Probability and Statistics

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

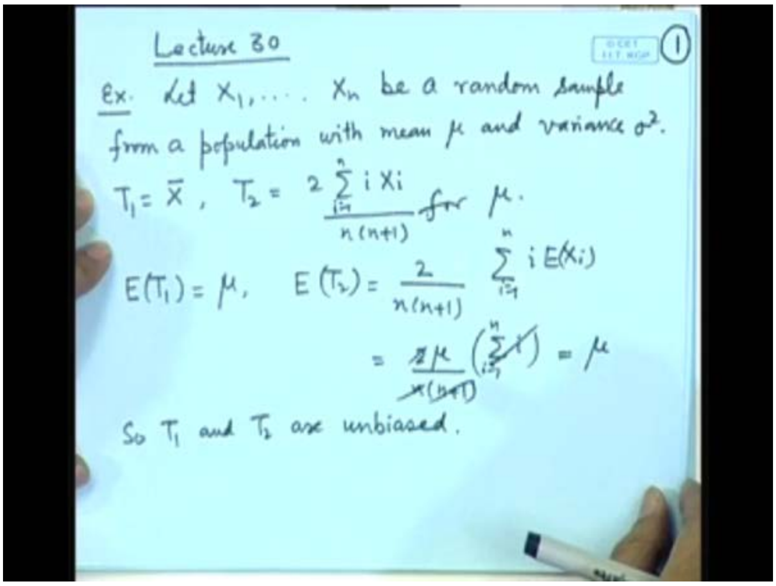
Module No. # 01

Lecture No. # 30

Estimation-IV

Yesterday, we have introduced the criteria that among given estimators, which estimator should be preferred. For example, if T 1 and T 2 are two unbiased estimators for the same parameter g theta, then we will prefer T 1 over T 2 if variance of T 1 is less than or equal to variance of T 2. In general, if I am considering any two estimators; that means, they need not be unbiased; in that case, we will compare the mean squared errors and the estimator with smaller mean squared error will be preferred.

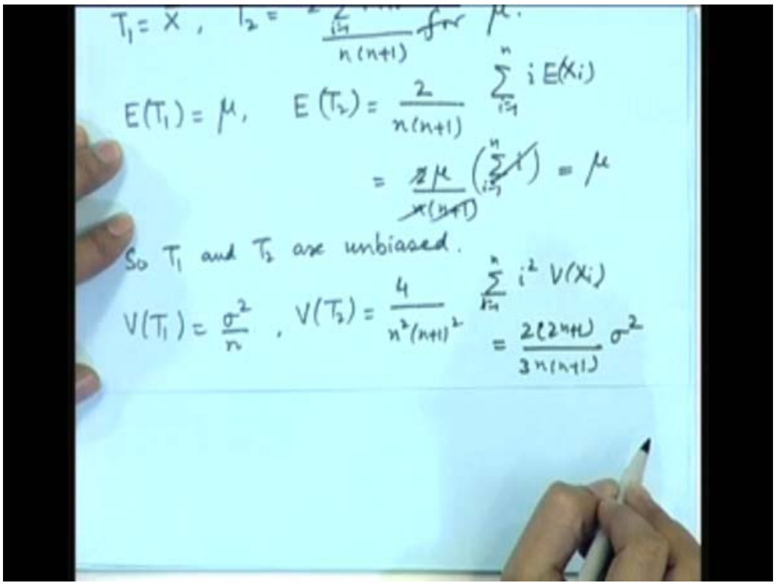
(Refer Slide Time: 00:52)



Let me give one example here. Suppose we have a random sample from a population with mean mu and variance sigma square. Now, let us consider estimators T 1 and T 2 for mu. Let us see. So, what is expectation of T 1? Naturally, it is equal to mu. What is expectation of T 2? You can apply the linearity property of the expectation. So, this becomes twice divided by n into n plus 1 sigma i expectation of X i. Now, expectation of

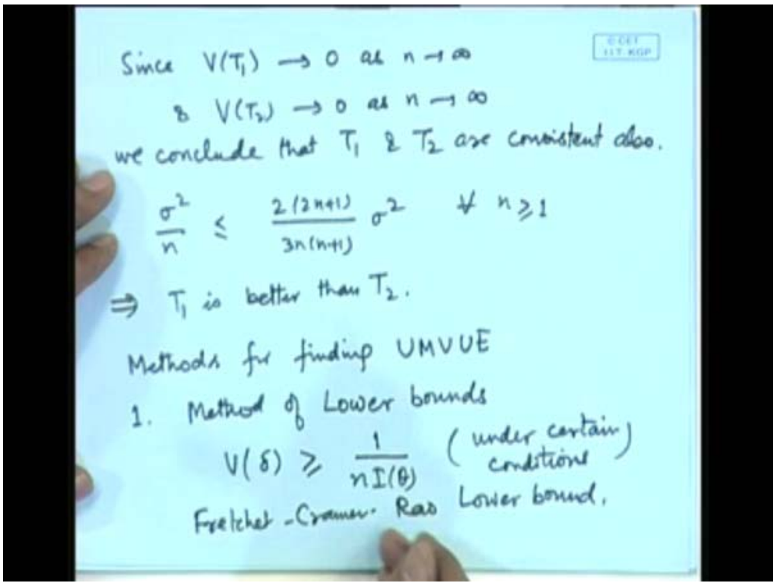
X i is mu. So, this reduces to 2 mu by n into n plus 1 sigma of i; i is equal to 1 to n. Now, this is nothing but n into n plus 1 by 2. So, this cancels out with this and you get that both T 1 and T 2 are unbiased estimators. So, T 1 and T 2 – both of them are unbiased.

(Refer Slide Time: 01:58)



Let us look at variances. So, what is variance of T 1? Variance of T 1 is sigma square by n. We have already shown that the variance of the sample mean is equal to sigma square by n. Let us consider the variance of T 2. Now, variance of T 2, because of the independence, becomes 4 by n square into n plus 1 square sigma i square variance of Xi. Now, variance of Xi is sigma square. And, this is sigma i square from 1 to n; that is, the sum of the squares of first n integers; that is, n into n plus 1 into 2 n plus 1 by 6. So, after simplification, this quantity turns out to be twice into 2 n plus 1 divided by 3 n into n plus 1 sigma square. So, now the question is that we can also check the consistency here. For example, both of them are unbiased and variance of T 1 goes to 0 as n tends to infinity; variance of T 2 also goes to 0 as n tends to infinity.

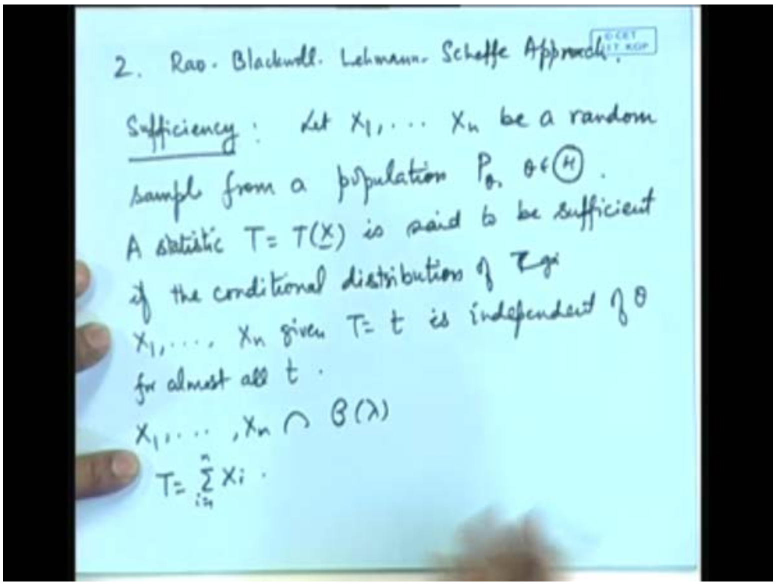
(Refer Slide Time: 03:08)



Since variance of T 1 goes to 0 as n tends to infinity and variance of T 2 goes to 0 as n tends to infinity, we conclude that T 1 and T 2 are consistent. So, we have two estimators; both of them are unbiased, both of them are consistent also. Again, which one we will prefer? So, we compare the variances. You can see easily that sigma square by n is less than or equal to 2 into 2 n plus 1 by 3 n into n plus 1 sigma square for all n greater than or equal to 1. Actually, for n is equal to 1, the two sides will be equal. So, this implies that T 1 is better than T 2.

Now, the question comes that among a set of given estimators, we can find by comparing the variances or mean squared errors. But, in the first place, how to find the best among them? Because the total set of estimators is infinite, we need certain other methodology. There are two methods for finding out the unbiased – methods for finding UMVUE. One method is the method of lower bounds. Under certain given conditions, variance of an unbiased estimator is greater than or equal to a prescribed number. It is 1 by n times I theta. This is under certain conditions. This is called Fretchet-Cramer Rao lower bound. So, if there is an estimator, which will have this variance equal to this, (Refer Slide Time: 05:33) that will be naturally minimum variance and biased estimator. Then, later on, the generalizations of this Fretchet-Cramer Rao bound have been done. And, we have the bounds when we have multi-parameter situations, when we can use higher order derivatives, etcetera. But, for application of these lower bounds, certain conditions need to be satisfied and the bounds may not always be attained.

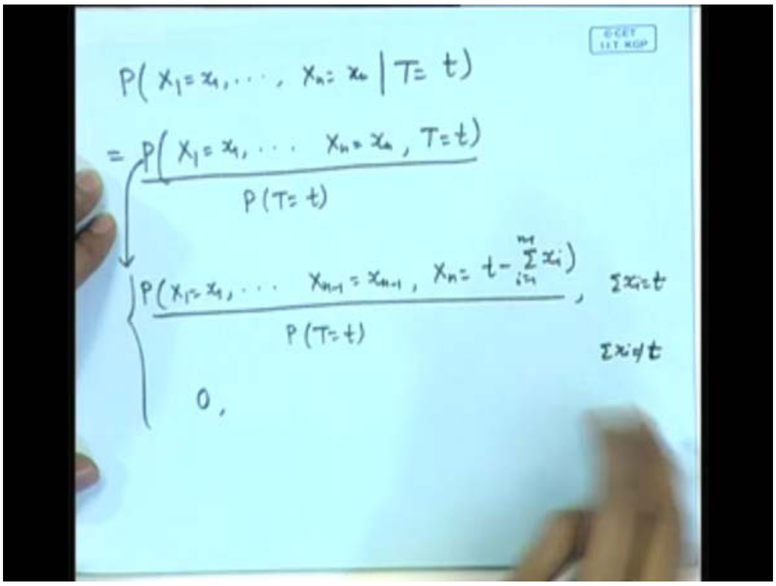
(Refer Slide Time: 06:04)



There is another approach that is called Rao-Blackwell Lehmann-Scheffe approach. We introduced two concepts; that is of sufficiency and completeness. Firstly, we define what sufficiency is. We have the regular model that X 1, X 2, X n be a random sample from a population say P theta, theta belonging to script theta. Then, a statistic T is said to be sufficient if the conditional distribution of X 1, X 2, X n given T is equal to t is independent of theta for almost all t.

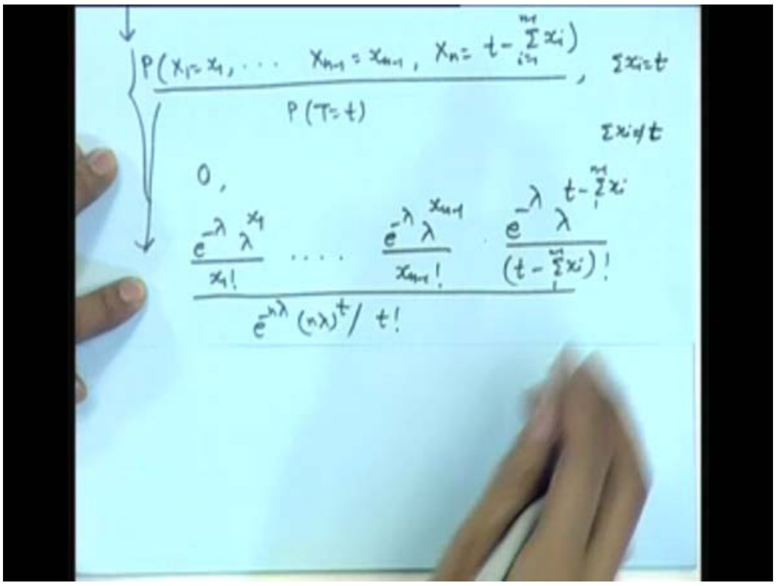
Let us take an example here. Suppose I consider X 1, X 2, X n, follow say Poisson lambda distribution. Let us define T to be sigma of X i; i is equal to 1 to n.

(Refer Slide Time: 08:08)



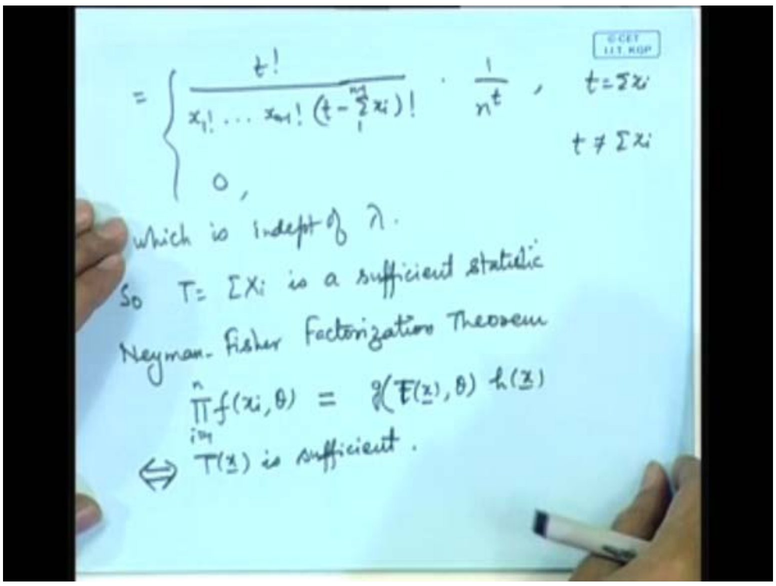
Let us consider the conditional distribution of x 1, x 2, x n given T is equal to t. So, this is equal to probability of X 1 is equal to x 1 and so on, X n is equal to x n, T is equal to t divided by probability of T is equal to t. Now, here T is sigma X i and we know that it follows Poisson n lambda. So, the denominator quantity can be written. How to find out the numerator quantity? We simplify this; we can write it as probability of X 1 is equal to x 1 and so on. X n minus 1 is equal to x n minus 1. And, X n is equal to t minus sigma x i; i is equal to 1 to n minus 1. This is valid if sigma x i is equal to t; otherwise, it is defined to be 0. Because it is conditional on t, for every value of t, we have to determine this. Now, the numerator quantity can be determined, because X 1, X 2, X n are independent Poisson lambda variables.

(Refer Slide Time: 09:32)



So, we can substitute these values here. e to the power minus lambda into lambda to the power X 1 by x 1 factorial and so on. And, the last term will be e to the power minus lambda lambda to the power t minus sigma xi 1 to n minus 1 divided by t sigma xi, 1 to n minus 1 factorial and divided by e to the power minus n lambda n lambda to the power t divided by t factorial. You can easily see that e to the power minus lambda term cancels out, because we have n terms here; and in the denominator, we have e to the power minus n lambda. The powers of lambda, that is, lambda to the power t here; and, in the denominator, we have lambda to the power t. So, that also cancels out.

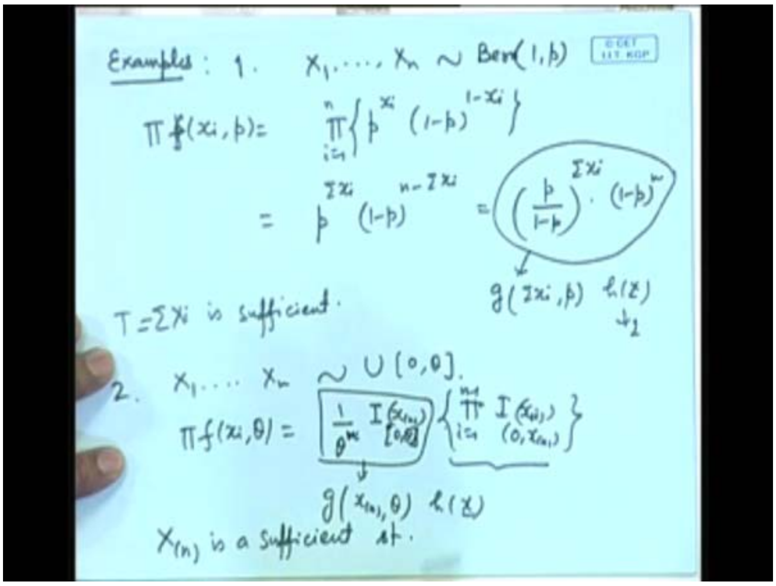
(Refer Slide Time: 10:26)



So, we are left with t factorial divided by x 1 factorial and so on x n minus 1 factorial t minus sigma xi; 1 to n minus 1 factorial and 1 by n to the power t when t is equal to sigma xi. And, it is equal to 0 if t is not equal to sigma xi. Now, you can see here this term does not depend upon lambda. This is independent of lambda. So, we conclude that T, that is, sigma Xi is a sufficient statistic. The role of sufficiency is quite important in statistical inference. In fact, it means that we can generate an alternative sample say x 1 prime, x 2 prime, x n prime given T is equal to t. That means, whatever information about the parameter can be drawn from the sample x 1, x 2, x n, all of that is contained in sigma xi. That means, there is no additional information in x 1, x 2, x n, which is not there is sigma xi. This allows us to make the data compact, because we need not keep record of all the individual observations, rather we keep record of only the sufficient statistics.

Now, this method of proving that sigma xi is sufficient involves finding out the conditional distribution and which may be quite cumbersome for various problems. And, another thing is that here we have to guess also that what would be a sufficient statistic. So, there is another result, which is known as Neyman-Fisher factorization theorem, which allows us to figure out what will be a sufficient statistic in a given problem. I will not state the theorem in a full form; rather we look at the practical aspect of it. We write down the joint density, that is, product of f x I, theta; i is equal to 1 to n. This is a joint probability density function (Refer Slide Time: 12:54) of x 1, x 2, x n. If this can be factorized as g t x, theta into h x, where the first term depends upon xi’s only through t and the second term is free from theta. Then, we say that this implies and implied by that T x is sufficient. The proof of this involves slightly major theoretic consideration. So, we skip the proof. But, this is a very practical way of obtaining sufficient statistics.

(Refer Slide Time: 13:41)

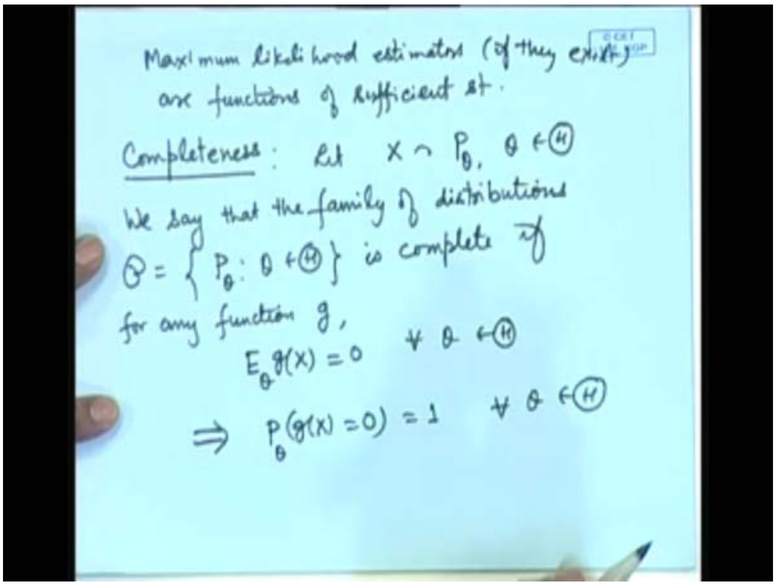


So, let us look at the applications of this. Let us consider say X 1, X 2, X n follows Bernoulli distribution. So, the joint distribution here, product i is equal to 1 to n p to the power xi 1 minus p to the power 1 minus xi; that is, p to the power sigma xi 1 minus p to the power n minus sigma xi. This we can write as p by 1 minus p to the power sigma xi multiplied by 1 minus p to the power n. Now, you can see here, this term is a function of sigma xi and p alone; and, h x we can take to be 1. So, this proves that sigma Xi is sufficient.

Let us look at the practical aspect of it. If we have conducted n Bernoullian trials, we will be interested and we want to draw certain inference on the proportion of the success. Then, you can see that sigma Xi is actually the number of successes here. So, that gives the full information about p. We do not have to keep track of individual xi's. Suppose we consider say uniform distribution. Then, the joint density is equal to 1 by theta to the power n. Now, one may say that if we write like this, then where is the variable coming in, which will be sufficient? But, this is not a complete description, because for complete description, we need to write down the range of the variables, which is each of the xi's, is from 0 to theta. So, we can write it in the terms of indicator function that x n is from 0 to theta. And, remaining xi's – they are between 0 to x n product i is equal to 1 to n minus 1. So, this part we can consider as g of x n and theta. And, this part we can consider as h x. So, by factorization theorem, we conclude that X n is a sufficient statistic.

Let us also correlate with the discussion that we had in the previous lecture about the maximum likelihood estimators. The derivation of the maximum likelihood estimator involved the full probability density function or probability mass function of x 1, x 2, x n, which we termed as the likelihood function. And, that function we maximized with respect to the parameter. Now, if you look at the factorization, then (Refer Slide Time: 16:56) this term does not play a role, because if I take l n for example, then this term will be separated out and the maximization problem reduces only to the maximization of this function. So, naturally, theta will become a function of T x alone.

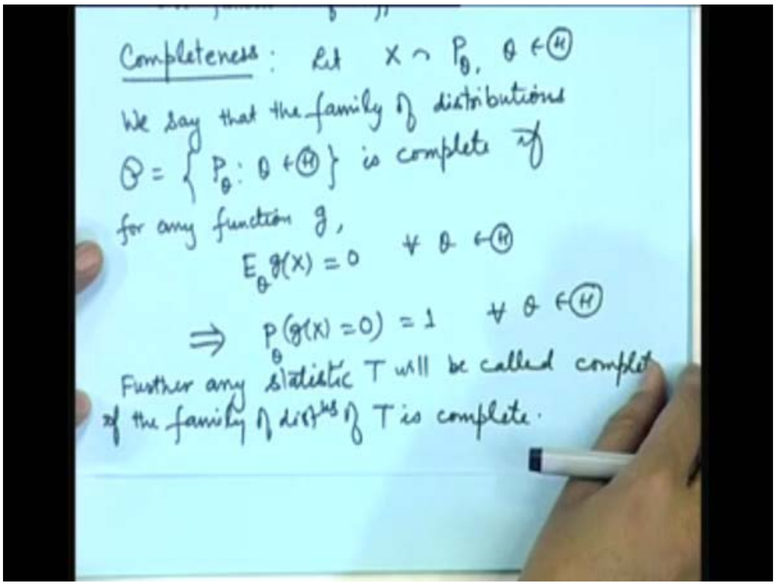
(Refer Slide Time: 17:16)



So, we conclude here that maximum likelihood estimators – if they exist – are functions of sufficient statistic. So, that brings us the importance of the sufficiency; that means, whatever inference we draw, finally, we can restrict attention to the sufficient statistic. We will look at further examples of this later.

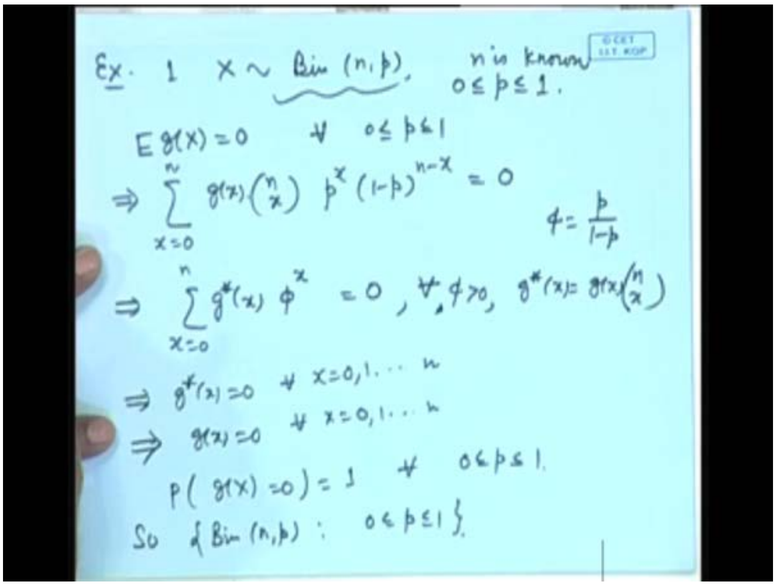
Let me introduce another concept called completeness. Let X follow a distribution say P theta; theta belonging to theta. So, we say that the family of distributions, P is equal to P theta is complete if for any function g, expectation of g x is equal to 0 for all theta, implies probability of g x is equal to 0 is equal to 1.

(Refer Slide Time: 19:08)



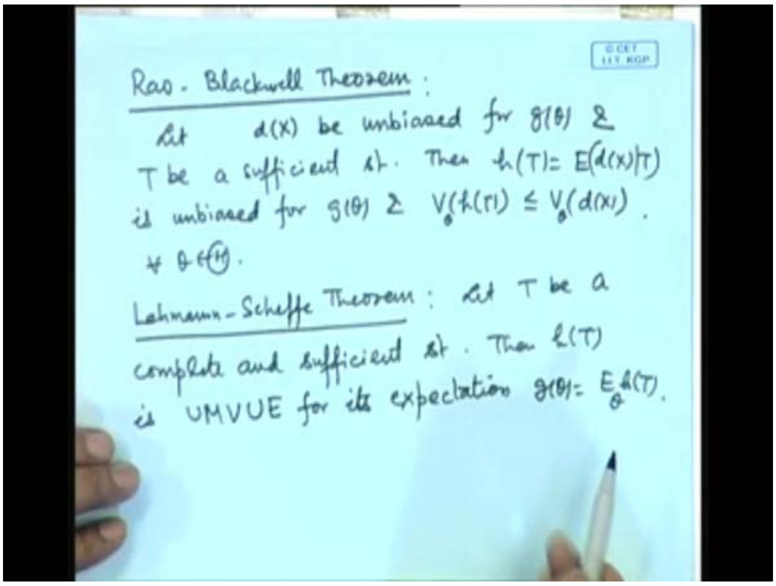
Further any statistic T will be called complete if the family of distributions of T is complete. So, let me give the example here and explain that what the meaning of this is. What we are saying is that whenever expectation of g x is 0, that function itself is 0. That means the only unbiased estimators of 0 are 0 itself. Now, this is a very important statement. And, let us see that why this is true for various distributions.

(Refer Slide Time: 20:00)



Let us take say x follows binomial (n, p), where n is known and p is the parameter. Let us look at expectation of g x is equal to 0. Now, this condition is equivalent to g x, n C x, p to the power x 1 minus p to the power n minus x is equal to 0 for x is equal to 0 to n. We may introduce a term called say phi is equal to p by 1 minus p. Then, this term I can write as sigma g star x phi to the power x is equal to 0 for x equal to 0 to n, where g star I have written as g x into n C x. Now, the left-hand side is a polynomial of degree n in phi and we are saying that it is vanishing for all phi. So, a polynomial will vanish identically on an interval provided all it is coefficient vanish; that means, g star x is 0 for all x is equal to 0, 1 to n, which implies that g x itself is 0 for all x is equal to 0, 1 to n. So, probability that g x is equal to 0 will be 1 for all p. So, this family of binomial distributions is a complete family of distributions. So, 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 unbiased estimators.

(Refer Slide Time: 22:19)

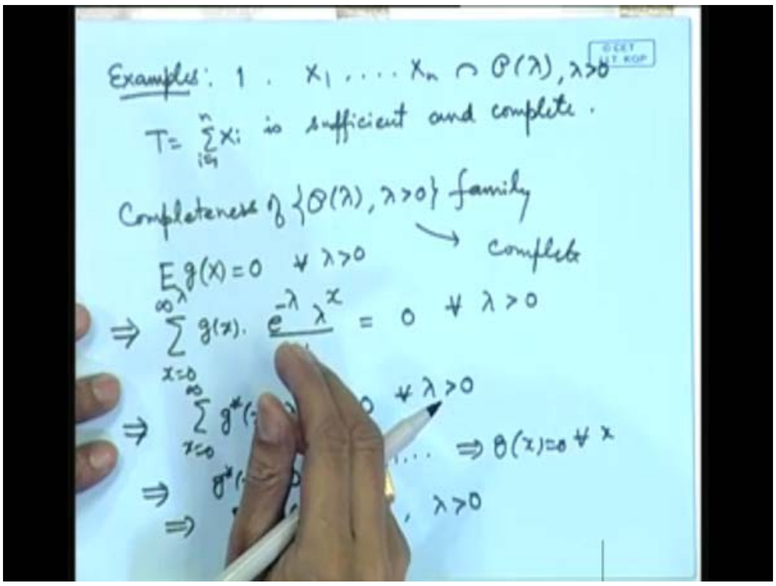


The Rao-Blackwell theorem says that let d x be unbiased for g theta and T be a sufficient statistic. Then, let us define say h T as expectation of d x given T, conditional expectation. Then, this is unbiased for g theta; and, variance of h T is less than or equal to variance of d x for all theta. Now,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

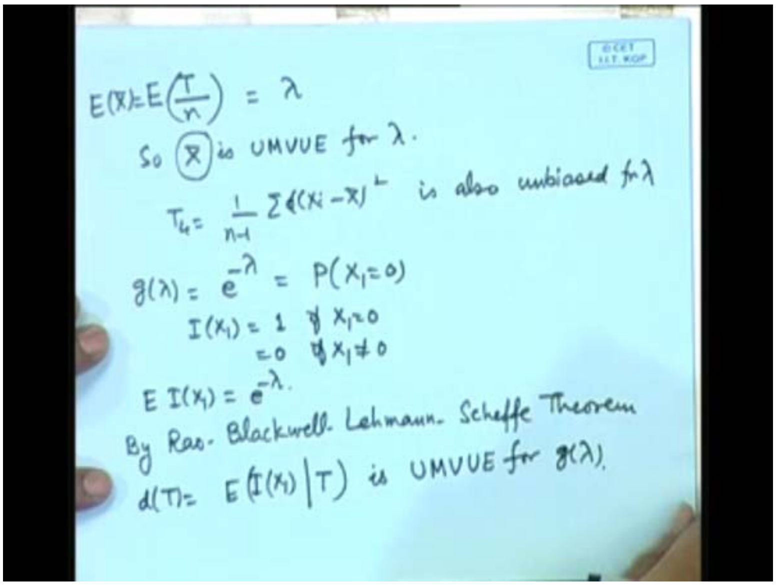
(Refer Slide Time: 25:55)



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. So, 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 distributions 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 statement is equivalent to g x e to the power minus lambda lambda to the power x by x factorial. So, since lambda is positive, 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. So, 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.

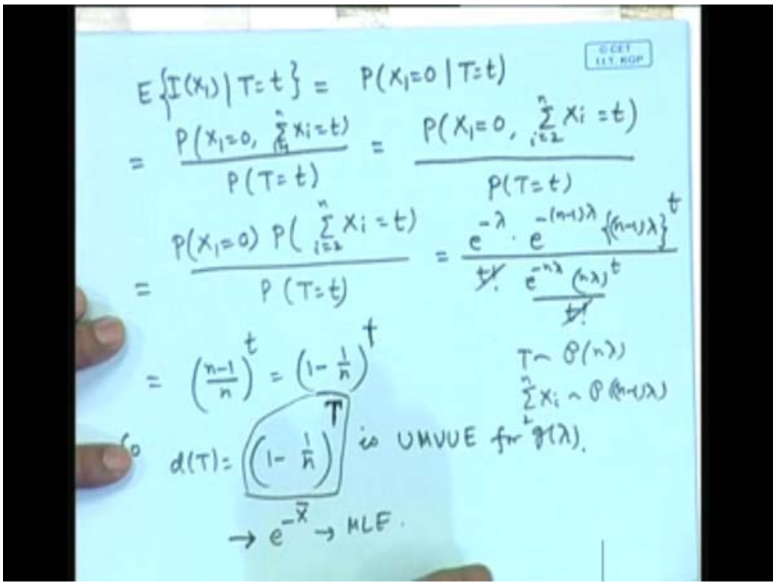
(Refer Slide Time: 28:36)



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 a question for this, because in this particular case, we had written earlier X 1 as 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 a sample variance; and, in the Poisson distribution case, lambda is a population variance. So, sample variance is unbiased. But, this is UMVUE (Refer Slide Time: 29:28). So, we do not have to consider any other estimator and we restrict attention to X bar for this one. As far 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 straight away unbiasedly estimable. We had taken an example of the probability of 0 occurrence. Now, we wrote that we can consider an estimator such as say I of X 1 is equal to 1 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: 31:13)

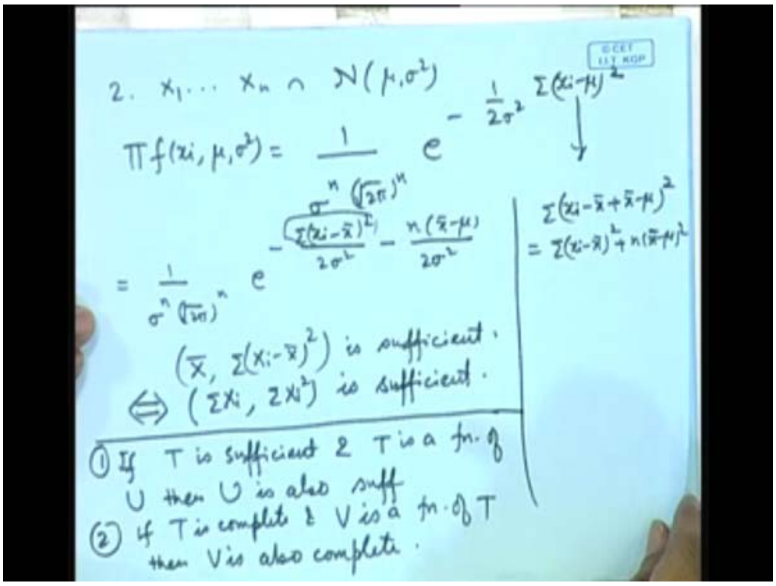


So, expectation of I X 1 given T is equal to small t; that is equal to probability of X 1 is equal to 0 given T is equal to t; that is, probability of X 1 is equal to 0, sigma Xi is equal to t divided by probability T is equal to t. We know the distribution of T; that is, Poisson n lambda. So, the denominator is determined. The numerator is determined if we make use of the condition that X 1 is 0. Then, this summation reduces to sigma X i, i is equal to 2 to n to be equal to t, because the 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 X i, 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 some of the independent Poisson random variable is again Poisson. So, this will be Poisson n minus 1 lambda; that is, T follows Poisson n lambda; sigma X i, 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 we 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.

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 here fully, 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 some results in this direction.

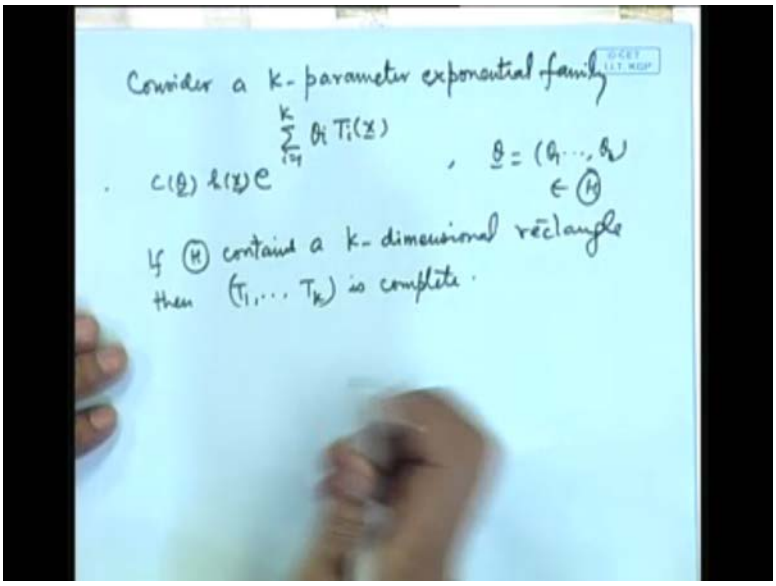
(Refer Slide Time: 34:39)



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 into 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 into e to the power minus sigma X i 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 X i minus X bar plus X bar minus mu square. That becomes sigma X i minus X bar whole square plus n times X bar minus mu square. And, the cross product term vanishes.

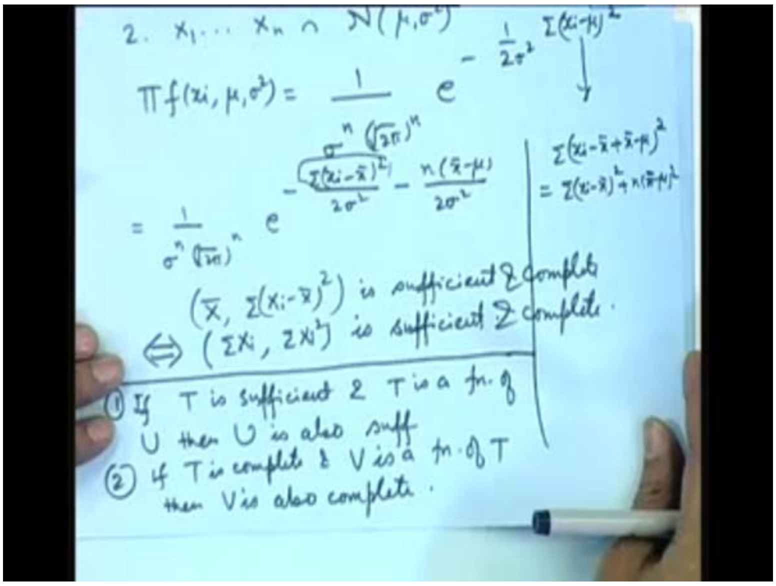
Now, you can see here that this is a function of sigma X i minus X bar whole square (Refer Slide Time 36:21). This is a function of X bar. So, we can say that X bar and sigma X i minus X bar whole square is sufficient. Any one-to-one function of a sufficient statistics will also be sufficient. In fact, 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 (Refer Slide Time: 37:30) sigma X i sigma, X i square. Now, the question comes about checking the completeness of this. That will involve the joint distribution of X bar and sigma X i minus X bar whole square, which we already know, the distribution of X bar is normal and the distribution of n minus 1 S square by sigma 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: 38:22)



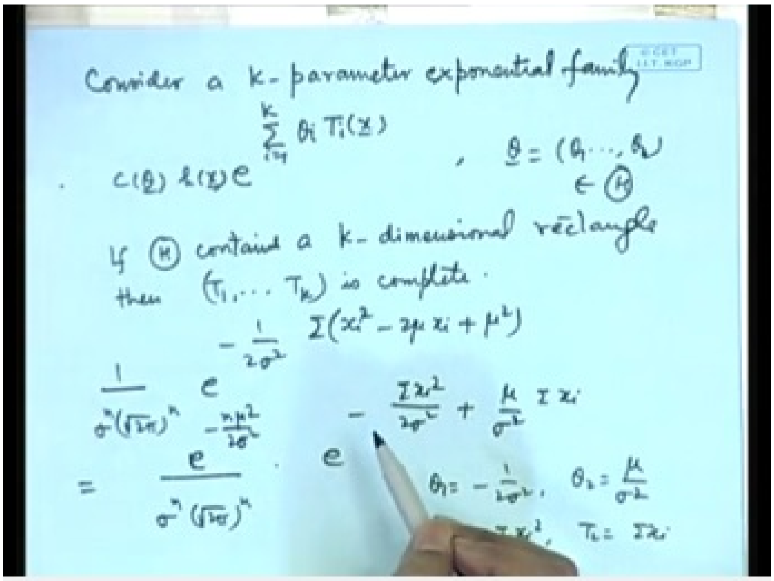
A general result in this direction is that consider a k-parameter exponential family; that means, the distribution is of the form e to the power sigma theta i T i x; i is equal to 1 to k multiplied by some C theta h x. This is called a k-parameter exponential family. This belongs to certain parametric space, 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.

(Refer Slide Time: 39:39)



If we utilize this one, then obviously, here (Refer Slide Time: 39:32) X bar and sigma X i 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 parameters is minus infinity to infinity and minus infinity to 0. And therefore, this can be written in this following fashion.

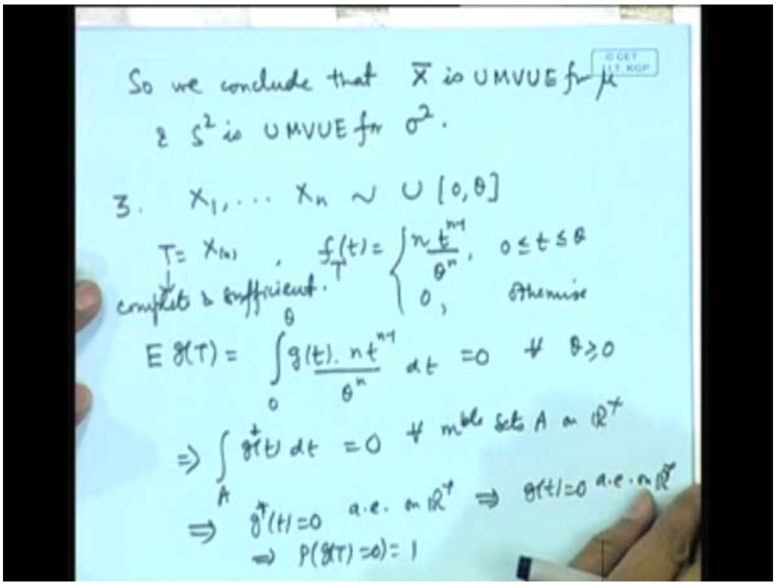
(Refer slide Time: 40:24)



1 by sigma to the power n root 2 pi to the power n into e to the power minus 1 by 2 sigma square sigma x i 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 into e to the power minus n mu square by 2 sigma square into e to the power minus sigma x i square by twice sigma square plus mu by sigma square into 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 x i square; and, T 2 we can take to be sigma x i. 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 x i square, sigma x i is a complete statistic.

The sufficiency is already established through factorization theorem. Therefore, we conclude that (Refer Slide Time: 41:35) sigma X i, sigma X i square is sufficient and complete; or, X bar and sigma X i 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 S square is sigma square; that is, sigma X i minus X bar whole square by n minus 1, which is a function of this (Refer Slide Time: 42:07). Therefore, that is also a UMVUE.

(Refer Slide Time: 42:13)

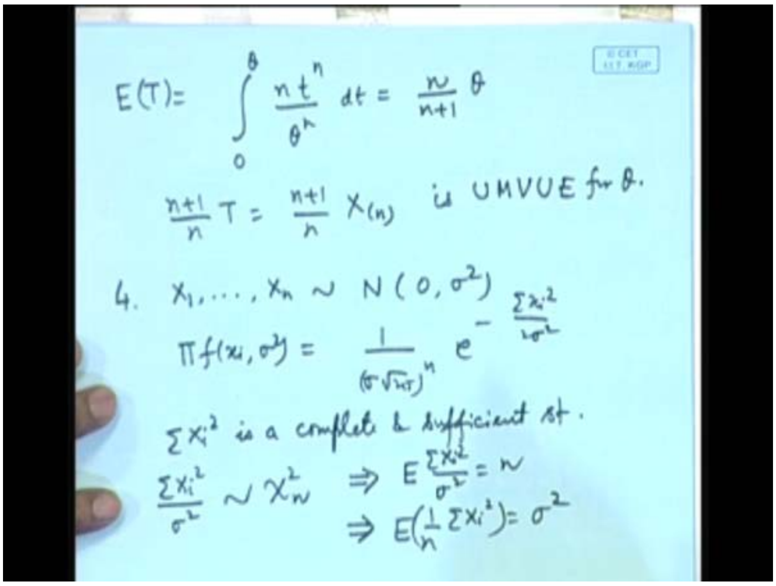


So, we conclude that X bar is UMVUE for mu; and, S square is UMVUE for sigma square. So, we can see here that once the determination of a complete sufficient statistic is done in a problem, then finding out the UMVUE is simply a problem of finding certain expectations.

Let us take another familiar example that is of a uniform distribution. Here we have seen the method of moment 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 a 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 d t 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 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 levy measurable sets on the positive real line; that means, we can say the integral of g star t, where g star t denotes this thing, (Refer Slide Time: 44:21) 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 that probability that g T is equal to 0 is 1. This proves that T is actually complete; we have already proved that it is sufficient.

(Refer Slide Time: 45:04)

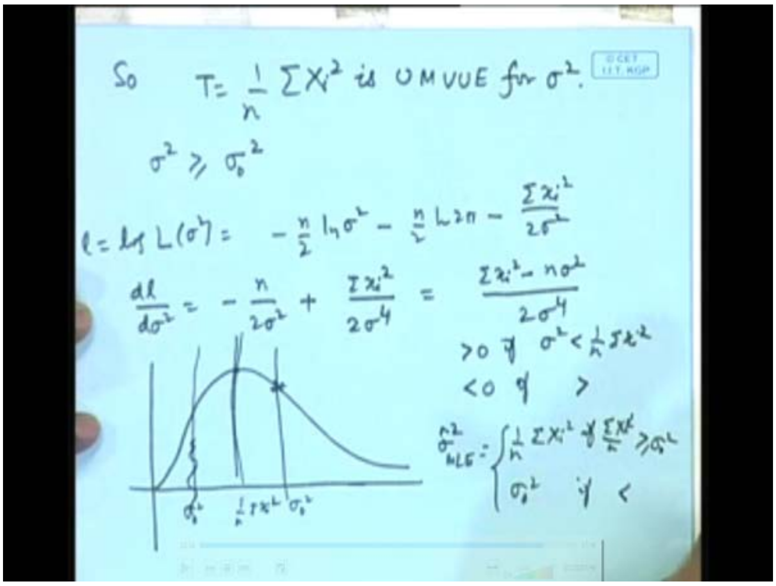


Now, let us look at expectation of T. Expectation of T is equal to n t to the power n by theta to the power n; 0 to theta d t; that is equal to n by n plus 1 into theta, which shows that the maximum likelihood estimator is actually a biased estimator. But, we can adjust this coefficient. So, we get n plus 1 by n into T; that is, n plus 1 by n into X n. This will become minimum variance unbiased estimator for theta. So, you can see here that this settles the issue of which estimator among the unbiased estimators must be chosen. The concept of completeness and sufficiency is quite significant in 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 densities becomes 1 by sigma root 2 pi to the power n into e to the power minus sigma x i 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 intervals. Therefore, sigma x i square is complete statistic. The sufficiency is clear from here. So, we conclude that sigma X i square is a complete and sufficient statistic.

Now, we look at the distribution of sigma X i square. The distribution of sigma X i square by sigma square is chi square on n degrees of freedom. This means that expectation of sigma X i square by sigma square is n; that is, expectation of 1 by n sigma X i square is sigma square. That settles the issue here. In fact, for this problem, if we find the maximum likelihood estimator, that will be this (Refer Slide Time: 48:13). The method of moment estimators will be this. And, this is also minimum variance unbiased estimator.

(Refer Slide Time: 48:24)

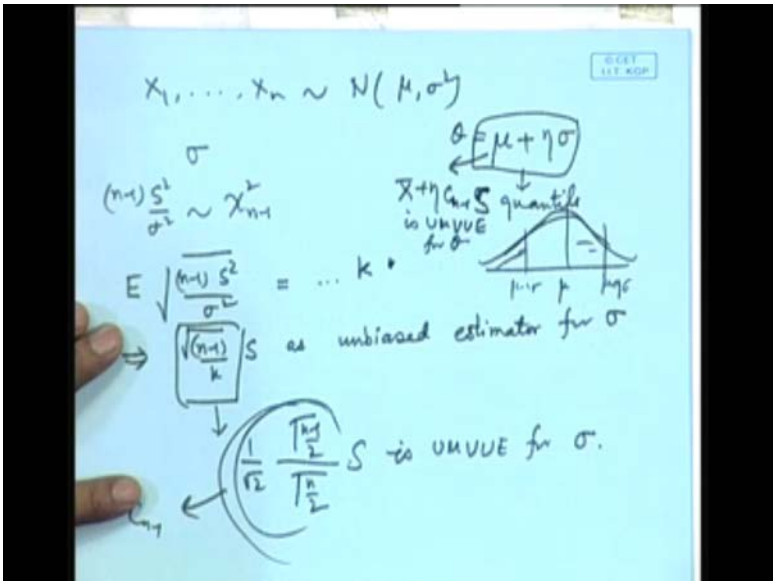


So, 1 by n sigma X i square is UMVUE for sigma square. Let us compare it with the previous work. When we considered normal mu sigma square distribution, we (Refer Slide Time: 48:42) concluded that one by n sigma X i minus X bar whole square is UMVUE. Now, here you see, that if the information about mu is there, then the UMVUE is changing; that means, making use of the given information about the parameter changes our inference. A layman will blindly without knowing the concept of sufficiency 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 as X bar and 1 by n minus 1 sigma X i 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 X i square. In fact, it is better than 1 by n minus 1 sigma X i 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 square. Obviously, this 1 by n sigma X i square, which was the MLE as well as UMVUE becomes slightly unreasonable estimator if it is observed that this value is less than 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 square minus sigma x i 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 x i square minus n sigma square by 2 sigma to the power 4.

You can easily see that if n sigma square is less than sigma x i square, this is positive; that is, if sigma square is less than 1 by n sigma x i 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 (Refer Slide Time: 51:51) and 1 by n sigma x i square is here, then this solution is alright; whereas, if this value is in this side, then the maximum is occurring at this point. Therefore, the maximum likelihood estimator for sigma head square becomes 1 by n sigma X i square if sigma X i square by n is greater than or equal to sigma naught square; and, it is equal to sigma naught square if it is less. So, here you can see the unbiased estimator does not belong to the given parametric space. Therefore, we have to discard some portion of it and get a modified estimator.

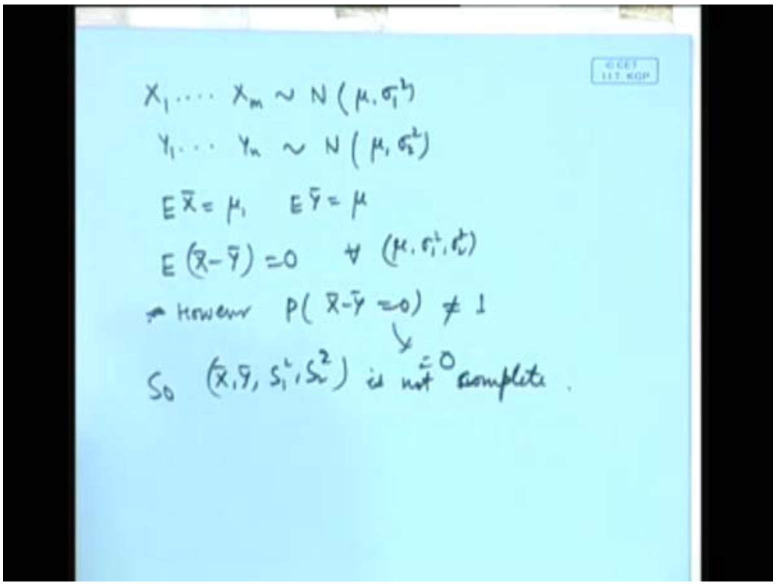
(Refer Slide Time: 52:43)



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 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 square root of this quantity and we obtain a multiple of a constant here. Now, we adjust that constant here. So, we get here 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 we 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 quantile; quantiles are locations on the distribution; we have defined it earlier (Refer Slide Time: 54:43). 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 (Refer Slide Time: 54:56). Then, easily we can see, because of the linearity, we can put X bar plus eta; and this particular term – let us denote it say C n minus 1 – C n minus 1 sigma; then, this will become UMVUE. C n minus 1 S. This will be UMVUE for this parametric function, (Refer Slide Time: 55:18) which we call say theta. So, you can see that for various kind of parametric functions, the UMVUE's can be derived once we have the complete sufficient statistic with us. The only disadvantage is that sometimes the complete sufficient statistic may not exist; that means, a statistic, which we are considering sufficient may not be complete.

(Refer Slide Time: 55:48)



An example for this situation is suppose I say X 1, X 2, X m follows normal mu sigma 1 square; Y 1, Y 2, Y n follows normal mu sigma 2 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 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 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.