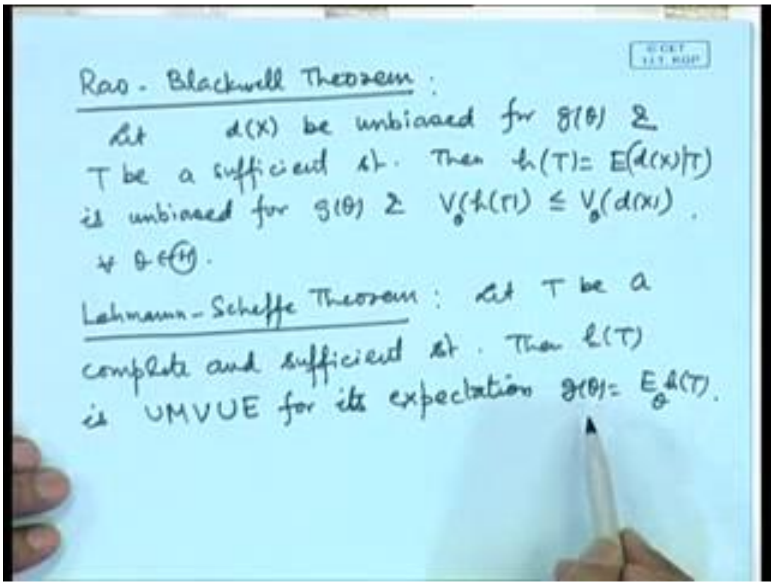
Probability and Statistics
Prof. Somesh Kumar
Department of Mathematics
Indian Institute of Technology, Kharagpur

Lecture - 60
Rao-Blackwell Theorem and Its Applications

We have actually an important result which is known as Rao-Blackwell, and Rao-Blackwell Lehmann-Scheffe Theorem; let me give that result. That will help us in obtaining the uniformly minimum variance and unbiased estimators.

(Refer Slide Time: 00:38)



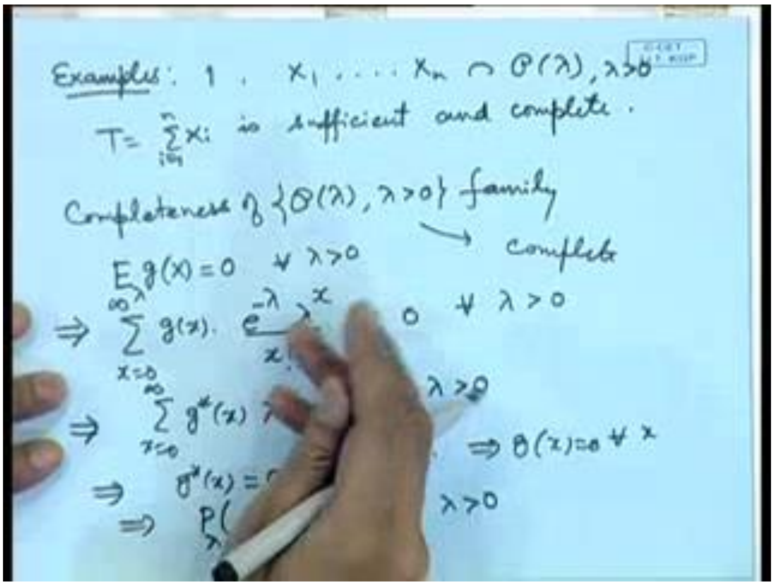
So, the Rao-Blackwell theorem says that- let d x be unbiased for g theta and T be a sufficient statistic, then let us define say h T as expectation of d x given T conditional expectation, then this is unbiased for g theta and variance of h T is less than or equal to variance of d x for all theta.

Now this means that, if there is a estimator which is not dependent upon the sufficient statistic I can transform it by conditioning upon the sufficient statistic and get something which is better. That means, it is always advisable to start with functions of sufficient statistic for making estimation. Now further a strengthening of this theorem is done if we use the concept of completeness also. Let T be a complete and sufficient statistic then h T is UMVUE for its expectation; that is g theta is equal to expectation of h T. Now this is extremely significant result, it means that whenever I have a complete sufficient statistic and I have to find out UMVUE of any parametric function then I consider an appropriate parametric function which will be based on the complete sufficient statistic and it will be unbiased; that is all

So, that will become UMVUE automatically, we do not have to do any further proof that we have to compare its variance with any other unbiased estimator etcetera; it will be automatically. The reason is that the property of the completeness that the only unbiased estimators of 0 is 0 itself; that means, for any given parametric function based on the complete sufficient statistic you cannot have two unbiased estimators. If they are two then they will be same with probability 1. And if there is any other estimator which is not dependent upon the complete sufficient statistic, then that can again be improved by taking conditioning so you will get the same one.

Therefore, it is advisable to restrict attention to functions of complete sufficient statistic. Now this gives us a very convenient tool for deriving unbiased estimators in various problems. So, let us go back to the examples which we have done earlier.

(Refer Slide Time: 04:13)

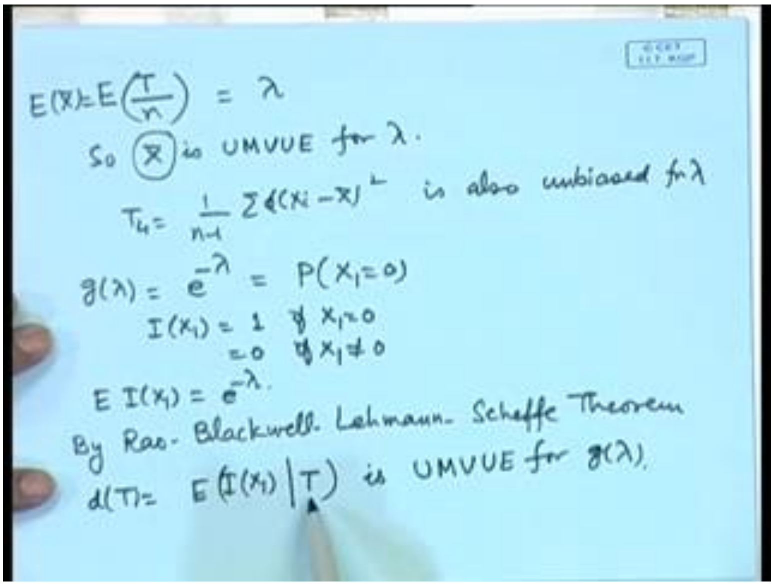


So, one of the first problems we considered was estimation of the parameter of a Poisson distribution. So now the question arises about the completeness and sufficiency. Here T is sufficient, what about completeness? The distribution of T is Poisson n lambda. So, if we can prove that the family of Poisson distribution is complete then T will also become a complete statistic. So, in place of Poisson n lambda we can prove the completeness of Poisson lambda family.

Let us write down expectation of say g x is equal to 0 this is statement is equivalent to g x e to the power minus lambda lambda to the power x by x factorial. Since lambda is positive this we can multiply on both the sides by e to the power lambda and g x by x factorial term I can combine as some g star x. Now the left hand side is a power series in lambda and we are saying that it is identically 0 on the positive half of the real line; that is possible only if g star itself is 0, that means all the coefficients must be 0 which is equivalent to saying that g x is 0 for all x, which implies that probability that g x is 0 is 1 for all lambda. That means this family is complete.

So, this means that sigma xi is sufficient as well as complete. Now that makes our problem extremely simple. Now based on T whatever estimator we take if we take its expectation then that estimator will become UMVUE for that.

(Refer Slide Time: 06:55)

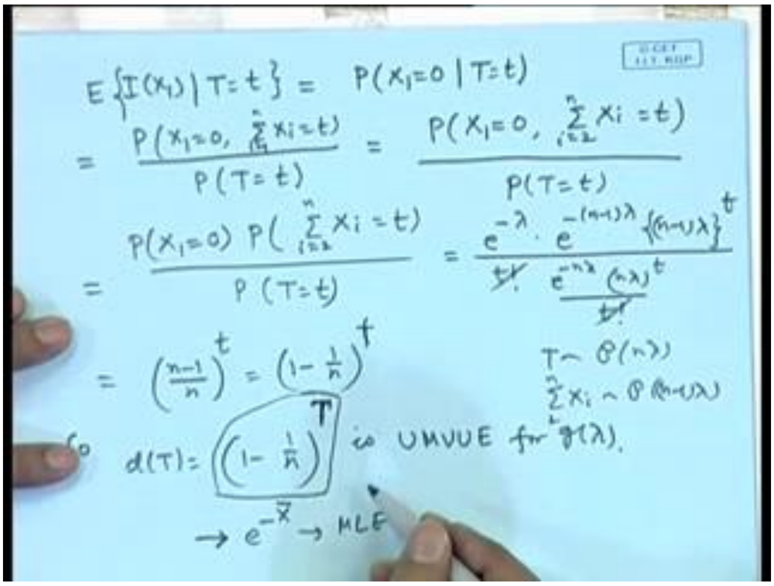


For example: expectation of T by n that is expectation of x bar that is equal to lambda. So, x bar is UMVUE for lambda. Now that answers the questions for this, because in this particular case we had written earlier x 1 has an estimator x 1 plus x 2 by 2 as another estimator. We could have also considered say T 4 as 1 by n minus 1 sigma xi minus x bar whole square this is also unbiased, because this is the sample variance and in the Poisson distribution case lambda is the population variance so sample variance is unbiased, but this is UMVUE. So, we do not have to consider any other estimator and we restrict attention to x bar for this one.

As for as the unbiased estimation is concerned we can take help of the complete sufficiency and get the best unbiased estimator. This concept is also useful to estimate certain parametric functions which are not a straight away unbiasedly estimable. We had taken a example of the probability of 0 occurrence. Now, we wrote that we can consider and estimator such as say I of x 1 is equal to one if x 1 is 0 and it is 0 if x 1 is not 0, then expectation of I x 1 is e to the power minus lambda.

Naturally this is not dependent upon the complete sufficient statistic. So, by Rao-Blackwell Lehmann-Scheffe Theorem let me write it as d T is equal to expectation of I x 1 given T this is UMVUE. So now, the question comes of determination of d T that can be determined by using the concept of conditional expectation.

(Refer Slide Time: 09:32)



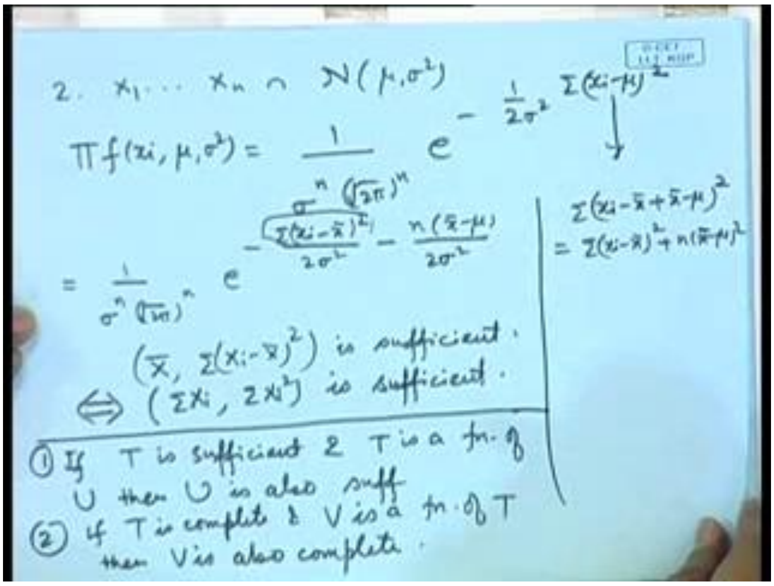
So, expectation of I x 1 given T is equal to small t that is equal to probability of x 1 is equal to 0 given T is equal to t. That is probability of x 1 is equal to 0 sigma xi is equal to t divided by probability T is equal to t. We know the distribution of T that is Poisson n lambda so the denominator is determined, the nominator is determined if we make use of the condition that x 1 is 0 then this summation reduces to sigma xi i is equal to 2 to n to be equal to t, because of first one is 0.

The advantage of writing like this is that the first one is independent of the second term, because this is x 1 and this is x 2 to x n. So, we can write it as a product x 1 is equal to 0 into product of sigma xi i is equal to 2 to n is equal to t divided by probability of T is equal to t. Once again we make use of the fact that sum of the independent Poisson random variables which is again Poisson, so this will be Poisson n minus 1 lambda that is T follows Poisson n lambda sigma xi 2 to n follows Poisson n minus 1 lambda and x 1 follows Poisson lambda. So, we can substitute these values here this is e to the power minus lambda e to the power minus n minus 1 lambda n minus 1 lambda to the power t divided by t factorial then e to the power minus n lambda n lambda to the power t by t factorial. So, these terms obviously cancel out and we are left with n minus 1 by n to the power t which can write as 1 minus 1 by n to the power t.

So, d T that is 1 minus 1 by n to the power T is UMVUE for g lambda. So, this concept of completeness and sufficiency is extremely useful for determination of the minimum variance unbiased estimators for given problems. Of course, one may wonder that this estimator looks somewhat different, because we are estimating e to the power minus lambda and what type of term we have got. But if you see carefully if I take the limit of this as n tends to infinity it is actually e to the power minus x bar, because this is nothing but n times x bar so this becomes e to the power minus x bar which was actually the maximum likelihood estimator.

So, this is another important point that in most of the practical situations asymptotically the minimum variance unbiased estimator and the maximum likelihood estimator will be same. There have been some results in this direction.

(Refer Slide Time: 12:58)



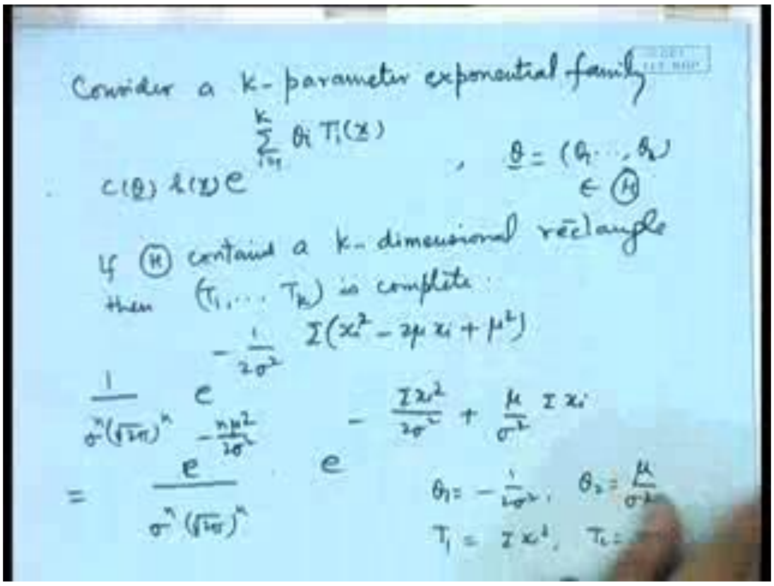
Let us take some other practical problems- say x 1, x 2, x n follows normal mu sigma square distribution. Now let us determine a complete sufficient statistics here, write down the joint distribution of x 1, x 2, x n so that is 1 by sigma to the power n root 2 pi to the power n e to the power minus 1 by 2 sigma square sigma xi minus mu square. Now if we write the term like this it is not easy to understand that what will be a sufficient statistic, because here all the observations are coming into picture.

So, we do slight algebraic simplification we can write it as 1 by sigma to the power n root 2 pi to the power n e to the power minus sigma xi minus x bar whole square by twice sigma square minus n x bar minus mu by 2 sigma square That means, we have added and subtracted sigma xi minus x bar plus x bar minus mu is square that become sigma xi minus x bar whole square plus n times x bar minus mu square, and the cross product term vanishes.

So, now you can see here that this is a function of sigma xi minus x bar whole square this is a function of x bar. So, we can say that x bar and sigma xi minus x bar whole square is sufficient any one to one function of a sufficient statics will also be sufficient. In fact, we can write the general thing that if T is sufficient and T is a function of U then U is also sufficient. On the other hand if T is complete and V is a function of T then V is also complete.

So, this implies that we can also write this as sigma xi sigma xi square. Now the question comes about checking the completeness of this, that will involve the joint distribution of x bar and sigma xi minus x bar whole square which we already know; the distribution of x bar is normal and the distribution of n minus 1 is square by sigma is square is chi square and they are independent so we can write a proof based on this. However, it may be quite complicated. Fortunately there is another result that if the distributions are in exponential family then a form of complete statistic can be determined.

(Refer Slide Time: 16:41)



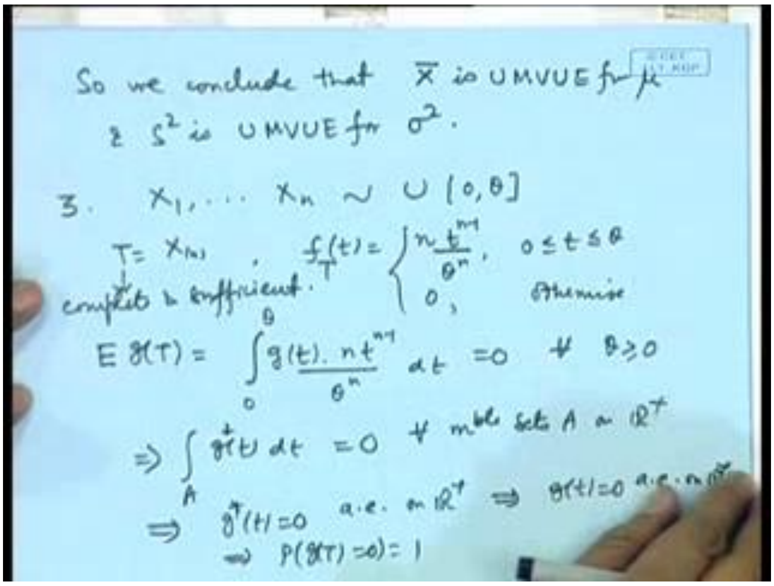
So, a general result in this direction is that consider a k- parameter exponential family; that means, the distribution is of the form e to the power sigma theta i T i xi is equal to 1 to k multiplied by sum c theta h x. This is called a k- parameter exponential family; this belongs to certain parameter express theta. If theta contains a k- dimensional rectangle then T 1, T 2, T k is complete. Sufficiency is of course obvious because of the factorization theorem, but this will also be complete. If we utilize this one then obviously, here x bar and sigma xi minus x bar whole square will be complete also. Because this can be considered as a two parameter exponential family; one of the parameters can be written as minus 1 by 2 sigma square and another parameter can be written as minus n by 2 sigma square.

So, the range of the parameter says minus infinity to infinity and minus infinity to 0 and therefore this will; I am sorry this can be written in this following fashion: 1 by sigma to the power n root 2 pi to the power n e to the power minus 1 by 2 sigma square sigma xi square minus 2 mu xi plus mu square. Now this we write down 1 by sigma to the power n root 2 pi to the power n e to the power minus n mu square by 2 sigma square e to the power minus sigma xi square by twice sigma square plus mu by sigma square sigma xi.

So, if you write in this particular fashion you can see that we can consider it as a two dimensional parameter theta 1 we can take to be minus 1 by 2 sigma square, theta 2 we can take to be mu by sigma square, T 1 we can take to be sigma xi square, and T 2 we can take to be sigma xi. So, the range of theta 1 is from minus infinity to 0 and the range of theta 2 is from minus infinity to infinity, so obviously this contains two dimensional rectangles. Therefore, sigma xi square sigma xi is a complete statistic. The sufficiency is already established through factorization theorem.

Therefore, we conclude that sigma xi sigma xi square is sufficient and complete or x bar and sigma xi minus x bar whole square is a complete and sufficient statistic, this is a one to one function of this. Now the estimation problem for finding out the minimum variance unbiased estimators becomes very simple. For example, expectation of x bar is mu, therefore x bar will be minimum variance unbiased estimator for mu. We have also proved that expectation of x square is sigma square that is sigma xi minus x bar whole square by n minus 1 which is a function of this, therefore that is also a UMVUE.

(Refer Slide Time: 20:31)

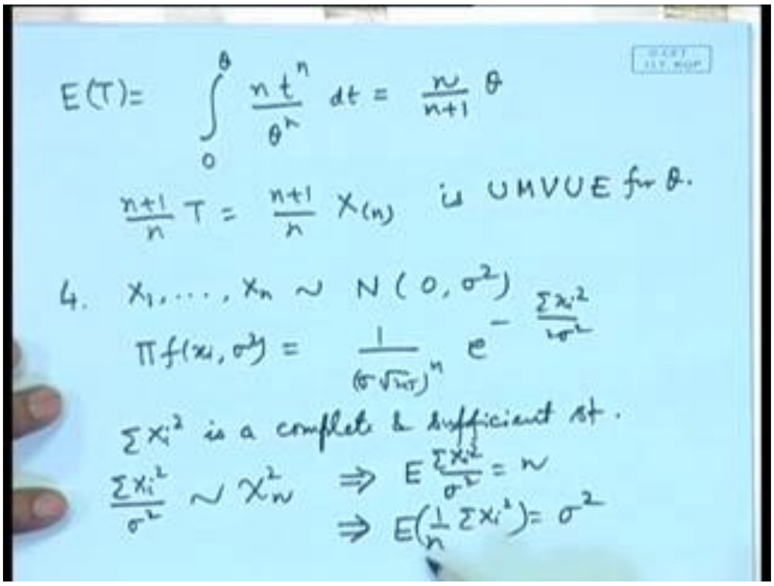


So, we conclude that x bar is UMVUE for mu and S square is UMVUE for sigma square. So, we can see here that once the determination of a complete sufficient static is done in a problem then finding out the UMVUE which is simply a problem of finding certain expectations. Let us take another familiar example there is of a uniform distribution. Here we have seen the method of movements estimator was 2 x bar which was unbiased MLE was x n. So, we write down the joint distribution, we have seen x n is actually sufficient, but what about its completeness. So let us write T is equal to x n, what is the distribution of this? It is n T to the power n minus 1 by theta to the power n 0 less than or equal to t less than or equal to theta 0 otherwise.

If we want to prove the completeness of T then let us take expectation of g T that is equal to g t n t to the power n minus 1 by theta to the power n dt from 0 to theta equal to 0 for all theta. Now you see here this is a function of t and we are saying the in integral over every interval of the form 0 to theta is 0. Now through the intervals of the form 0 to theta we can generate all the (Refer Time: 22:31) measurable sets on the positive real line. That means, we can say the integral of g is star t where g star t denotes this thing is equal to 0 for all measurable sets A on R plus, which is implying that g star t itself must be 0 almost everywhere on R plus. This implies g t is 0 almost everywhere on R plus. So, this implies at probability that g T is equal to 0 is 1. This proves that T is actually complete you have already prove that it is sufficient.

Now let us look at expectation of T.

(Refer Slide Time: 23:23)



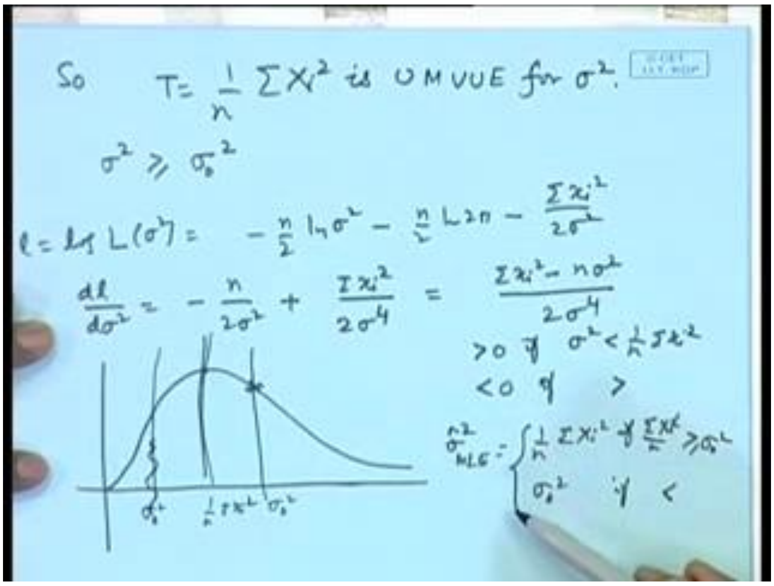
So, expectation of T is equal to n t to the power n by theta to the power n 0 to theta dt; that is equal to n by n plus 1 theta which shows that the maximum likelihood estimator is actually a biased estimator, but we can adjust this coefficient. So, we get n plus 1 by n T that is n plus 1 by n x n this will become minimum variance unbiased estimator for theta. So, you can see here that this settles the issue of that which estimator among the unbiased estimators must be chosen.

The concept of completeness and sufficiency is quite significant in a statistical inference. Here we have given examples for the estimation problems, later on we will see when we do the testing or confidence interval there also the same statistic plays a role. Here we give some more examples of estimation problems. Suppose in place of mu the mean of a normal distribution is given to be 0, let us see how this modifies the given problem. Let us write down the joint density because all the information will be derived from the distribution itself. So, product of individual density is becomes 1 by sigma root 2 pi to the power n e to the power minus sigma xi square by 2 sigma square.

Now here you do not have to do anything, you just observe that the distribution belongs to one parameter exponential family; the parameter is minus 1 by 2 sigma square and it is from minus infinity to 0, the range of minus 1 by 2 sigma square which obviously contains one dimensional interval. Therefore, sigma xi square is complete statistic. The sufficiency is clear from here. So, we conclude that sigma xi square is a complete and sufficient statistic. Now we look at the distribution of sigma xi square. The distribution of sigma xi square by sigma is square is chi square on n degrees of freedom. This means that expectation of sigma xi square by sigma is square is n, that is expectation of 1 by n sigma xi square is sigma is square.

That settles the issue here. In fact, for this problem if we find the maximum likelihood estimator that will be the method of movements estimator will be this and this is also minimum variance unbiased estimator.

(Refer Slide Time: 26:43)



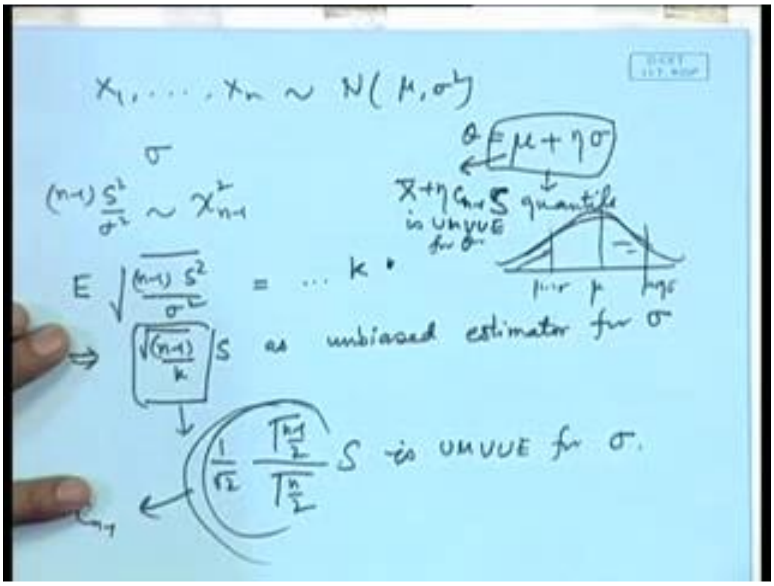
So, 1 by n sigma xi square is UMVUE for sigma square. Let us compare it with the previous work when we considered normal mu sigma is square distribution we concluded that 1 by n sigma xi minus x bar whole square is UMVUE. Now here you see that if the information about mu is there then the UMVUE is changing; that means, making use of the given information about the parameter changes our inference. A Lehmann will blindly without knowing the concept of sufficiency, he may just say that once we know that for mu we have x bar and for sigma square we have s square then he can always use in any given problem the estimators has x bar and 1 by n minus 1 sigma xi minus x bar whole square. Whereas, here you see that if we know that mu is equal to 0 then first of all for mu there is no estimation problem, because we know that the value is 0 and for sigma square also a more efficient estimator is 1 by n sigma xi square. In fact, it is better than 1 by n minus 1 sigma xi minus x bar whole square because of the UMVUE thing that we have considered here.

This also shows that whatever information is coming in the form of the likelihood function; that means the data and the parameter space that should play full role in deriving any inference in particular for estimation. Suppose here we have another restriction say sigma square is greater than or equal to sigma naught a square. Obviously, this 1 by n sigma xi square which was the MLE as well as even be we becomes slightly unreasonable estimator if it is observed that this value is less then sigma naught square. Let us see how to modify the maximum likelihood estimator.

We consider the log of the likelihood function that is equal to minus n by 2 log sigma is square minus sigma xi square by 2 sigma square. So, if we differentiate this dl by d sigma square we get minus n by 2 sigma square plus sigma xi square by 2 sigma to the power 4 which is nothing but sigma xi square minus n sigma square by sigma to the power 4. We can easily see that if n sigma square is less then sigma xi square this is positive, that is if sigma square is less than 1 by n sigma xi square and it is less than 0 if sigma square is greater than this. That means the form of the likelihood function is that it increases up to a certain value then decreases.

Now if sigma naught square is here and 1 by n sigma xi square is here then this solution is alright, whereas if this value is on this side then the maximum is occurring at this point. Therefore, the maximum likelihood estimator for sigma hat is square becomes 1 by n sigma xi square if sigma xi square by n is greater than or equal to sigma naught square and it is equal to sigma naught square if it is less. Here you can see the unbiased estimator does not belong to the given parameter space, therefore we have to discard some portion of it and get a modified estimator.

(Refer Slide Time: 31:02)



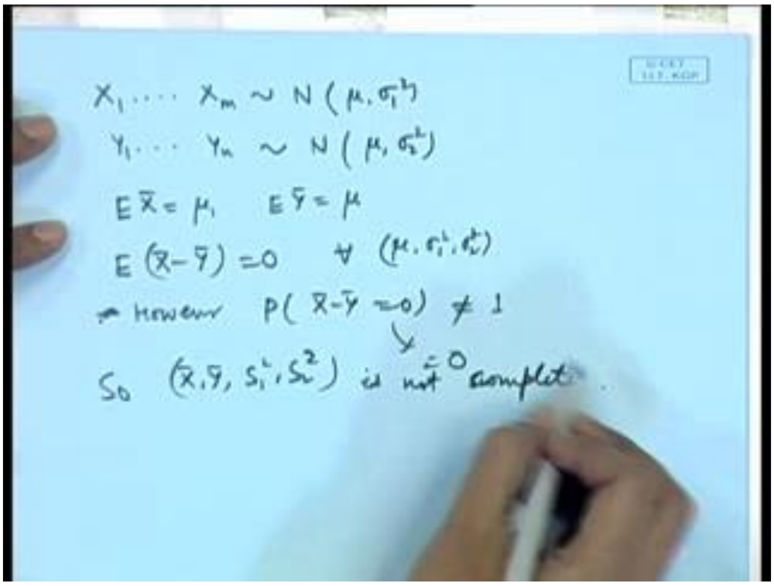
Suppose in the same problem x 1, x 2, x n follows normal mu sigma square and we are interested in the estimation of sigma. Now if maximum likelihood estimation is to be done then immediately we can take the square root of the estimator of sigma is square, but that will not preserve the unbiasedness. So, we can then make use of the concept of completeness and sufficiency. Here we know that S square by sigma square follows chi square on n minus 1 degrees of freedom. So, we take expectation of a square root of this quantity and we obtain a multiple of a constant here.

Now, we adjust that constant here, so we get here n minus 1 by k square root n minus 1 by k s as unbiased estimator for sigma; that is a standard deviation. This quantity which I have written here is actually 1 by root 2 gamma n minus 1 by 2 divided by gamma n by 2; which can be done after certain calculations, because expectation of this will involve evaluation of a gamma function which can be easily done and this term will come. So, what will get that this is UMVUE for sigma.

Now suppose we are interested in a parametric function say mu plus eta sigma, which is nothing but a quantiles are locations on the distribution we have defined it earlier. So, suppose this is a normal distribution mu. So, mu plus eta sigma may be a particular quantile mu minus eta sigma may be another quantile etcetera and we may be interested to find a UMVUE of this. Then easily we can see because of the linearity we can put x bar plus eta and this particular term let us denote it by say c n minus 1; c n minus 1 sigma then this will become UMVUE, a c n minus 1 as this will be UMVUE for this parametric function which we call say theta.

So, you can see that for various kinds of parametric functions the UMVUE is can be derived once we have the complete sufficient statistic with this. The only disadvantage is that sometimes the complete sufficient statistic may not exist; that means a statistic which we are considering a sufficient it may not be complete.

(Refer Slide Time: 34:08)



A example for this situation is suppose I say x 1, x 2, x m follows normal mu sigma 1 is square y 1, y 2, y n follows normal mu sigma 2 is square, then expectation of x bar is mu expectation of y bar is also mu. So, expectation of x bar minus y bar is 0 for all parametric functions. However, probability that x bar minus y bar is 0 is not 1. In fact, this probability is actually equal to 0. So, x bar y bar S 1 square, S 2 is square is not complete.

In this case we cannot make use of the Rao-Blackwell Lehmann-Scheffe Theorem. In fact, there is another result here which says that the UMVUE for mu does not exist. So, that may happen sometimes. However, this concept is extremely useful as we have seen.

In the next lecture we will be discussing the interval estimation; that means, in place of a single value as an estimate for a certain parametric function we will give a range of values and we will say that with a certain confidence the parameter lies into that range.