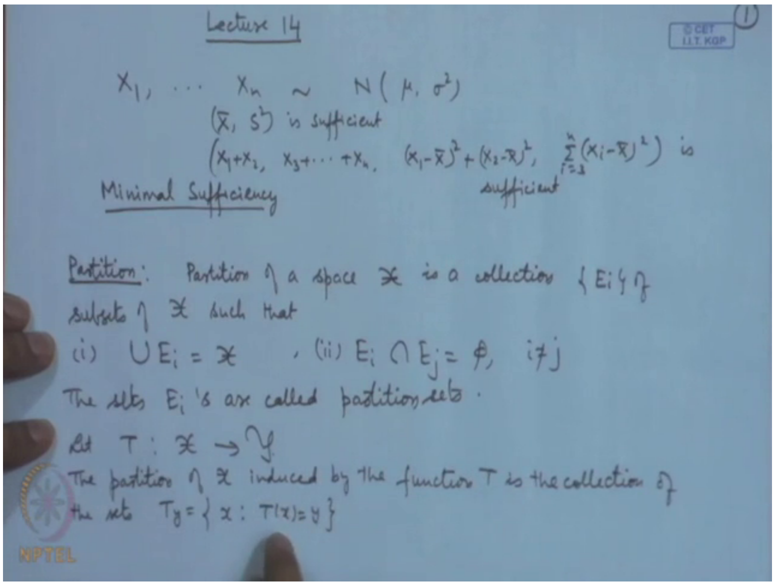
Indian Institute of Technology, Kharagpur

Lecture No. # 14

Minimal Sufficiency, Completeness

We have considered the concept of sufficiency and I related it to the Fisher's information measure. We showed that if a statistic is sufficient then the information contained in the sufficient statistic is the same as the information contained in the whole sample. However, we have also seen that for a given problem there can be various sufficient statistics.

(Refer Slide Time: 00:51)

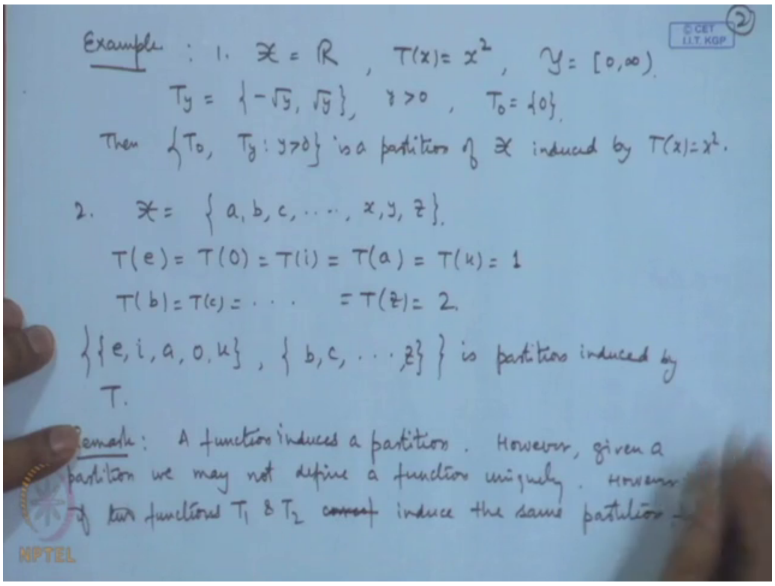


Suppose, I consider a random sample X 1, X 2, X n. Say X 1, X 2, X n from say normal mu sigma square, then by using factorization theorem we showed that X bar and x square is sufficient, but using the same argument we may also say that X 1 plus X 2, X 3 plus and so on X n and similarly we can say X 1 minus X bar square plus X 2 minus X bar square and so on. Say X i minus bar whole square I equal to 3 2 n this also sufficient and like that we can write several sets of sufficient statistics.

So, now which one to use then naturally it should occur that the one which leads to the maximum reduction of the data should be used, that leads to the concept of minimal sufficiency. However, this concept I will introduce through the concept of partition. So, let me introduce the concept of partition. So, a partition of a space say X is a collection say E i of subsets of x such that union of E i is equal to x and E i intersection E j is equal to phi far i not equal to j. That means, the it is a mutually exclusive sets and the union is equal to the full.

That means mutually exclusive and exhaustive subsets of a space that is called a partition. The sets E i they are called partition sets. Let T be a function from say X into another space y, then the partition of X induced by the function T is the collection of the sets say T y which is defined as a set of all those values x such that T X is equal to y. That means, corresponding to distinct values of y look at the inverse image set and each of these sets even you consider then that forms a partition.

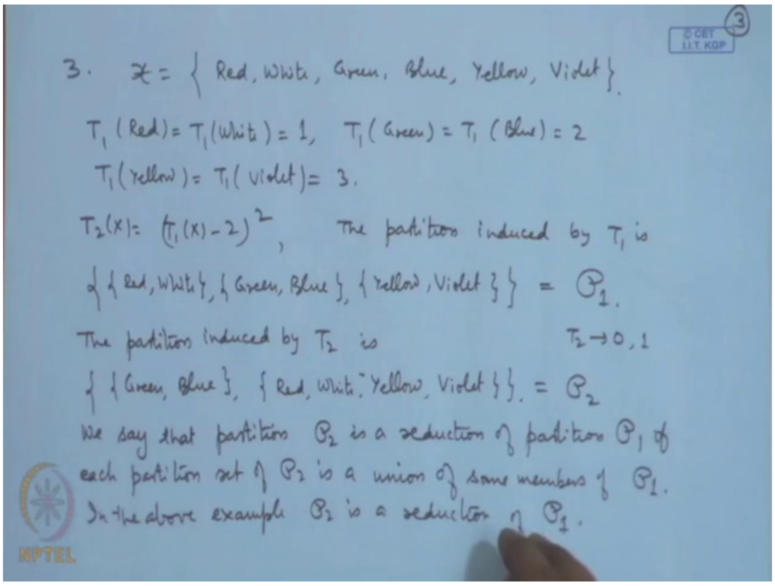
(Refer Slide Time: 04:08)



Let me explain through a example here. Say, I consider X to be a set of real numbers and let us consider the function T x to be say x square. Then what is y then? y is the 0 to infinity then if u consider say T y then it will be equal to minus root y root y for all y greater than 0 and T 0 of course, 0. Then T 0 and T y for y greater than 0 this is a partition of X induced by T x. Let me give another example. Suppose, I consider say x to be the set of say alphabets a b c d and so on up to x y z then I define say T of say e t of o, t of i and say T of a u say is equal to for example, I write 1.

And T of say remaining things remaining characters like b c and so on. T of z is equal to say 2 then if I consider the inverse images of 1 and 2 respectively then what you will get e, i, a, o, u and another set will be remaining b c and so on. Then this is the partition induced by this function T. Now, we can say that a function induces a partition, but given a partition you may not necessarily able to define a function uniquely. However, given a partition we may not define a function uniquely. However if 2 functions say T 1 and T 2 induce the same partition then they must be 1 to 1 functions of each other.

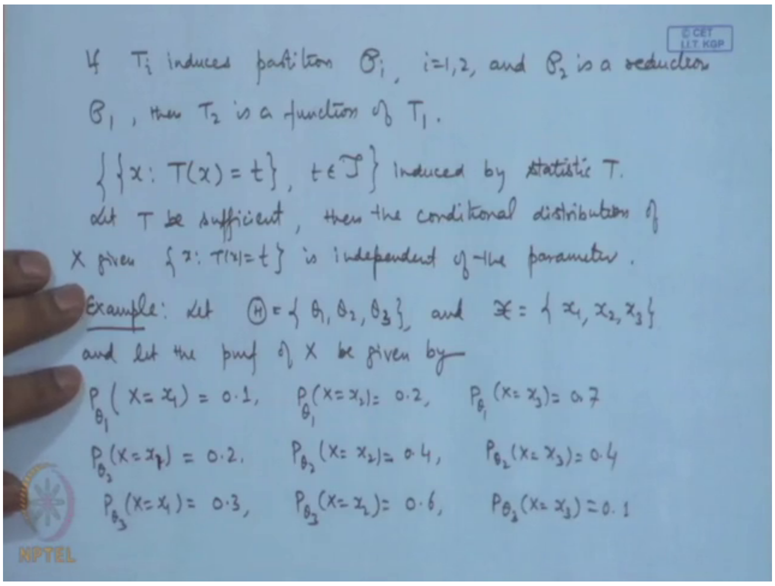
(Refer Slide Time: 08:06)



Let me take another example. Let us consider say X as a colors red, white, green, blue, yellow and say violet. I define a function T 1 of say red and T 1 of white. I assign the value say 1 say T 1 of green and T 1 of blue is equal to say 2, T 1 of yellow and T 1 of violet suppose I define it to be 3 and I define say T 2 X is equal to T 1 x minus 2 whole square. Now, if you define this corresponding T 1 value is equal to 1, T 2 will become 1 and corresponding to T 1 is equal to 3 also it will become 1, corresponding to the value T 1 equal to 2, T 2 will be 0. So, the partition induced by the partition induce by T 1 is red, white in 1 set, green and blue in another set and yellow and violet in another set because the value 1 correspondence to red and white, the value 2 corresponds to green and blue and value 3 corresponds to yellow and violet.

So, let me call this partition p 1. Let us also find out the partition induced by T 2. Now, let us see T 2 is taking values, T 2 can take value 0 that is when T 1 is equal to 2 and when T 1 is equal to 1 or 3, T 2 is taking value 1. So, the value 0 is corresponding to green and blue. So, the partition which is induced by T 2 once it consists of green and blue and 1 is corresponding to that red, white, yellow and violet. Red, white, yellow and violet. Let me call this partition p 2. So, we say that partition p 2 is a reduction of partition p 1 if each partition set of p 2 is a union of some members of p 1. Now, here you observe. The partition sets in p 2 this set is the same as this set and the second set here is a union of 2 sets of p 1. So, in the above example. P 2 is a reduction of p 2 P1 . So, now, from here we can make out if the partition induced by T 2 is a reduction of the partition induced by T 1 then T 2 is a function of T 1.

(Refer Slide Time: 12:23)

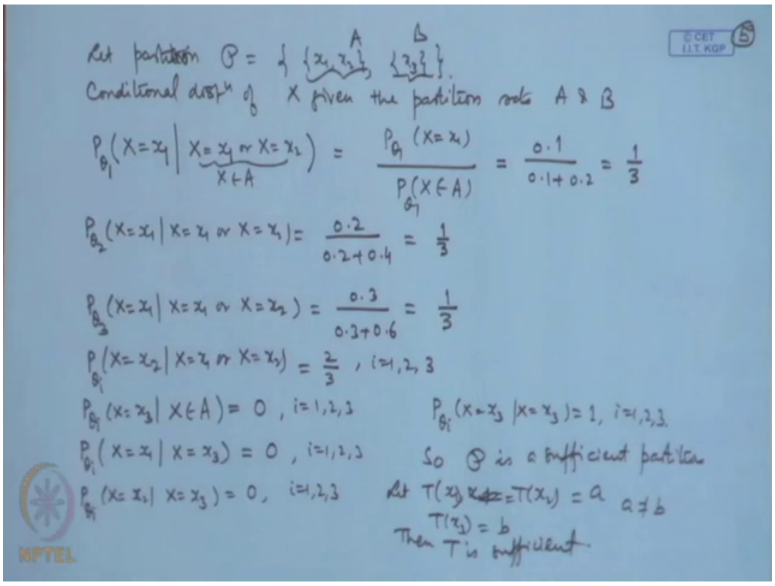


So, we can make this statement that if T i induces partition say p i for i is equal to 1 2 and p 2 is a reduction of p 1 then T 2 is function of T 1. So, now when we talk about a sufficient statistic t then the corresponding partition sets will be T x equals to T where T rallies over the set of values of T this is a partition sets induced by statistic T. Let T be sufficient. Now, the definition of sufficiency says that the conditional distribution of the data that is X 1, X 2 and X n given T must be independent of the parameter. So, we can say that the conditional distribution.

The conditional distribution of x given x such that p x equals to t is independent of the of the parameter. So, let me take 1 example. Let the parameter space consists of 3 points theta 1, theta 2 and theta 3 and let us consider say the variable space consisting of 3 point X 1, X 2, X

3 and let the probability mass function of x be given by probability X is equal to X 1. Now, when theta equals to theta 1 that is equals to 0.1, probability X equals to X 2 is equal to say 0.2, probability X is equal to X 3 is equal to 0.7. So, when the parameter value is theta 1 this is a probability distribution. When theta 2 is the parameter value the probability distribution is given by say 0.2, 0.4 and say 0.4 and when theta 3 is a true parameter value suppose a probability distribution is given by 0.3, 0.6 and 0.1. Let me define a partition here.

(Refer Slide Time: 15:54)



Let us consider partition X 1, X 2 and X 3. Now, consider the conditional distribution of X given the partition sets. Here we have 2 sets. 1 set we call it a and another I call b. So, what is the probability of X is equal to X 1 given say X equal to X 1 or X equals to X 2 that is given that X belongs to a. Now, here it will be dependent upon the. That means, you have to calculate these probability under different configurations. Let me consider say theta equals to theta 1. When I take probability of X equal to X 1 given X belonging to a then we apply the conditional formula this will be equal to probability of X is equal to X 1 divided by probability X belonging to a that is when x equal to X 1 or X equal to X 2.

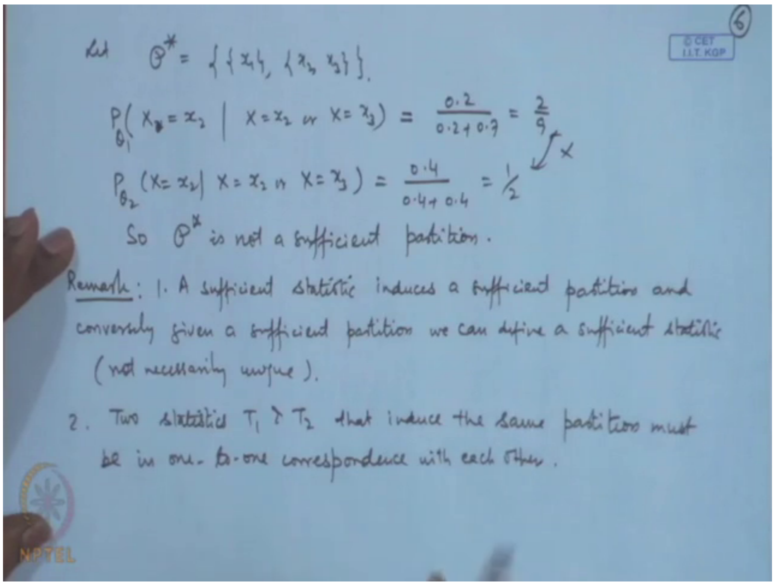
Now, when theta equals to theta 1 probability of X equal to X 1 is 0.1 and probability of X 2 equals to X 2 equals to 0.2. So, these value turns out to be 0.1 divided by 0.1 plus 0.2 that is equal to 1 by 3. Now, if I consider say theta 2 X equals to X 1 given X equals to X 1 or X equals to X 2. Again you notice here when theta equals to theta 2 the corresponding probabilities are 0.2 and 0.4

So, these value will turn out to be 0.2 divided by 0.2 plus 0.4 that is once again 1 by 3 and in a similar way if I calculate when theta is equal to theta 3 this is turning out to be 0.3 by 0.3 plus 0.6 that is again 1 by 3. So, what we have found that the conditional probability of X is equal to X 1 given X belongs to a is free from theta. However, we need to check other configurations also. That means, what is the probability of X is equal to X 2 given X equal to X1 or X equal to X 2.

Once again when theta is equal to theta 1, this value will become 0.2 divided by 0.1 plus 0.2. When theta is equal to theta 2 it will become to 0.4 divided by 0.2 plus 0.4 when theta is equal to theta 3 these probability will equal to 0.6 divided by 0.3 plus 0.6. That means, it is equal to 2 by 3 in all the cases. That means, it is free from value of theta. Similarly, if I consider X is equal to X 3 given x belonging to a then this is going to be 0 because when you take the numerator probability of X equal to X 3 intersection X belongs to a that is going to be 0. So, this is equal to i equal to 1, 2, 3.

Then, similarly if I consider probability of X is equal to X 1 given X belonging to b that is x equals to X 3 then that is also 0 and this will be true for all i is equal to 1, 2, 3. If I consider probability of X is equal to X 2 given X equal to X 3 that is also going to be 0 for i equals to 1, 2, 3. If I consider theta X equals to X 3 given X equal to X 3 that is going to be 1 for i equals to 1, 2, 3. So, we can say that this is a sufficient partition. So, if I consider a function T which is assigning 1 value to say T X 1 and T X 2 as sum value and T X 3 as another value. Let me call it as a and this value as b where a is different from b then p must be sufficient. That means, a function which partitions the variables space into 2 parts, 1 part is consisting of X 1 and X 2 and another 1 is consisting of X 3 then that will be sufficient. Just to tell that this is not a unique way of looking at it for example, in place of X 1, X 2, X 3 where I have clubbed X 1 and X 2 suppose I have clubbed in a different way.

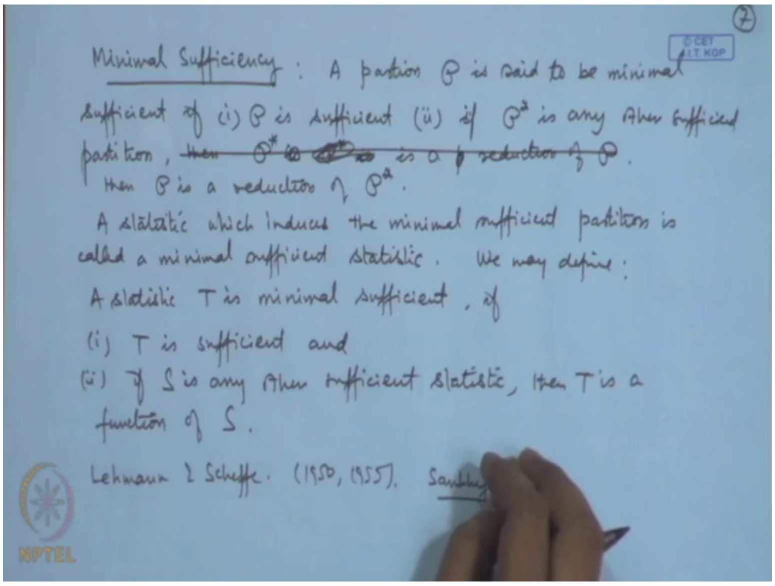
(Refer Slide Time: 21:17)



For example, if I had clubbed let me consider another partition let me call it p star suppose I have put X 1 and say X 2 and X 3. Let us see what is the probability of say X 2 given X equals to X 2 or X equals to X 3. Now, when theta equals to theta 1 this probability will be equal to now X 2 given X 2 equals to X 3. So, this will become equal to 0.2 divided by 0.2 plus 0.7, 0.2 divided by 0.2 plus 0.7 that is 2 by 9. Now if I consider theta 2 same probability then for theta 2 these values are 0.4 and 0.4. So, this value will be turning out to be half, this is not same as this.

So, p star is not a sufficient partition. Consequently any statistic which will induce these partition will not be sufficient. Let me a sufficient statistic induces a sufficient partition and conversely given a sufficient partition we can define a sufficient statistic. Of course, this is not necessarily unique and also 2 statistic T 1 and T 2 that induce the same partition must be in 1 to 1 correspondence with each other.

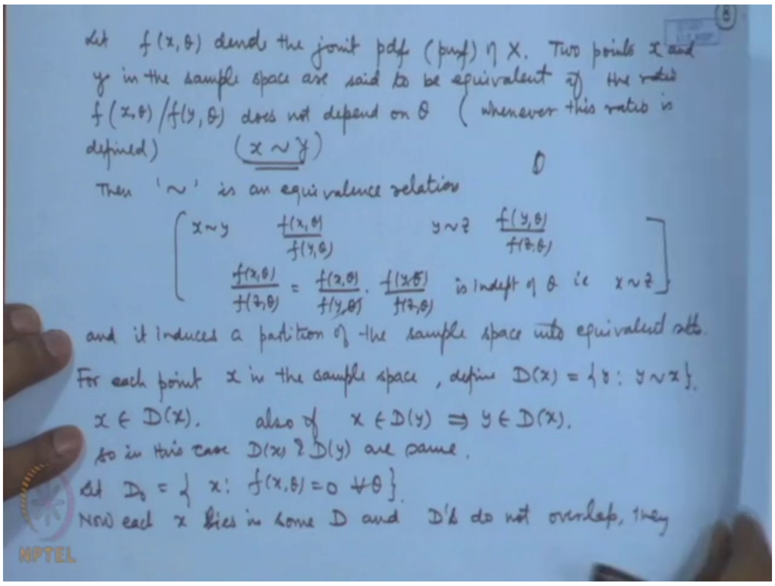
(Refer Slide Time: 24:33)



Now, we define the concept of minimal sufficient partition and minimal sufficient statistic. 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. I am sorry this written wrongly if p star is any other sufficient partition then p is a reduction of p star. So, let me explain. You 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 a minimal sufficient that is the maximal reduction of the data. Now, the question is how to determine a minimal sufficient statistic or a minimal sufficient partition in a given problem. This problem is settled by Lehmann and Schaffer in 1950 and 1955 in papers in Sankhya. We consider the case when the distribution is either discreet or continuous.

(Refer Slide Time: 28:33)



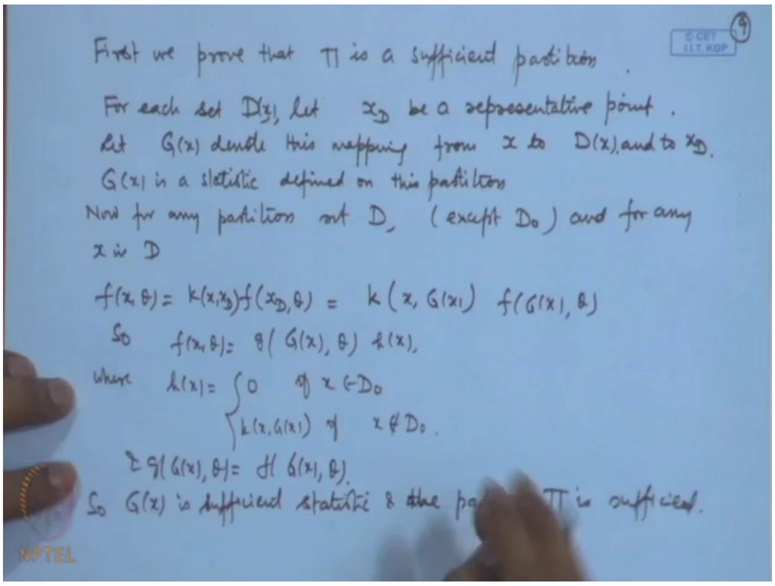
So, let us consider f X theta. Let f X theta denote the joint probability density function or probability mass function of x. That means, we have observation of X 1, X 2, X n which we are calling as X here. Now, 2 points X and y in the sample space or said to be equivalent if the ratio if X theta by f y theta does not depend on theta. Of course, when we write the ratio of the densities of 2 different variable points then there is a possibility that either the numerator or denominator may be 0 or both maybe 0. So, in that case we qualify this statement by saying whenever this ratio is defined. So, this the we say that X and y are equivalent and we use a 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 parameters 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 2 terms. So, that is also free from. So, this is independent of theta that is we can say X is equivalent to j. 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 1 set in this partition class then within that class all the points will be equivalent and if I take 2 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 will 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 are the probability mass function we put it in another set. Let d naught be the set of all those points for which f X theta equals to 0 for all theta. 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.

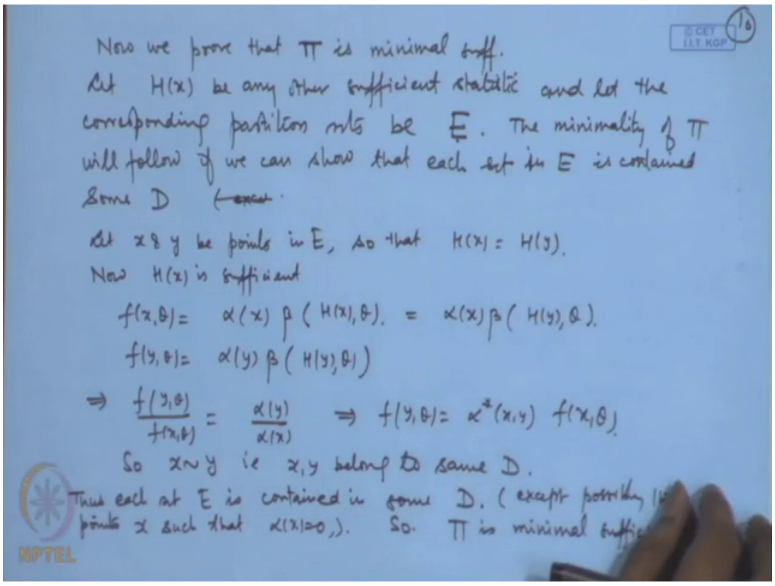
(Refer Slide Time: 34:32)



First we proof that pi is a sufficient partition. Now, let us consider for each set d. Let X D be a representative point. 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 1 point X D there. So, G X is a statistic define on this partition. Now for any partition the set d of course, I am not considering D naught 0 and for any X in d let us write f X theta.

Now, X belongs to this and 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 into 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 does not belong to d 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: 38:31)

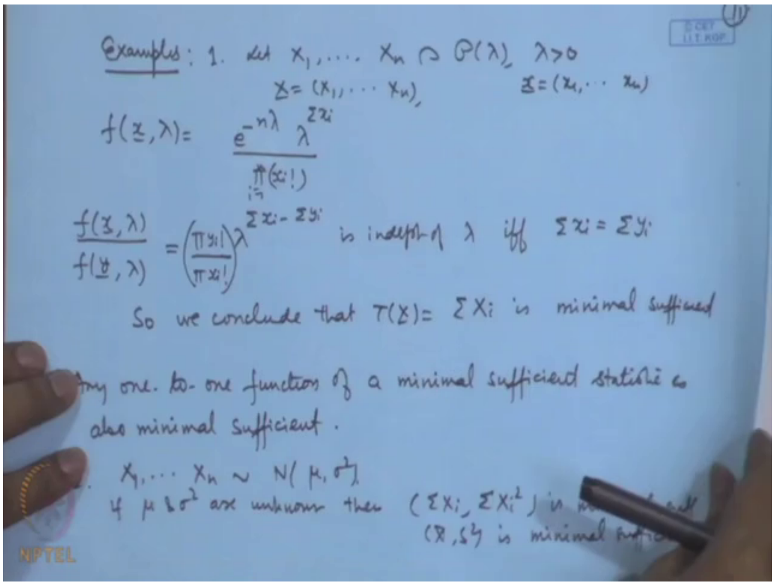


Now, let us consider. Now, we prove that pi is minimal sufficient. For that let us consider another say H X. Let H X be any other sufficient statistic and let the corresponding partition sets be, let me call them e the partition sets induced by pi d and the partition sets induced by E let me call by h let may be 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 the 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 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 statistic. 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 partition that will induce this condition and the corresponding sufficient corresponding statistic then we will find out that will be minimal sufficient.

(Refer Slide Time: 43:28)



So, let me explain through some examples. Let us consider the cases of standard estimations suppose I consider Poisson lambda estimation and we denote by X the X 1, X 2, X n by small x we denote the points. So, consider the f x lambda here that is e 2 the power minus m 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 m lambda will cancel out, 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, by the previous results that we have approved of Lehmann and Scheffe we conclude that T X is equal to sigma X i is minimal sufficient. Of course, you can say that any 1 to 1 function of a minimal sufficient statistic is also minimal sufficient. Let me just take up the cases of sufficient statistic that we worked out in the previous classes. We had seen binomial distribution, nomial distribution, exponential distribution etcetera. Let us look at each of those cases and see what were the sufficient statistic. Consider this case X 1, X 2, X n the random sample from normal mu sigma square and sigma square is known.

Now, in this case as joint distribution that we wrote was of the form 1 by root to the power n e to the power sigma X i square by 2 E to the power minus mu square plus 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 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 sufficiency statistics. So, like that if we consider minimal problems like in the second case we have taken mu naught is known and in this case we figure out that sigma X i minus mu not whole square is not 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.

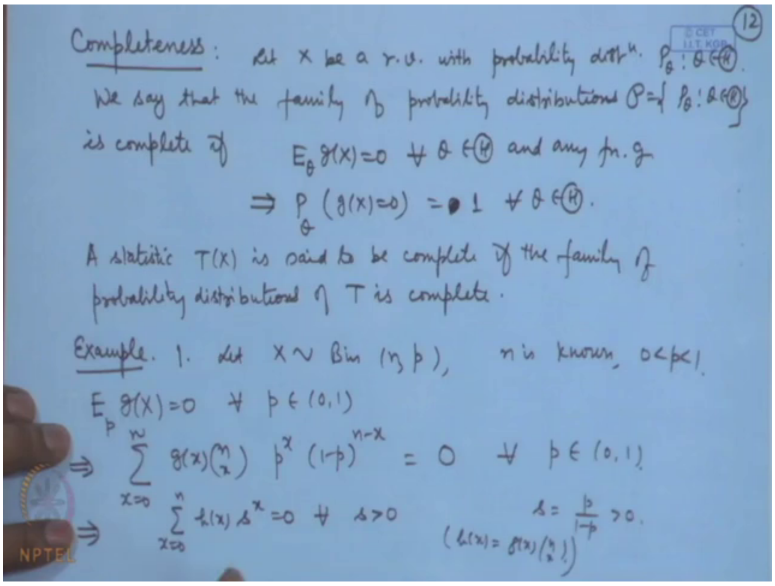
So, in most of the problems where we have applied factorisation theoram we actually have a factorization. So, if we write down the ratio then the term which is consisting of parameter theta there then it is related to G of T x theta divided by G of T y theta. So, this ratio if you consider and obtain the condition when it is going to be free from the parameter that will give the minimal sufficient statistic

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 X square is minimal sufficient and we cananswer various other questions. Let me just tell few of this here. Let us consider say exponential distribution with parameter lambda. Here if I write down the ratio we will get sigma X i as a minimal sufficient.

If we consider exponential distriburtion with location parameter then X 1 will be turning out to be minimal sufficient. If we consider say 2 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 or reduce to the ordered statistic that will be minimal sufficient.

If we consider uniform distribution on the interval 0 to theta then Xn 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: 50:19)

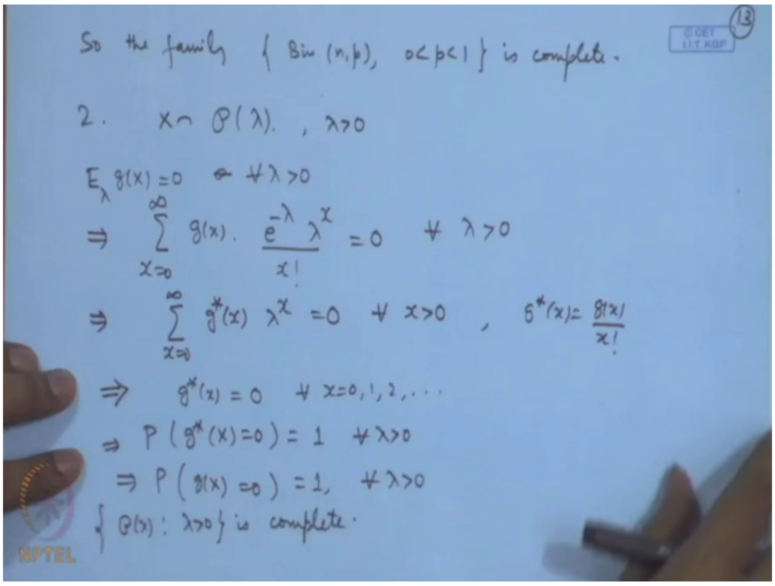


Let me introduce another concept that is completeness. Let X be a random variable with probability distribution p theta, theta belonging to theta. So, we say that the family of probability distributions p that is equal to p theta, theta belonging to theta is complete if expectation theta G X is equal to 0 for all theta belonging to theta and any function G implies that probabilities 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 equvalent to G X n c 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 function G X into n c X.

So, now, if you see this left hand side this is a binomial of degree n in s and I am saying it is vanishing identically over an interval this implies that H X must be 0 for all X. Now, here for all X means because X can take value 0 1 to n this means that probability that H X is equal to 0 is 1 for all p. So, family of binomial distributions is complete.

(Refer Slide Time: 54:19)

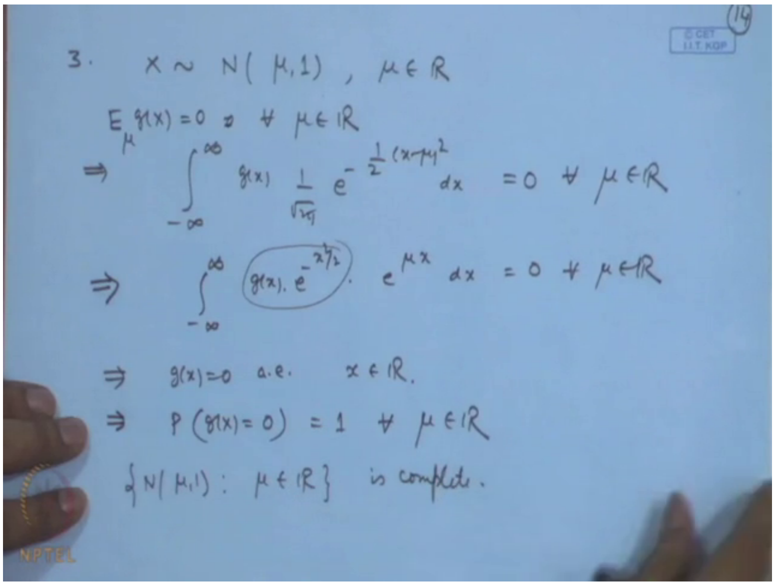


So, the family of binomial distributions n p where p lies within 0 to 1 is complete or we can say here X is a complete statistic. Let us take another example. So, x followes poisson lambda then lambda is a positive parameter here. Let us write down the statement expectation of G X is equal to 0 for all 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 at 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 on both the sides. This statement becomes equivalent to say G star X into lambda to the power x where g star is nothing, but G x by x factorial. Once, again if you have noticed 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 vanises identically over an interval all the coefficients must vanish. So that means, this is equal to 0 for all X is equal to 0, 1, 2 and so on. Therfore 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.

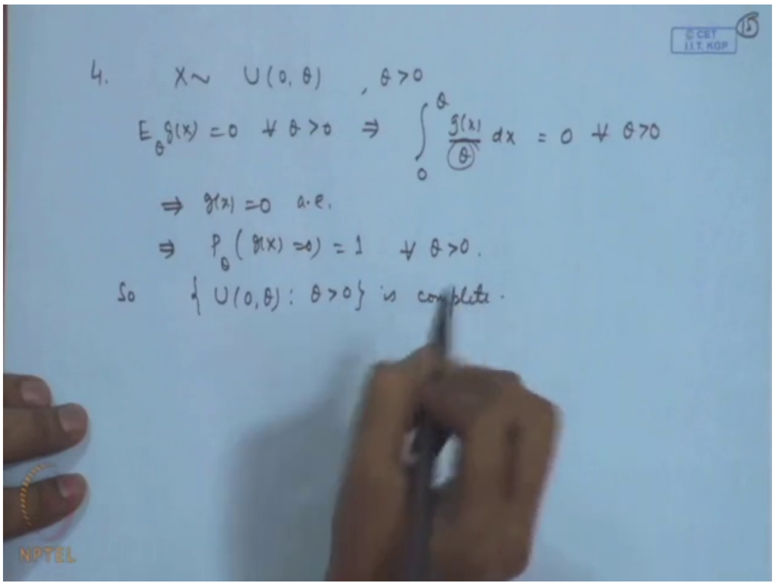
(Refer Slide Time: 56:26)



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 mu. Now, this is all 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 2 to the power minus X square by 2 e to the power mu x D X is equal to 0 for all mu belonging to R.

Now, this is nothing, but the bilateral or biirriate 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. So, the family of the normal distributions is complete family.

(Refer Slide Time: 58:01)



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 D X 0 to theta is equal to 0. Now, this term I can adjust here. So, what we are saying is the integral of G X is 0 for all values over all the intervals of the form 0 to theta. Therfore, we can using the Libeg Integration theory we can say that the function G X itself is 0 almost everywhere. That means, probability that G X is equal to 0 must b 1 for all theta. So, the family of uniform distributions is also complete.

So, what in the next lecture I will give a general frame work for the completeness. We will also define bounded completeness and the consiquence of the sufficiency completeness is that we can easily derive uniformly minimum variance unbiased estimators. So, we will give these applications in the next lecture.