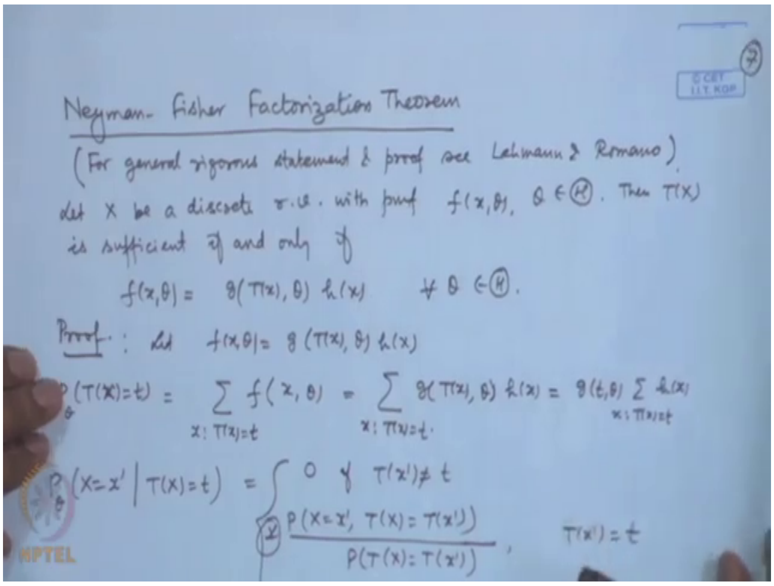
Statistical Inference
Prof. Somesh Kumar
Department of Mathematics
Indian Institute of Technology, Kharagpur

Lecture – 24
Sufficiency – II

That we notice when we are proving that t is equal to sigma xi sufficient in the 2 examples that I have consider is that, we are already guessing what is a sufficient statistics. Now in many given problems it may not be obvious that what is sufficient and therefore, this definition of taking conditional distribution to prove that a given statistic is sufficient maybe to combustion. And, moreover it may give rise to like we consider conditional distribution of x 1 x 2 x n given say x 1 minus x 2 plus x 3 minus x 4 and so on.

It may turn out that this is not free, we may take sigma xi square, it may not be free from theta. Then how to get or you can say, how to get guess an sufficient statistic. Fortunately for this there is a important result called factorization theorem, which readily produces the sufficient statistics.

(Refer Slide Time: 01:27)



So, this is known as Neyman Fisher factorization theorem, named after RA Fisher and Jerzy Neyman, who proved in around 1939. We are not going to give a very very rigorous statement and proof of this theorem, which will be applicable to all situations, rather we will consider a discrete case here and to write the proof here, for general rigorous statement and proof see the book of say Lehmann and Romano. We are considering a discrete case here, let X be a discrete random variable, with probability mass function say f x theta theta belonging to discrete theta, then T X is sufficient if and only if, f x theta is equal to g T x theta into h x for all theta. So, we are calling this as the factorization theorem.

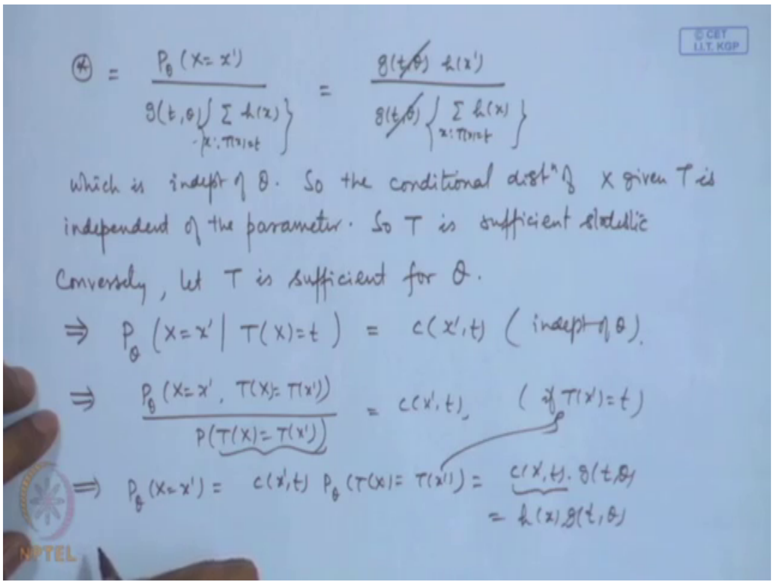
What I am saying is the distribution can be written as product of 2 terms, g and h, where h is a term where parameter does not appear. In the term g, the parameters theta appears but appearance of x is through T alone. So, if that is happening then we say T is sufficient. So, the factorization means that the part of the distribution where the parameter is involved should involve only the sufficient statistics in the form of variables and the other term should be free from the parameter.

Let us look at a proof of this. So, we are considered the discrete case here. So, let us assume say f x theta is equal to g of T x theta h x. Now, let us consider say probability that T x is equal to t; that is equal to now, this is a probability which is involving a function of the random variable x. So, this can be considered as the sum over the probability mass function, over those values of x for which T x is equal to t. If I am assuming this factorization I can write it as g of T x theta into h x, x such that T x is equal t.

Now in the term h x this T x is equal t condition is not it is coming whereas, here T x is equal to t will be true for all the values. So, this term can be taken out of the summation sign, this can be written as g t theta sigma h x over x; such that, T x is equal to t. So, if I consider probability of X is equal to say x prime given T X is equal to t; that means, conditional distribution of X given T, this is equal to 0, if T x prime is not equal to t. And, in other case it is equal to probability of X is equal to x prime T X is equal to T x prime divided by probability of T X is equal to t which is actually equal to T x prime here, because I am taking T x prime is equal to t.

So, if you consider this now, probability X is equal to x prime T X is equal to T x prime it is same as probability X is equal to x prime. So, this becomes equal to, let me consider only this star portion; let me call this is a star here I will consider this portion.

(Refer Slide Time: 06:15)



So, this is star portion is equal to P theta X is equal to x prime divided by g t theta sigma h x x; such that, T x is equal to t; that is equal to g t theta h x prime divided by g t theta h x sigma x, such that T x is equal to t. Now this term cancel out. So, you look at this conditional distribution of X given T. Now this term is free from the parameter, independent of theta.

So, the conditional distribution of X given t is independent of the parameter. So, T is sufficient by the definition of the sufficiency. Let us take the converse part of this theorem. Let us assume that T is sufficient for theta. If we assume that t is sufficient of theta; that means, I am saying the conditional distribution of X given T X is a function of only x prime and t that is independent of theta.

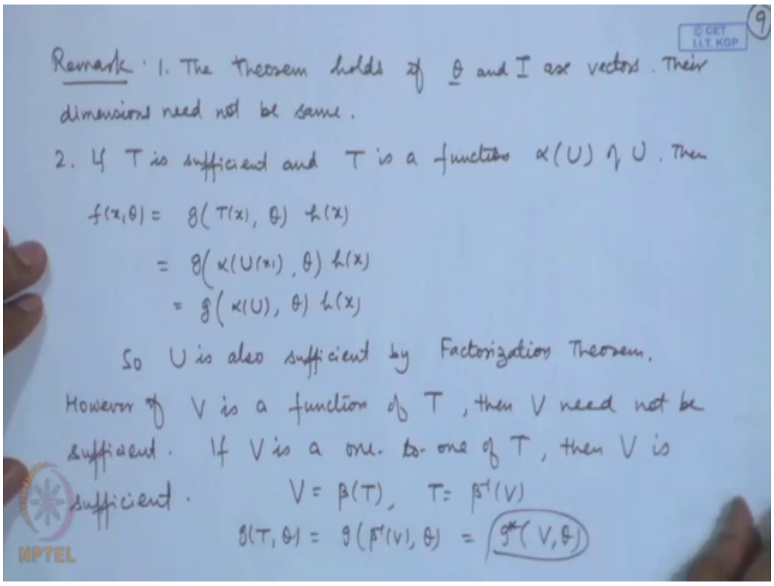
But this left hand side you can write as P theta X is equal to x prime T X is equal to T x prime divided by probability T X is equal to T x prime, that is equal to c x prime t, if T x prime is equal to t. In other case of force, it is equal to 0, so we do not write that that case here. This means that probability of X equal to x prime is equal to c x prime t and this term. Now what is this term? This term will be simply c x prime t g t theta because I am taking T x prime is equal to t.

Now this is nothing, but the factorization because this term I can write as h x g t theta. So, I have considered this discrete case here because it is easy to write this conditional probability. If the distribution are continues then probability of T X is equal to t does not make sense because that will be 0. So, we are use the conditional density function form. So, the general proof which is given in the Lehmann and Romano, this takes care of all these cases.

Another point which I would like to mention here, here I have taken theta to be a scalar, but suppose theta is a vector here then what will be the change here? If we make this assumption this theta will become vector, this theta will become vector here. Here also will become a vector, this will become a vector, this will become a vector. So, there is no change in the argument here that means, if the expectation holds t will be sufficient.

Let us look at the converse part. In the converse part, you are saying that this is free from theta, so that I will become vector here and it will not make any difference and here we will write it as a function of t and theta where theta is a vector parameter. So, this result holds even if theta is a vector parameter and another thing is about T also. I am writing here T as a one dimensional term, but that is also not must here T also can have several components like it could be T 1, T 2, T k, similarly theta can be theta 1, theta 2, theta n. So, the let me write this as a remark here.

(Refer Slide Time: 10:59)



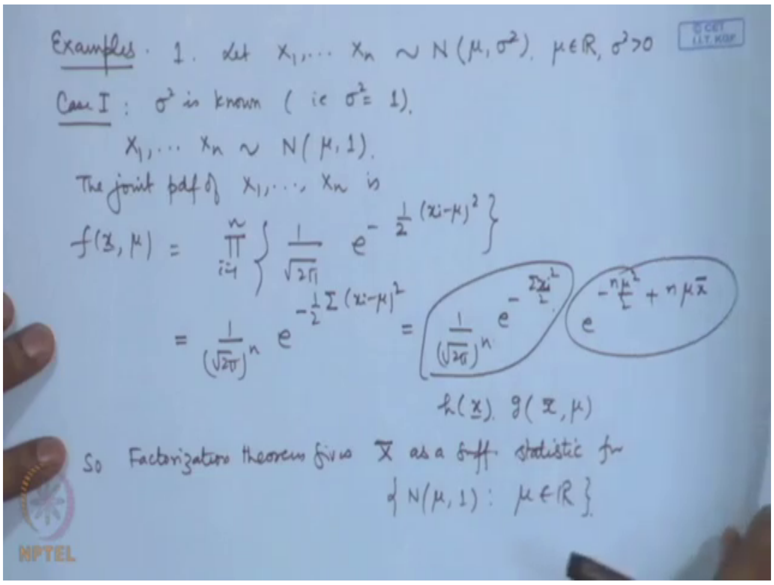
The theorem holds if theta and T are vectors and another point is that their dimensions need not be the same, their dimensions need not be same. Now let us revisit our statement. I said that if T is a function of U, then U is also sufficient. Now in the factorization theorem, if I substitute t as function of U, then I will be writing it is something like h of U. If I put that thing then it will mean that U is also sufficient by the same argument.

So, let me add that here. If T is sufficient and T is a function; say alpha U, then let us look at the density function, f of x theta is equal to g of T x theta into h x, this we can write as g of. Now T it is a function of alpha U So, this we can write as g of a function of U So, we can just alpha U we can write. So, U is also sufficient by factorization theorem; however, if V is a function of T, then V need not be sufficient. If V is a 1 to 1 function of T, then V is sufficient. This proof is again simple, if we say V is a 1 to 1 functions say, V is equal to beta of T, then we can say T is equal to beta inverse of V.

In that case g T theta you can write as g of beta inverse V theta; that means, it is a function of v and theta. So, V is also sufficient. Now the definition of sufficiency can be used to prove whether a given statistic is sufficient or not sufficient, because even find out the conditional distribution of a given statistics and you can see whether it is free from the parameter or not; however, you should know what statistic you are checking whereas, the factorization theorem yields a sufficient statistic because it is appearing there.

Now if you want to prove that something is not a sufficient statistic, then factorization theorem will not be useful, because to show that it is it cannot be represented is more difficult than saying that it is a function. So, both of the that is the original definition as well as the factorization theorem have different uses.

(Refer Slide Time: 15:15)



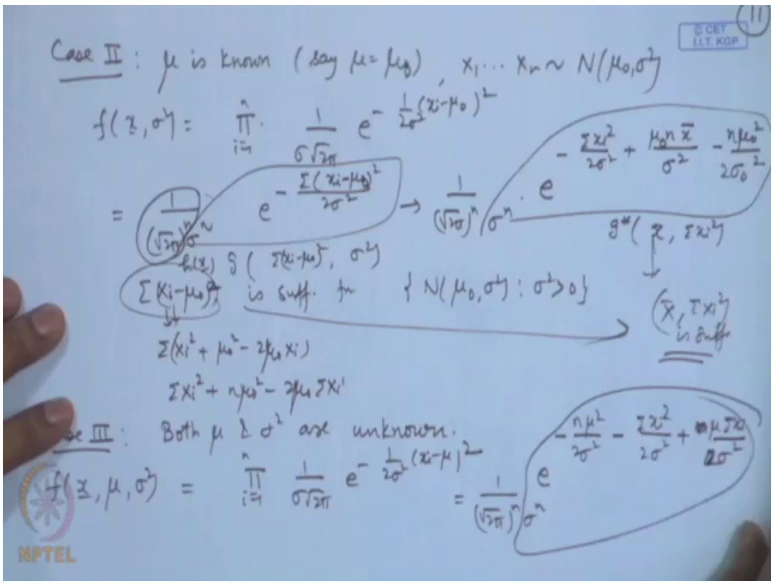
Let me give some examples here. So, let X 1 X 2 X n follows a normal distribution with mean mu and variance sigma square, I will consider different cases. As I mentioned to you that the sufficiency is a property of the family of distributions, it is not a property of a variable or a property of the parameter. It is a property which is holding for the family. So, here we are saying mu belongs to r sigma square is positive. Let us take special cases.

Suppose I say sigma I square is known, say sigma I square is equal to 1; that means, I am saying X 1 X 2 X n follows normal mu 1 distribution. Now let us write down the joint distribution of, joint distribution of X 1 X 2 X n. So, that is equal to product 1 by root 2 phi e to the power minus 1 by 2 x i minus mu square is equal to 1 to n.

So, this if you see 1 by root 2 phi to the power n, e to the power minus sigma 1 by 2 x minus mu square. This we can write as 1 by root 2 phi to the power n, e to the power minus sigma x i square by 2, e to the power minus n mu square by 2 plus n mu x bar, because if I take the cross product that is twice mu xi with the minus sign.

So, 2 2 will cancel out, minus minus will become plus. So, we get mu sigma xi that I write as n mu x bar. Now if you write this function h of x and this part we consider as a function of x bar and mu, then it is exactly in the form of factorization theorem. We have one term which is free from the parameter and other term which is depended up on the parameter, depends on the variable only through x bar. So, by factorization theorem gives x bar as a sufficient statistic. Now this family is normal distributions with variances known. So, here the sufficient statistics is x bar in a rough way we can say x bar is sufficient for mu.

(Refer Slide Time: 18:35)



Let us take the second case, I take mu is known. If mu is known say mu is equal to mu naught. In that case, the distribution of X 1 X 2 X n is normal mu naught sigma I square. So, the joint distribution of X 1 X 2 X n will become 1 by sigma root 2 phi e to the power minus 1 by 2 xi minus mu naught sigma I square, that is equal to 1 by root 2 phi to the power n sigma to the power n e to the power minus.

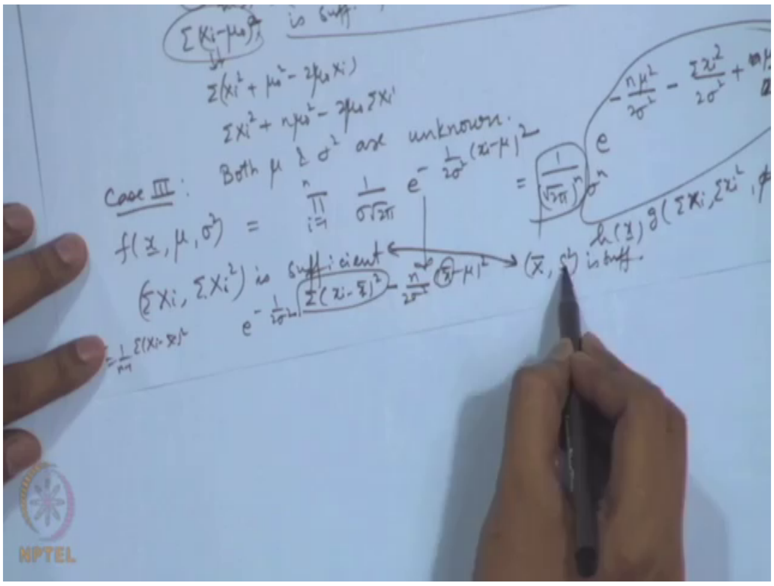
Now this term we write it as sigma xi minus mu naught square by twice sigma square. Here you see I cannot separate out xi like in the case of sigma known case. So, we can say here that sigma Xi this term I write as say h x and remaining term I write as g of sigma xi minus mu naught square and sigma square.

So, sigma Xi minus mu naught square is sufficient for the family of normal distributions. Now we may do this factorization in different way also. We may right here 1 by root 2 phi to the power n, sigma to the power n and as before let us expand this. So, we get minus sigma xi square by 2 sigma i square plus mu n x bar by, so this mu is mu naught here sigma square and e to the power minus n mu naught square by sigma twice sigma naught square.

If you look at this breakup, then I can consider this as a function of x bar and sigma xi square. So, we can also conclude that X bar and sigma Xi square is sufficient which is of course true here. But, if you see this one this is a larger reduction then this because here the sufficient statistic is two dimensional. Here you have sufficient statistic as one dimensional and of course, you can see here that this itself is a function of sigma Xi and sigma Xi square. Because, if I expand this I get sigma Xi square mu naught square minus 2 mu naught Xi. So, this is equal to sigma Xi square plus n mu naught square minus 2 mu naught sigma Xi, so this is actually a function of this. So, we will prefer this because this is a higher level of data reduction because this is one dimensional, this is a two dimensional.

Let us take the case where both mu and sigma square are unknown. Now notice here if both are unknown then I have to consider the joint distribution by treating both mu and sigma square as the parameters. So, this is a two dimensional parameter case here and the product of the individual distributions of X 1, X 2, X n it is equal to 1 by 2 sigma xi square xi minus mu square. You expand this, this is equal to 1 by root 2 to the power n sigma to the power n, e to the power minus n mu square by 2 sigma xi square minus sigma xi square by twice sigma square plus n mu or you can say mu sigma xi by sigma square.

(Refer Slide Time: 23:13)

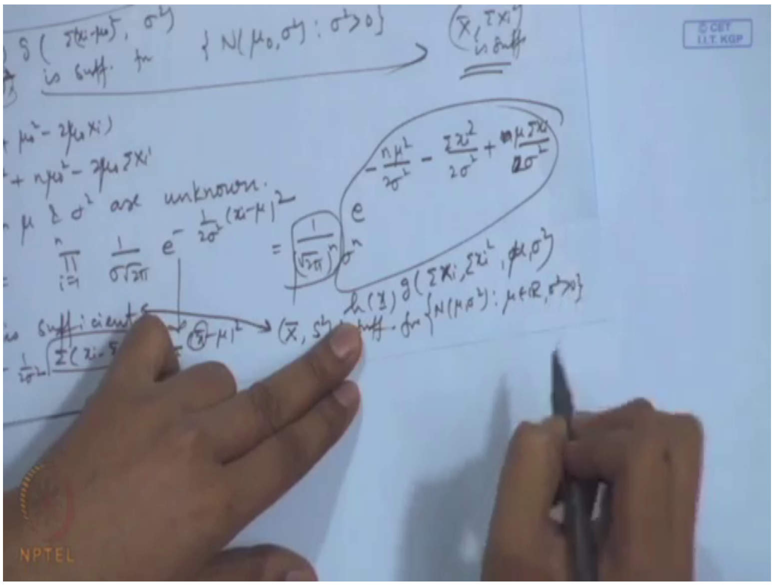


So, this term you can see if it is a function of this term is a function of sigma Xi, sigma xi square and mu and sigma square and this term you can call h x. So, we conclude that sigma Xi, sigma Xi square is sufficient. Here you can see that further reduction is not possible; however, we can consider it in a slightly in a different way as follows. We may write this as e to the power minus 1 by twice square sigma xi minus x bar whole square minus n by 2 sigma square x bar minus mu square; that means, I have added and subtracted x bar term here.

In that case this is actually sigma xi whole square this is x bar. So, we can conclude that X bar and s square, where we have used earlier the notation s square for 1 by n minus 1 sigma Xi minus X bar whole square that is the sample variance, so this is sufficient. Now you see there is no difference in this statement, if I considered x sigma Xi and Xi square then this is 1 to 1 function of X bar S square because, from here I can obtain this and from here I can obtain this.

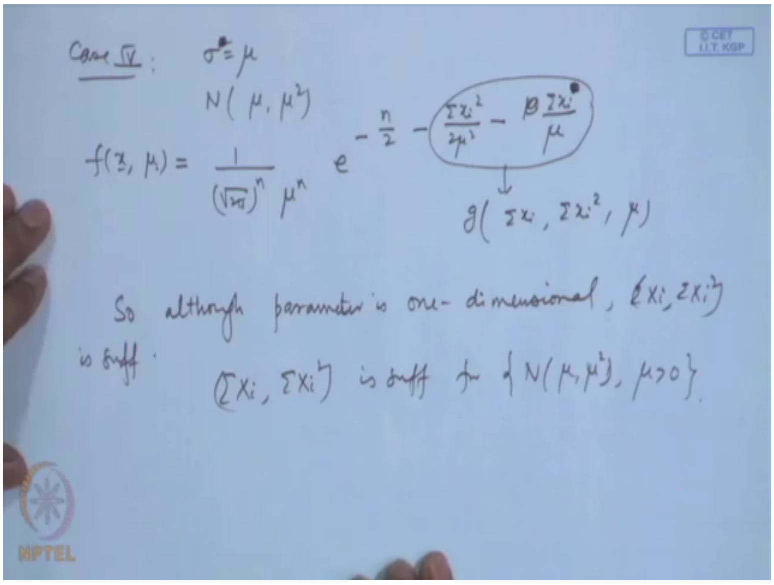
So, we can say that when the both parameters in a normal distribution are unknown, then the sample mean and the sample variance are sufficient. Now many times we are using it as a misnomer that, expiry sufficient for mu and S square is sufficient for sigma square.

(Refer Slide Time: 25:03)



Actually we have to say this is sufficient for the family normal mu sigma square, mu belonging to r and sigma square greater than 0. I will just explain this discrepancy may occur if we do not maintain this family here.

(Refer Slide Time: 25:19)



For example I take another case say sigma is equal to sigma square is equal to sigma is equal to say sigma is equal to mu, then what happens to the density, it is the distribution is normal mu mu square. If I have this then you can look at this breakup here, that joint density although it is a function of x and mu alone because sigma is vanishes here. This is equal to 1 by root 2 phi to the power n, mu to the power n, e to the power minus n by 2 minus sigma xi square by twice mu square minus mu sigma xi square by mu square that is mu.

So, here you see this is a function of sigma xi, sigma xi square and mu. Although the parameter is one dimensional, the sufficient statistics is two dimensional. So, although parameter is one dimensional, sigma Xi sigma Xi square is sufficient. So, again the statement is again the same, that is sigma Xi sigma Xi square is sufficient for this family normal mu mu square; of course, mu is greater than 0 here.

So, sufficiency is a property of the family of distributions. In the next lecture we will consider few more examples that is how to apply the factorization theorem to derive the various sufficient statistics. And, we will look at the maximal data reduction by means of sufficiency that is the concept of minimal sufficient statistics.

So, in the next lecture will be considering these concepts.