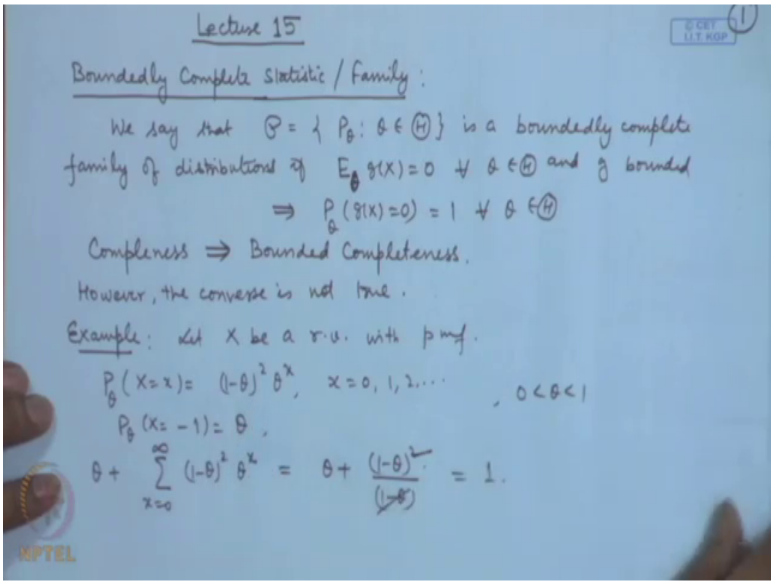
Statistical Inference
Prof. Somesh Kumar
Department of Mathematics
Indian Institute of Technology, Kharagpur

Lecture – 29
UMVU Estimation, Ancillarity – I

In the last lecture, I introduced the concept of minimal sufficiency and completeness of certain statistics or again these are also the properties of the family of distributions.

(Refer Slide Time: 00:40)



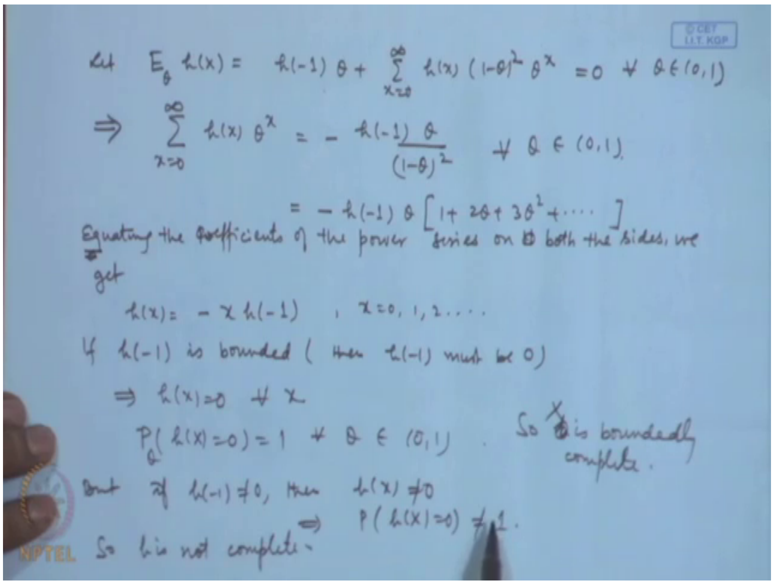
Now, before we proceed further I will define a related concept that is called boundedly complete, boundedly complete statistics or boundedly complete family of distributions. So, we say that P is equal to P theta is a boundedly complete family of distributions. If expectation of g X is equal to 0 for all theta and g bounded implies that probability of g X is equal to 0 is 1 for all theta.

So, the difference from the definition of completeness is that there we wrote any function g. So, expectation g X equal to 0 for all theta and any function g if that implies that the probability that the function is 0 with probability 1 then it was complete. If I impose the condition that g is bounded, then it will imply that probability of g X equal to 0 is 1 then it will be called a boundedly complete family of distributions. So, we can say that completeness implies bounded completeness. However, the converse is not true I will give an example here.

Let X be a random variable with probability mass function given by P theta X is equal to X is equal to 1 minus theta square theta to the power x for x equal to 0 1 2 and so on. And P theta X is equal to minus 1 is equal to theta here theta is between 0 to 1. Now, you can easily see that theta plus sigma 1 minus theta square theta to the power x, x equal to 0 to infinity that is equal to theta plus 1 minus theta square divided by 1 minus theta, because this is infinite geometric series with common ratio theta. So, this cancels out you get to 1.

So, this is a proper probability distribution. You can say it is a shifted geometric kind of distribution. Let us show whether it is complete or not ok.

(Refer Slide Time: 04:07)



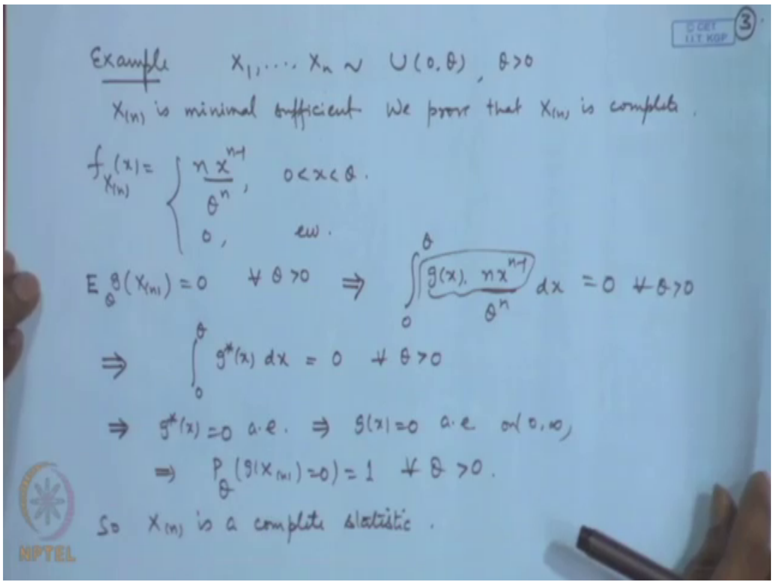
So, consider a function h X then its expectation can be written as h of minus 1 into theta plus sigma h X into 1 minus theta square theta to the power x, x equal to 0 to infinity. Now, suppose we equated to 0 for all theta in the interval 0 to 1.Now this term I take to the right hand side and then we divide by 1 minus theta square. So, it is reducing to h X in to theta to the power x, it is equal to minus h of minus 1 theta divided by 1 minus theta square this is for all theta then the interval 0 to 1.

Further, this 1 minus theta square in the denominator. So, if I bring it to the numerator it becomes 1 minus theta to the power minus 2 and I can expand because theta is in the interval 0 to 1. So, this we can write as minus h of minus 1 into theta and this expansion can be written as 1 plus 2 theta plus 3 theta square and so on. Now, if I consider these 2 terms, the left hand side is a power series in theta and the right hand side is also a power series in theta. So, if I create the terms we get equating the coefficients of the power series on both the sides, we get h X is equal to minus x into h of minus 1 for x equal to 0 1 2 1 so on.

Now if h of minus 1 is bounded then h of minus 1 must be 0 because if h of minus 1 is not 0 then this function is unbounded, because it will be X into some constant. So, for boundedness the only possibility is that h of minus 1 is 0 which will imply h of X is equal to 0 for all x. That means, probability of h X is equal to 0 will be 1 for all theta and the interval 0 to 1. So, h is boundedly complete not X is boundedly complete. But if h of minus 1 is not 0 then h of X is also not 0 this implies probability that h X is 0 cannot be 1. So, h is not complete because expectation of h X is 0, but h X will not be 0 with probability 1.

So, this is an example of a boundedly complete family of distributions which is not complete.

(Refer Slide Time: 07:52)



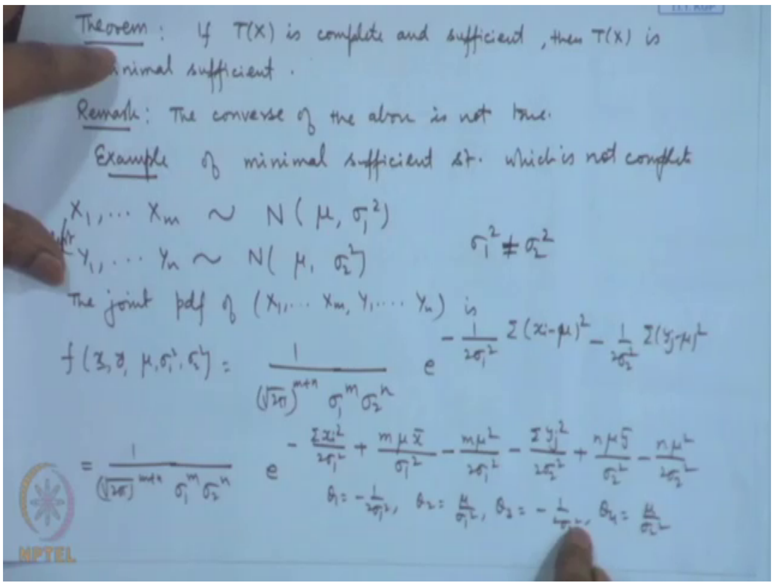
Now, there are relationships between sufficiency and completeness also there is a general way of determining complete statistics. For example, if the distributions are in the exponential family I have already given the example of binomial distribution Poisson distribution. So, in the Poisson distribution family is complete. If I consider sufficient statistics or minimal sufficient statistics, that is running out to be sigma x i which is again having Poisson distribution with parameter n lambda. So, if Poisson lambda is complete Poisson n lambda is also complete.

So, sigma xi is complete. So, we can conclude that in most of the standard examples that we have discussed the corresponding sufficient or minimal sufficient statistics will also be complete. Let me just take the example of non regular family say; let me consider say X 1 X 2 X n following uniform 0 theta distribution then X n is minimal sufficient we prove that X n is complete. Let us consider the distribution of X n that is n x to the power n minus 1 by theta to the power n 0 less than x less than theta it is 0 elsewhere.

So, if I consider expectation of say g of X n is equal to 0 for all theta then this statement is equivalent to g x n x to the power n minus 1 by theta to the power n dx from 0 to theta is equal to 0 for all theta. Now, this is equivalent to saying a function of x over all the intervals 0 to theta is integrated to 0. Again by the Lebesgue integration theory it implies that g star must be 0 almost everywhere this g star function I have taken to be this. So, this implies that g X is equal to 0 almost everywhere on 0 to infinity. This implies that probability that g X n is equal to 0 is 1 for all theta. So, X n is a complete a statistic.

So, there is a relation between minimal sufficiency and complete sufficiency. In fact, we have the following theorem.

(Refer Slide Time: 11:04)

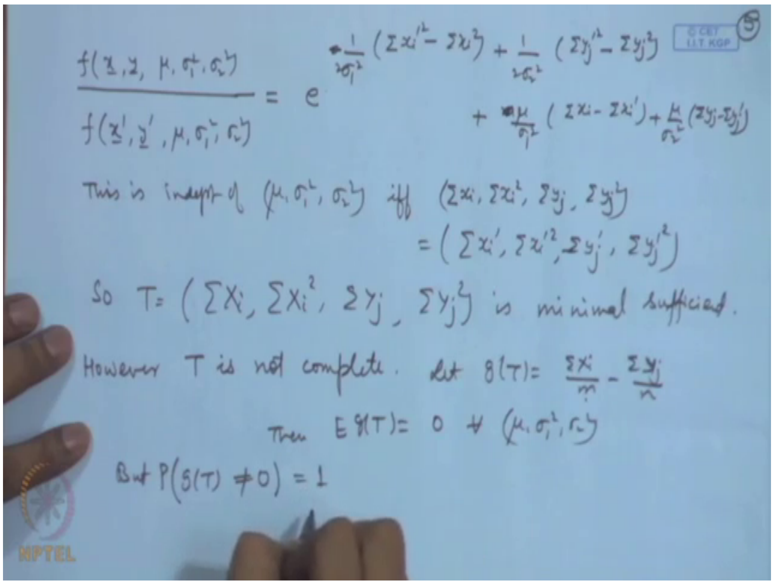


If T X is complete and sufficient then T X is minimal sufficient. However, the converse of the above statement is not true that is we may have an example of say minimal sufficient statistic which is not complete. Let us take say X 1 X 2 X m a random sample from normal with mean mu and variance sigma 1 square and Y 1 Y 2 Y n. This is another independent sample from normal with mean mu and variance sigma 2 square; here sigma 1 square and sigma 2 square are different.

Let us consider the joint distribution of X 1 X 2 X m and Y 1 Y 2 Y n. The joint pdf of X 1 X 2 X n Y 1 Y 2 Y n that is equal to 1 by root 2 pi to the power m plus n sigma 1 to the power n sigma 2 to the power n e to the power minus 1 by 2 sigma 1 square sigma x i minus nu square minus 1 by 2 sigma 2 square sigma yj minus mu square. This we can simplify as 1 by root 2 pi to the power m plus n sigma 1 to the power n sigma 2 to the power n e to the power minus sigma xi square by 2 sigma 1 square plus m mu X bar by sigma 1 square minus mu square by 2 sigma 1 square minus sigma yj square by 2 sigma 2 square plus n mu y bar by sigma 2 square minus n nu square by 2 sigma 2 square.

So, if we apply the ratio by writing down this joint pdf at 2 points x, y and say x prime, y prime then these terms will get cancelled out and we will be left with sigma x i square minus sigma x i prime square in 2 parametric function X bar minus y bar into the parametric function, X bar minus X bar prime y bar minus y bar prime and sigma yj square minus sigma yj prime square.

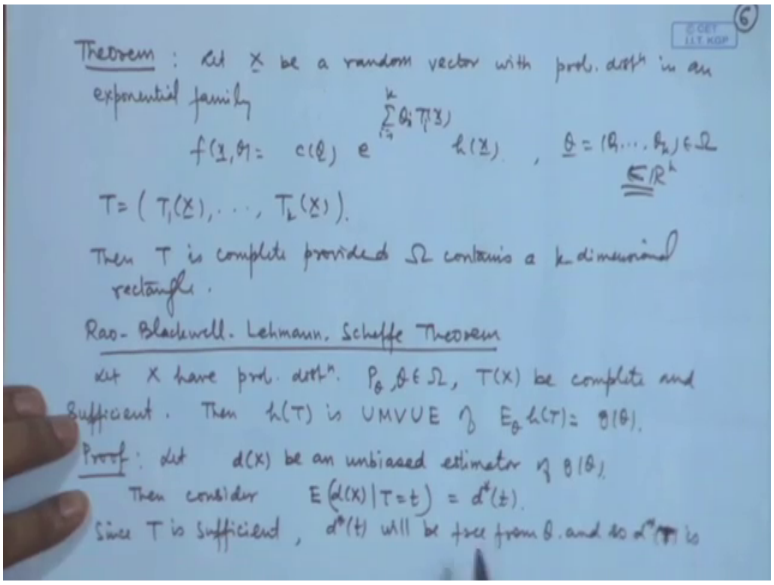
(Refer Slide Time: 15:03)



So, if we write down this function here say f x y mu sigma 1 square sigma 2 square divided by say fx prime y prime mu sigma 1 square sigma 2 square then that is equal to e to the power 1 by 2 sigma 1 square sigma x i prime square minus sigma x i square plus 1 by 2 sigma 2 square sigma yj prime square minus sigma yj square then plus m mu or mu by sigma 1 square sigma x i minus sigma x i prime plus mu by sigma 2 square sigma yj minus sigma yj prime. So, this is independent of mu sigma 1 square and sigma 2 square if and only if we have sigma x i sigma x i square sigma y i sigma y i square is equal to sigma x i prime sigma x i prime square sigma yi yj prime and sigma yj prime square.

So, T is equal to sigma x i sigma x i square sigma y j sigma yj square is minimal sufficient. However, T is not complete. Let us consider g T as a sigma x i by m minus sigma yj by n. Then expectation of g T is equal to 0 for all mu sigma 1 square sigma 2 square because expectation of x i and expectation of yj is mu. So, it is mu by m minus n mu by n. But g T is not 0; actually probability that g T is not 0 is 1, probability that g T is equal to 0 is actually 0. So, T is not complete. So, this is an example of a minimal sufficient statistic which is not complete. To determine complete statistics in general settings or to prove the completeness in general settings of exponential family one only needs to check the kind of parameter space that we are having.

(Refer Slide Time: 18:14)



So, we have the following general theorem which I will state without proof for the proof one can look at the book of Layman testing of hypothesis book. So, let X be a random vector with probability distribution in an exponential family. Say we write it in the form c theta e to the power sigma theta T x into hx. So, here c theta is a function of parameter h X is function free from parameter and parameter is occurring in the exponent, here theta is equal to theta 1 theta 2 theta k that is it is belonging to R k. Let me say it belongs to omega and omega is a subset of R k.

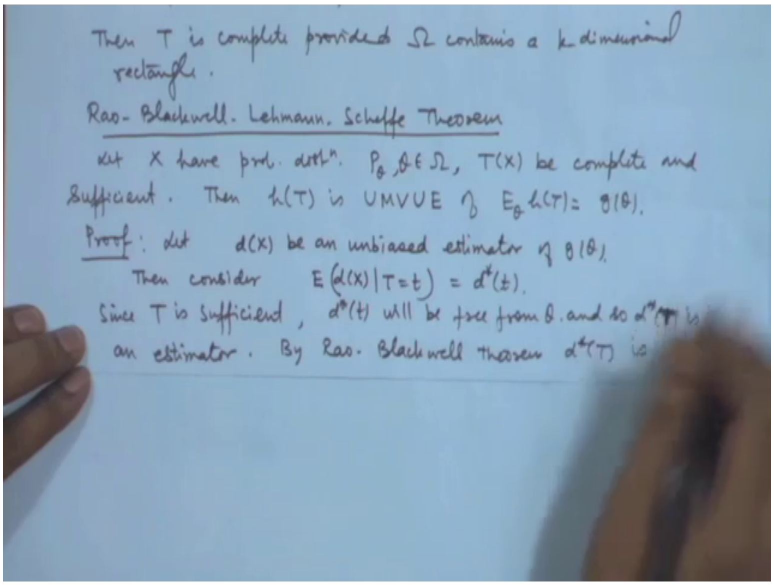
Let us write T as T 1 X and so on T k X. Then T is complete provided omega contains a k dimensional rectangle. If you look at the previous example here this is actually a 3 parameter distribution here. Here what we are getting is 1 by 2 sigma 1 square or you can say 1 by sigma 1 square mu by sigma 1 square then 1 by sigma 2 square and mu by sigma 2 square. However, they are not independent. Actually the parameter is 4 dimensional. If we write theta 1 is equal to say minus 1 by 2 sigma 1 square theta 2 is equal to mu by sigma 1 square theta 3 as equal to say minus 1 by 2 sigma 2 square and theta 4 is equal to say n say mu by sigma 2 square then this is a 4 dimensional parameter.

But there is dependency upon that for example, given theta 1 theta 2 and theta 3 we can determine theta four. So, the parameter space does not contain a 4 dimensional rectangle and that is why we could actually show that this is not complete T was not complete here. We have seen the application of sufficiency in estimation problems. We saw that if we have an unbiased estimator we can certainly improve upon it by conditioning upon the sufficient statistics, the result was known as the Rao Blackwell theorem. Now, if we couple this concept with the completeness we get a stronger result. In fact, we can reduce the problem to determination of the uniformly minimum variance unbiased estimator; the resulting result which is actually associated in the name of Lehmann Scheffe.

So, I will come couple the 2 results Rao Blackwell and Lehmann Scheffe and we call it Rao Blackwell Lehmann Scheffe theorem. Let X have probability distribution P theta; theta belonging to say omega and T X be complete and sufficient, then h T is UMVUE of expectation of h T; let us call it say g of theta. That means, for any estimable unbiased estimate function g theta if I have an unbiased estimator which is dependent upon the complete sufficient statistic then that will be actually UMVUE. Let us look at the proof of this.

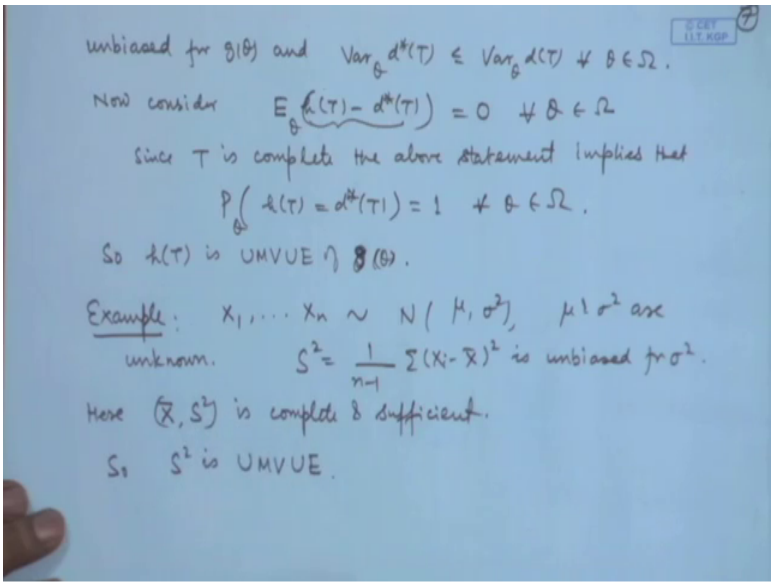
Let say d X be an unbiased estimator of g theta. Then you will have consider expectation of d X given T; let me denote it by say d star. Since, T is sufficient d star t will be free from theta, because the conditional distribution of X given T is independent of theta. Therefore, this expectation will not contain any term of theta and we can call it d star t and so, d star t is d star T suppose I had capital T here, this is an estimator.

(Refer Slide Time: 24:17)



Now, we have already seen that by Rao Blackwell theorem d star T is also unbiased for g theta and variance of d star was less than or equal to the variance of d T.

(Refer Slide Time: 24:38)



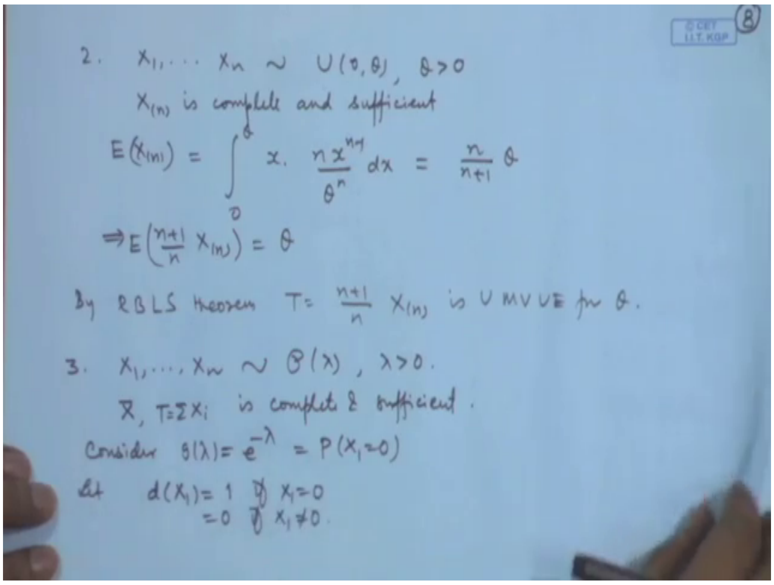
Now, consider expectation of h T minus d star T then that is 0 because both of these are unbiased for g theta. Now, this is a function of T and T is complete. Since, T is complete the above statement implies that h T must be equal to d star T with probability 1. Essentially it proves that h T is a unique unbiased estimator of g theta. So, h T is UMVUE.

Actually g star d star is also UMVUE, but these 2 we differ only on a set of measure 0. Now, this result is extremely useful for finding out the UMVUEs. We have seen actually in the earlier method of lower bounds that many times whatever best unbiased estimator we are able to think of the variance of that is not attaining the lower bound, whether we are considering the Fisher Rao Cramer lower bound, Bhattacharya lower bound or Chapmans Robbins Keifer low lower bound etcetera. In many of the cases we saw that the variance of the unbiased estimator was bigger than the lower bound the corresponding lower bound. However, this method when we are considering a function of complete and sufficient statistic, it immediately proves that the corresponding estimator will become uniformly minimum variance unbiased estimator.

Essentially what it is doing? It will actually show that the corresponding unbiased estimator is actually the only unbiased estimator available except of course, on a set of probability 0. So, since it is unique certainly it is UMVUE. So, if we go back to various problems where the lower bound was not attained, for example, if you consider normal mu sigma square where mu is unknown and we were considering the estimation of sigma square. So, let us consider say X 1 X 2 X n follows normal mu sigma square mu and sigma square are unknown and we have this S square as 1 by n minus 1 sigma x i minus X bar whole square this is unbiased for sigma square.

Now, in this problem X bar and S square is complete and sufficient. So, S square is UMVUE. We had noticed here that in this particular case, the lower bound that was attained by the method of Fisher Rao Cramer, it was lower than the variance of S square. The variance of S square was 2 sigma to the power 4 by n minus 1 and lower bound was 2 sigma to the power 4 by n, but here in this method UMVUE proving is easy, because we are just looking at the expectation of S square, since it is equivalent it is a function of the complete sufficient statistics. So, it becomes UMVUE.

(Refer Slide Time: 28:59)



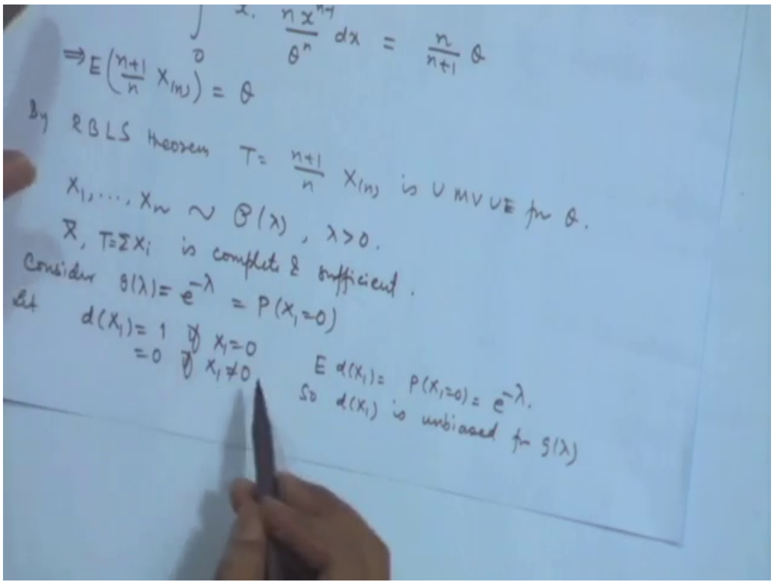
Let us take other related examples also. X 1 X 2 X n following uniform 0 theta. Here we have shown that X n is complete and sufficient.

Now, if we look at expectation of X n that is x into n x to the power n minus 1 by 2 theta to the power n x then this is equal to n by n plus 1 theta. That means n plus 1 by n X n is unbiased for theta. Now, this is a function of complete sufficient statistics. So, by Rao Blackwell Lehmann Scheffe theorem we conclude that n plus 1 by n X n this is UMVUE for theta.

We have also seen the standard distributions like Poisson distribution; where for lambda we are able to derive the UMVUE, but for lambda square we are not able to derive or if I can see that e to the power minus lambda. Then we were not able to derive the UMVUE, but using this method we can derive. Let me explain this here.

Let us consider say X 1 X 2 X n following Poisson lambda distribution lambda positive. Now, here X bar or you can say sigma X i this is complete and sufficient. Suppose I am considering g lambda is equal to e to the power minus lambda which I had explained actually this is probability of X 1 is equal to 0 that is the proportion of 0 occurrences in a given problem. Let us define say d X 1 is equal to 1 if X 1 is 0 and it is equal to 0 if X 1 is not equal to 0.

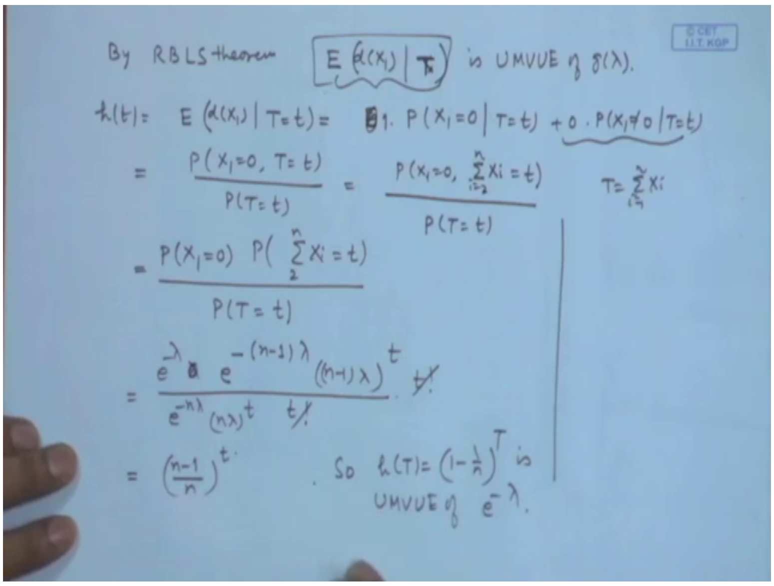
(Refer Slide Time: 31:38)



Then, if I consider here expectation of d X 1 then that is equal to probability of X 1 is equal to 0 that is equal to e to the power minus lambda. So, d X 1 is unbiased for g lambda.

However, this is not UMVUE because this is not a function of the complete sufficient statistic.

(Refer Slide Time: 32:07)



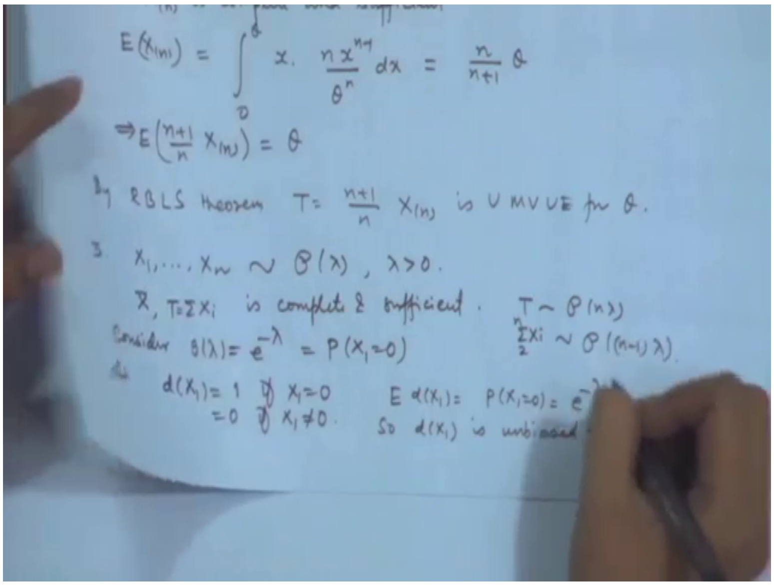
So, if I apply the Rao Blackwell Lehmann Scheffe theorem, if I consider Rao Blackwell Lehmann Scheffe theorem; if I consider expectation of d X 1 given T T sigma X i or X bar we can write then this is UMVUE of g lambda. So, the only thing remaining is that determination of this function, we can determine it easily.

Let us denote it by h T, expectation of say d X 1 given T is equal to small t. Then this is equal to expectation of now, d X 1 takes only 2 values 1 and 0. So, it is equal to probability of X 1 is equal to 0 given T is equal to t, because d X 1 is equal to 0 then probability of X 1 is not equal to 0, but when u 0 multiplied then that value will not matter; X 1 not equal to 0 given T is equal to t. So, this term is vanishing.

So, we need to only determine this conditional probability that is probability X 1 is equal to 0; T is equal to t divided by probability T is equal to t that is equal to probability X 1 is equal to 0. Now, this T is nothing but sigma X I; i is equal to 1 to n. If I say X 1 is equal to 0 then we can say sigma of X i from 2 to n is also equal to t.

Now, here you notice that the sum of independent Poisson's is Poisson.

(Refer Slide Time: 33:58)



So, the distribution of T will be Poisson n lambda and distribution of sigma X I, i is equal to 2 to n that will be Poisson n minus 1 lambda. So, if we use this here X 1 and sigma x i from 2 to n these will be independent. So, this can be written as the product of this probability. So, it becomes probability of X 1 equal to 0 in to probability of sigma X i from 2 to n is equal to t divided by probability T is equal to t. So, that is equal to e to the power minus lambda lambda to the power 0; so, that term will not come. Then this is following Poisson n minus 1 lambda. So, it is becoming e to the power minus n minus 1 lambda n minus 1 lambda to the power t divided by t factorial and then probability T is equal to t that is e to the power minus n lambda n lambda to the power t into t factorial.

So, these terms get cancelled out and we are left with n minus 1 by n t. So, h T is equal to 1 minus 1 by n to the power of T; this is UMVUE of e to the power minus lambda. So, this Rao Blackwell Lehmann Scheffe theorem is extremely useful to determine the UMVUE for various functions where the method of lower bounds is not applicable. I will be introducing in the next classes.