Statistical Inference

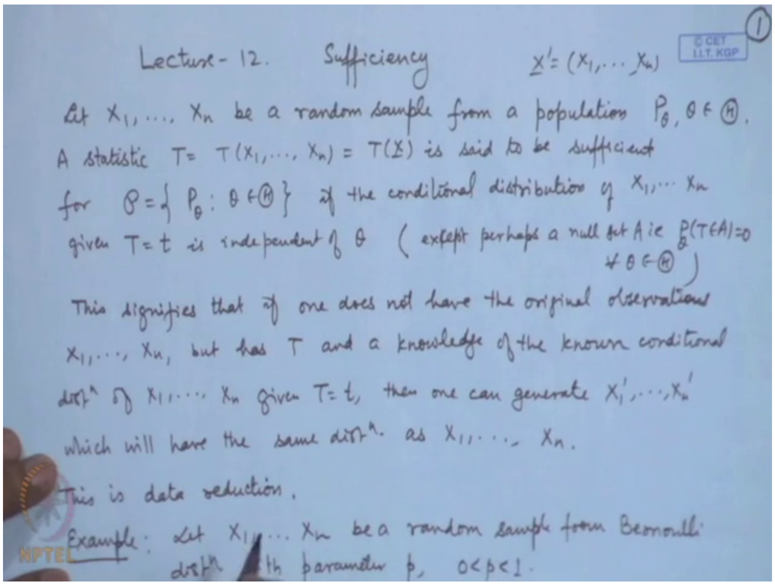
Prof. Somesh Kumar
Department of Mathematics
Indian Institute of Technology, Kharagpur

Lecture No. # 12
Sufficiency

Now, I start with a new concept that is called sufficiency. In the context ofstatistical inference, there is a concept which is useful to retain the necessary data without losing any information.What is the literal meaning of the word sufficiency? The literal meaning of the word sufficiency is that it is enough sufficient means enough. So, usually we are dealing with the statistical model that we deal in theinference problem is that, we say let X 1, X 2, X n, be a random sample meaning there by that we have data on n observations or you can say n data points are available to us.

Now, in many of the practical problems it becomes difficult to retain the data, because it may occupy lot of storage spacewhether it is on computer or it is in the form of hard copy of the data and then there is a danger of losing the data. It will be always interesting to say thatlet us keep the minimum things, such that whatever information or whatever useful inferences we want to make we are not suffering in that; that means, we do not want lose any important part of it.A formal specification of this concept is called sufficiency or sufficient statistic inthe context ofstatistical inference.

(Refer Slide Time: 02:10)



So, let us introduce the formal definition of sufficiency.As before, we have a random sample. So, let X 1, X 2, X n, be a random sample from a population say p theta, theta belonging to say script theta a statistic. So, statistic we have already defined a statistic means a function of observations. So, t that is T X 1, X 2, X n, which we also write as T X; that means, we are denoting x as X 1, X 2, X n. So, T x is said to besufficient. Now, what do you mean by sufficient for what. So, we usually mention the word sufficient for the family of probability distributions.

In loose terms we also say sufficient for the parameter theta meaning there by that, whatever be the parameter under consideration, many times in the problems we will have one dimensional parameter, two dimensional parameter etcetera.In that case we will have to consider specifically what parameter is being considered. So, the formal definition I am writing for the family of probability distributions meaning there by that, whatever parameters are under consideration this could be a scalar or a vector parameter. So, this is said to be sufficient if the conditionaldistribution of X 1, X 2, X n, given T is equal to say small t is independent of theta of course, except perhaps a null set a that is on a set a where T takes probability 0.

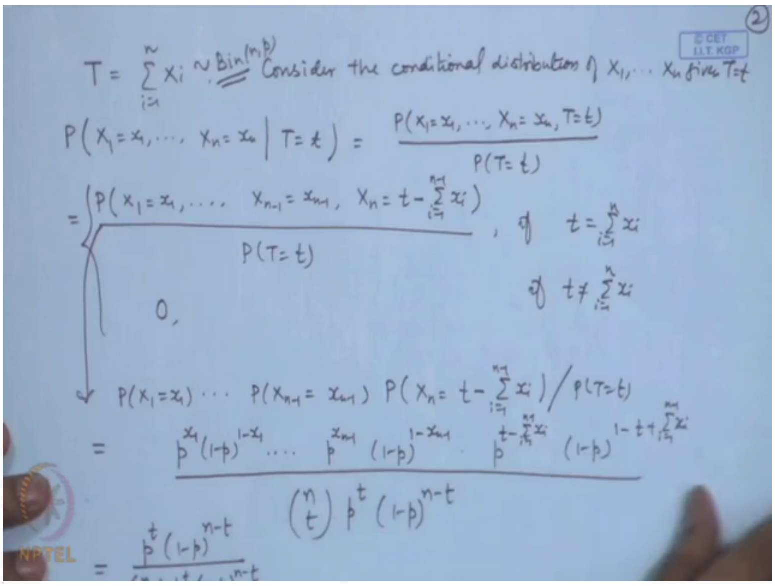
So, this is a exceptional case, but in general the distribution of the random sample given the statistic, if it is independent of the parameter then we say that this T is independent then we say that this T is a sufficient statistic. Now, what is the physical interpretation of this definition that, the distribution of X 1, X 2, X n, is free from theta and then we say it is sufficient.What does it mean?It means that now if the distribution is free from theta; that means, the distribution of X 1, X 2, X n, given T is completely known.

So, suppose we know T, we know the distribution of t.Now, this conditional distribution of X 1, X 2, X n, given T since it is free from theta then that is also known therefore, if I merge these 2 distributions that is a conditional distribution of X 1, X 2, X n, given T and the distribution of T, I get the joint distribution of X 1, X 2, X n, and T from there I get the distribution of X 1, X 2, X n. It means that even if I may not have the initial X 1, X 2, X n, with us, but we cangenerate that distribution once again, because of the information or you can say the distribution of X 1, X 2, X n, given T being free from the parameter and t is known to us.

This signifies that if one does not have the original observations X 1, X 2, X n, but has T and knowledge of the known conditional distribution of X 1. X 2, X n, given T then one can generate say X 1 prime, X 2 prime, X n prime, which will have the same distribution as X 1, X 2, X n. So, this is called data reduction. As we will show later on that in most of the practical problems, the sufficient statistics will become like one dimensional or two dimensional things although you may have any number of observations. So, this data reduction is helpful and we will show statistically also that, basing our decisions on the sufficient statistics isalso useful; that means, if there is any inference made in the terms of estimation testing of hypothesis etcetera.If I am making inference based on the sufficient statistics we are better off.

So, let me explain this example say binomial distribution example let me takesuppose, I have X 1, X 2, X n, be a random sample from say Bernoulli distribution with parameter p here p lies between 0 and 1.

(Refer Slide Time: 09:03)



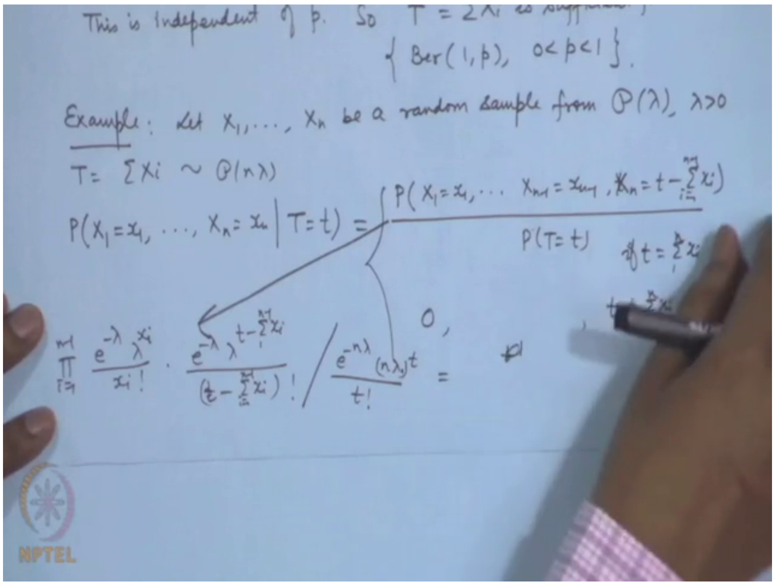
Let us consider say T is equal to sigma Xi, i is equal to 1 to n let us look at the conditional distribution of, consider the conditional distribution of X 1, X 2, X n, given T that is equal to X 1 is equal to X 1 and so on. X n is equal to X n given T is equal to t that is equal to probability of X 1 is equal to small X 1 and. So, on X n is equal to small X n T is equal to t divided by probability of T is equal to t. Now, that isequal to since, T is equal to sigma X i, if small X 1 plus small X 2 plus small X n is equal to t, then only this probability will becalculated in other cases this will be simply equal to 0. So, that is equal to probability of X 1 is equal to small X 1 and. So, on X n minus 1 is equal to small X n minus 1 and X n is equal to t minus sigma Xi, i is equal to 1to n minus 1.If t is equal to sigma X i is equal to 1to n otherwise this is 0.

Now, here we can make use of the fact that X 1, X 2, X n are independently distributed Bernoulli random variables. So, if they are independent this probability of the joint occurrence will be equal to the product of these probabilities. So, this term let me write this term is anyway 0. So, this term is equal to probability of X 1 is equal to small X 1 and. So, on X n minus 1 is equal to small X n minus 1, probability of X n is equal to t minus sigma X i, I is equal to 1to n minus 1 that is equal to p to the power X 1, 1 minus p to the power 1 minus X 1 and. So, on p to the power X n minus 1, 1 minus p to the power 1 minus X n minus 1 p to the power t minus sigma X I, I is equal to 1to n minus 1, 1 minus p to the power 1 minus t plus sigma X i, I is equal to 1to n minus 1 and divided by probability T is equal to t.

Now, what is the distribution of t, if X 1, X 2, X n are Bernoulli's independent then this will be binomial n p. So, probability T is equal to t that will be equal to n c t, p to the power t,1 minus p to the power n minus t. Now, we can easily see these terms this p to the power terms if you add you will get p to the power t similarly, if you add 1 minus p exponents you will get n minus sigma X i, that will cancel out with plus sigma X i, you get n minus t. So, you get it as p to the power tinto 1 minus p to the power n minus t divided by n c t, p to the power t into 1 minus p to the power n minus t.

Now, this term simply cancels out. So, we get it as 1 by n c t. So, this conditional distribution then.

(Refer Slide Time: 13:29)



We can express as probability of X 1 is equal to small X 1 and. So, on X n is equal to small X n given T is equal to t that is equal to 1 by n c t, for t is equal to sigma X i and it is equal to 0, if t is not equal to sigma x i. You look at this term there is no theta, no parameter appearing here, p is not appearing here. So, this isindependent of p. So, t is equal to sigma X i is sufficient for the family of Bernoulli distributionswe may also say it as that t is sufficient for p.Now, note here the physical significance of sufficiency.

If we are observing X 1, X 2, X n as p independent Bernoulli random variables; that means, their observations related to success or failure in n Bernoullian trials for example, you are looking at a game of say dart and we are consideringhitting a target and we make naims at the target, then what is important whether individual hits whether this say second one hit correctly, third one did not hit correctly is it important information or out of n total attempts how many are correct; that means, x that is some of x i's. Now, here you see in this concept of sufficiency exactly sigma X i is turning out to be sufficient therefore, this is the relevant information and whatever individual information about X 1, X 2, X n is there that is not necessary to be retained.

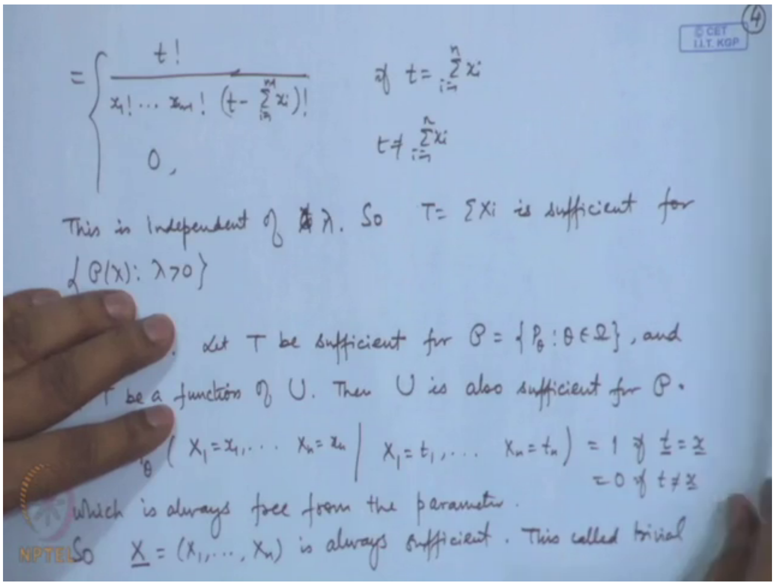
In fact, now if we know this and we know the distribution of t that is binomial n p.We can generate another random sample let us call it say X 1 prime, X 2 prime, X n prime, which will have Bernoulli 1 p distribution. Let me explain through another examplesay let X 1, X 2, X n be a random sample from Poisson lambda distribution, where lambda is positive.Once again let us define t is equal to say sigma X i. now, you can proceed in the same way like in the binomial case we can consider X 1 is equal to X 1 and. So, on X n is equal to X n given T is equal to t.

So, arguing as before we get it as X 1 is equal to X 1 and. So, on X n minus 1 is equal to X n minus 1, X n is equal to t minus sigma X i, I is equal to 1to n minus 1 divided by probability, T is equal to t, if t is equal to sigma X i, 1to n it is equal to 0 if t is not equal to sigma X i 1to n. So, once again this term will be equal to e to the power minus lambda, lambda to the power X i by X i factorial for I is equal to 1to n minus 1 and the last 1 is e to the power minus lambda, lambda to the power t minus sigma X i, 1to n minus 1 divided by t minus sigma X i, I is equal to 1to n minus 1 factorial.

Now, this willbecome e to the power minus n minus 1 lambda and then e to the power minus n lambda and also we have in the denominator t.Now, this will follow Poisson n lambda, because Poisson distribution is additive. So, if we are considering a random sample each one following Poisson lambda then sigma X i will follow Poisson n lambda.

So, we can write e to the power minus n lambda, n lambda to the power t by t factorial. So, this e to the power minus n lambda cancels out and if you look at lambda to the power X 1 plus, X 2 plus, X n minus 1, that cancels here you get lambda to the power t and in the denominator also we have lambda to the power t here so, what we get here this t factorial will go in the numerator.

(Refer Slide Time: 18:58)

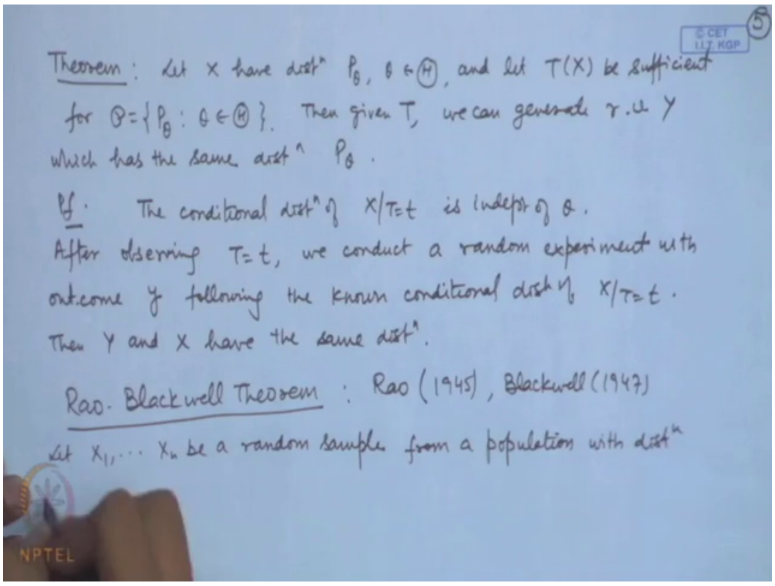


So, let me write it here, this is equal tot factorial divided by X 1 factorial, X 2 factorial X n minus 1 factorial, t minus sigma X i, I is equal to 1to n minus 1 factorial, if t is equal to sigma X i and it is equal to 0 if t is not equal to sigma X i, i is equal to 1to n. Once again you notice here that this is independent of t. So, t is equal to sigma X i this is independent of lambda sorry. So, t is equal to sigma X i is sufficient for the family of Poisson distributions, we may also saythat sigma X i is sufficient for the parameter lambda.

Now, we can make certain statements here if I am considering conditional distribution of X 1, X 2, X n given t and suppose t is a function of u, then if I consider the conditional distribution of X 1, X 2, X n given u then that will also be free from the parameter, because if that is not free from the parameter then the conditional distribution of X 1, X 2, X n given t will also not be free from the parameter. Therefore if t is sufficient and t is a function of u then u is also sufficient and of course, if I have a 1-to-one function of t, then that will also be sufficient. So, let me give some remarks here. Let t be sufficientfor a family of distributions and let t be a function of u then u is also sufficient for p.

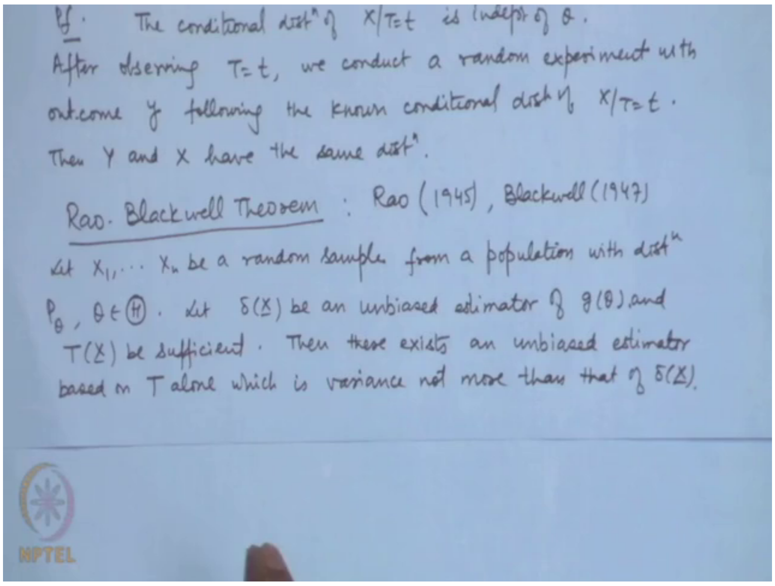
Another point that you notice here, that if I consider the conditional distribution of X 1, X 2, X n given X 1 is equal to small X 1, X 2 is equal small X 2, X n is equal to small X n, then that is always independent of parameter.We can write conditional distribution of say X 1 is equal to X 1, X n is equal to X n given say X 1 is equal to t 1 and. So, on X nis equal to t n this is equal to 1 if this t vector is same as x vector otherwise it is 0. So, this is naturally free from the parameterfree from the parameter. So, the sample x is always sufficient. So, the full sample is always sufficient, we will be interested in getting some sort of reduction over there.This is known as trivialsufficient statistics, trivial sufficient statistics.

(Refer Slide Time: 23:21)



Let me formally prove that given a sufficient statistics, you can generate the original sample. So, let x have distribution says p theta, theta belonging to say script theta.

(Refer Slide Time: 23:21)

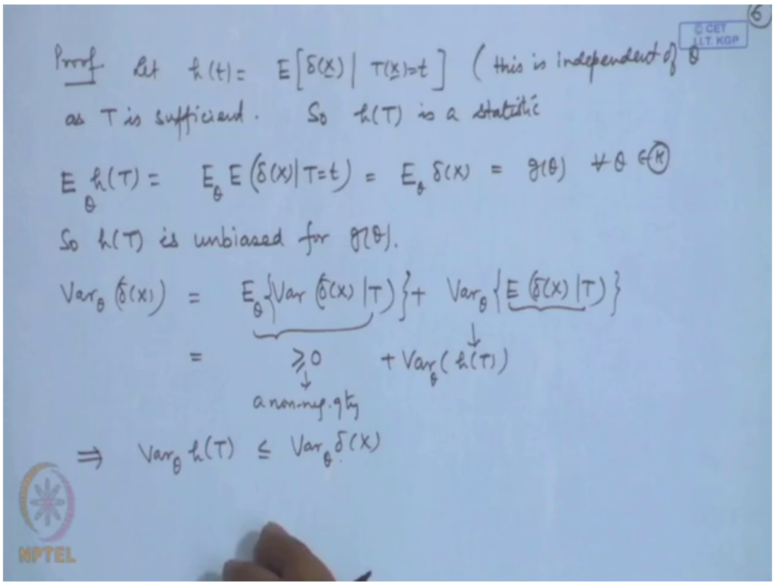


And let T X be sufficient, then given T we can generate random variable y which has the same distribution p theta, that is the same distribution of x. So, the conditional distribution of x given T is independent of theta. So, after observing T is equal to t, we conduct a random experiment with outcome sayy following the known conditional distribution of x given t and then y and x have the same distribution.

Now, another important significance you can say of sufficient statistics is that, if we are considering any one by the estimator I can have another one by estimator which is based on the sufficient statistics and its variance will be less than or equal to the variance of the initial estimator.This famous result is known as Rao Blackwell theorem; it isnamed after the Indian statiscian c r Rao, who proved this result in 1945 and David Blackwell 1947. Let X 1, X 2, X n be a random sample from a population with distribution p theta, theta belonging to say script theta let delta x b an unbiased estimator of parametric function say g theta and t be sufficient, then there exists an unbiased estimator based on t alone which has variance not more than that of delta x.

Now, this is a very significant statement in a given problem, if I have a sufficient statistics then I can always base our unbiased estimators on that statistics. So, that I will do better than if I do not base it; that means, I will be utilizing the full information in the sample for making my statistical inference.

(Refer Slide Time: 28:11)



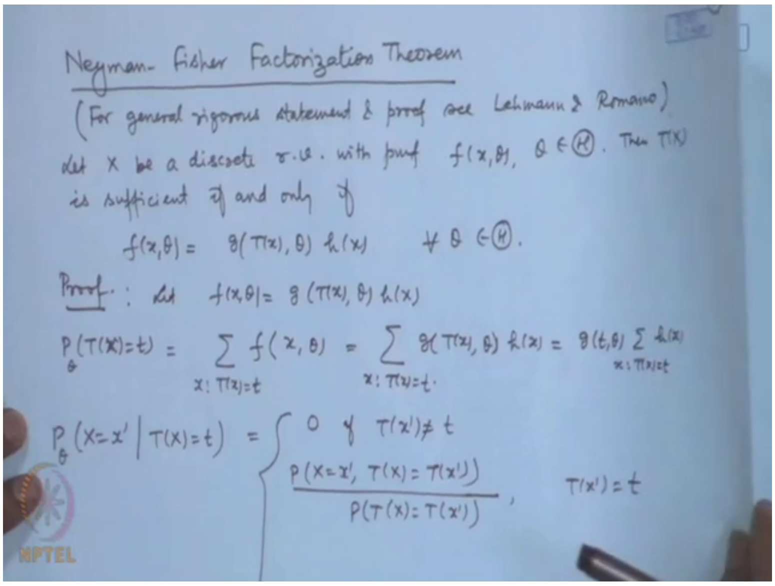
The proof is. In fact, not very difficult let us consider h t to be expectation of delta x, given T x is equal to T since, we know the conditional distribution of x given T is independent of theta therefore, this expectation is going to be a function of t alonethis is independent of theta as T is sufficient. So, h T is a statistic and I can consider it for my estimation purpose. Let us consider expectation of h T now, expectation of h T is simply expectation of expectation delta x given T.Now, this is nothing, but expectation of delta x that is equal to g theta. So, this new estimator that I have used h T is unbiased. So, this is unbiased for g theta.

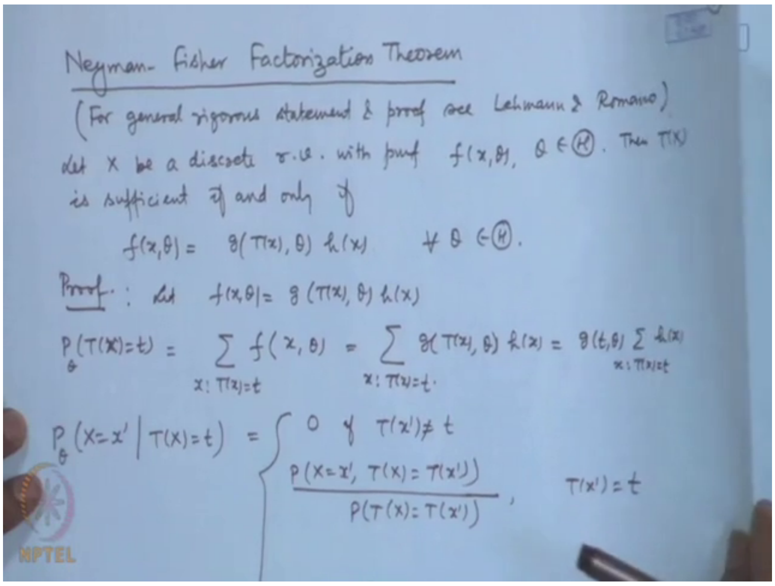
Further let us consider say variance of delta X. Now, this variance of delta X I can express as expectation of variance delta x given t, plus variance of expectation delta x given T. Now, this is equal to this quantity if you see this is a non-negative quantity and expectation of delta x given T we have defined to be h T. So, this is equal to variance of h T. So, what we are getting variance of delta x is equal to variance of h T plus a non-negative quantity; that means, variance of h T is going to be less than or equal to variance of delta x, you will give applications of this result a little letter.

Another important pointthat we notice when we were proving that t is equal to sigma X i is sufficient in the two examples that I have considered 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 I have given statistic sufficient may be too cumbersome 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 X i 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 an important result called factorization theorem, which readily producesthe sufficient statistics.

(Refer Slide Time: 32:08)





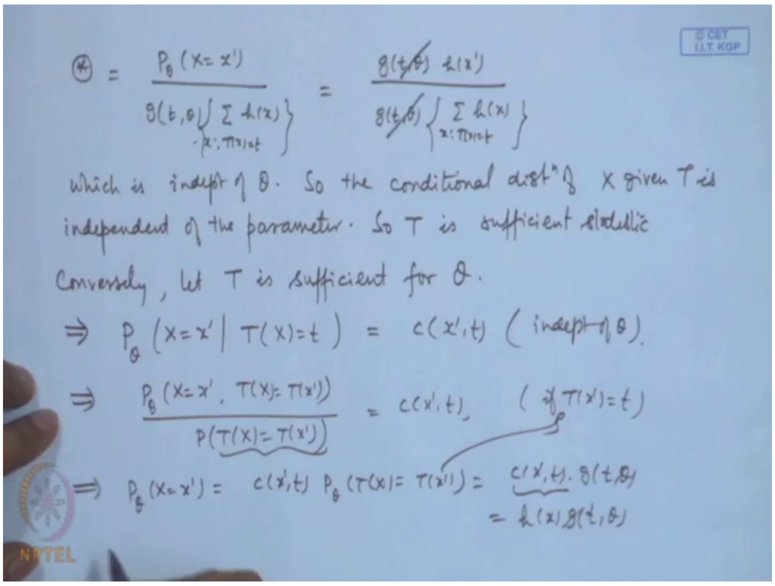
So, this is known as Neyman fisher factorization theorem named after r a fisher and Jersey Neyman who proved in around 1939. We are not going to give a very rigorous statement and proof of this theorem which will be applicable to all situations rather, we will consider a discreet case hereandto write the proof here, for general rigorous statement and proof see the book of say Lehmann and Romano. We are considering a discreet case here, let x be a discreet random variable with probability mass function say f x theta, theta belonging to script theta then T x is sufficientif 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 two terms g and h.Where h is a term where parameter does not appear in the term gthe parameter 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 considering the discreet 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 to t now in the term h x this T X equal to T condition is notit 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 of x equal to x prime, T X equal to T X prime, it is same as probability x equal to x prime. So, this becomes equal to let me consider only this star portion let me call this as a star here I will consider this portion.

(Refer Slide Time: 36:56)



So, this is star portion is equal to p theta x is equal tox 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 cancels out. So, you look at this conditional distribution ofx 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.

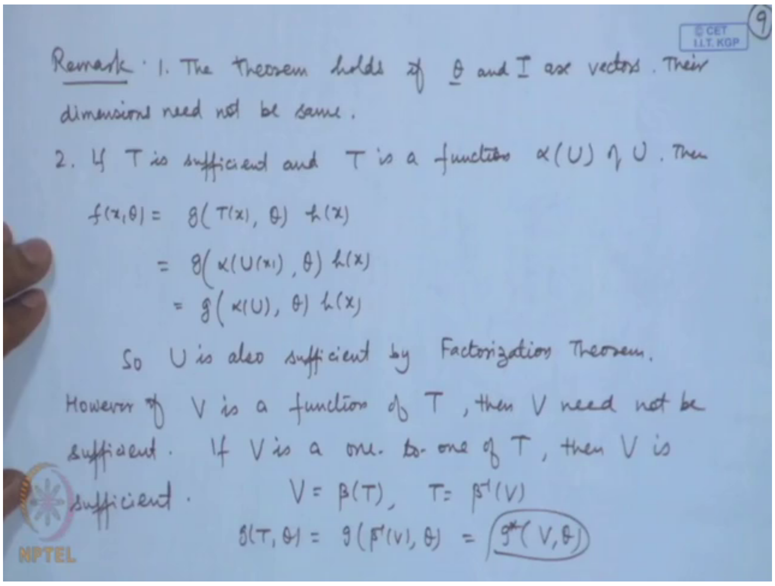
Let us take the converse part of this theorem, let us assume that t is sufficient for theta. If we assume that T is sufficient for theta; that means, I am saying the conditional distribution of x given T x is a function of only x prime and t, that is independent of theta. But this left hand side you 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 course, it is equal to 0. So, we do not write the 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 thisdiscreet case here, because it is easy to write this conditional probabilities, if the distributions are continuous then probability of T X equal to t does not make sense, because that will be 0 we have to 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 theta 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 factorization holds T will be sufficient. Let us look at the converse part.In the converse part we are saying that this is free from theta. So, theta will become vectorhere 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 1 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.

(Refer Slide Time: 41:40)



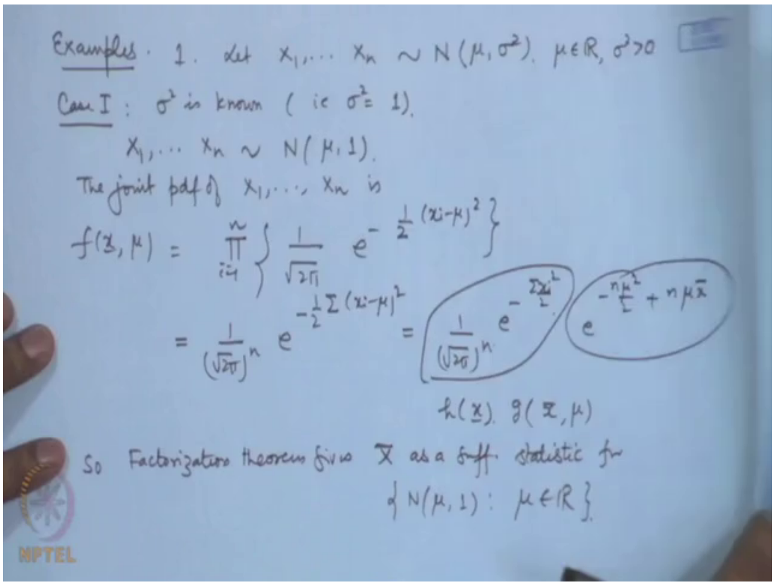
So, thelet me write this as a remark here, the theorem holds if theta and t are vectors and another point is that theirdimensions need not be the same. Now, let usrevisit our statement I said that if T is a function of u then u is also sufficient.Now, in the factorization theorem if I substitute t as a function of u, then I will be writing it as 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 of 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 one-to-one function of T then V is sufficient.Now, this proof is again simple, if we say V is a one-to-one function 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 we can find out the conditional distribution of a given statistic 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 yield say sufficient statistic, because it is appearing there now, if we want to prove that something is not a sufficient statistics, 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 that is the original definition as well as the factorization theorem have different uses.

(Refer Slide Time: 45:56)

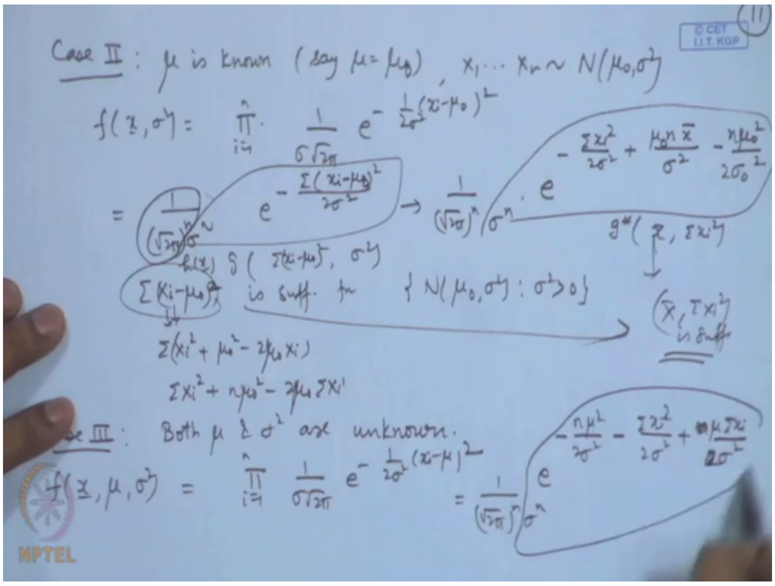


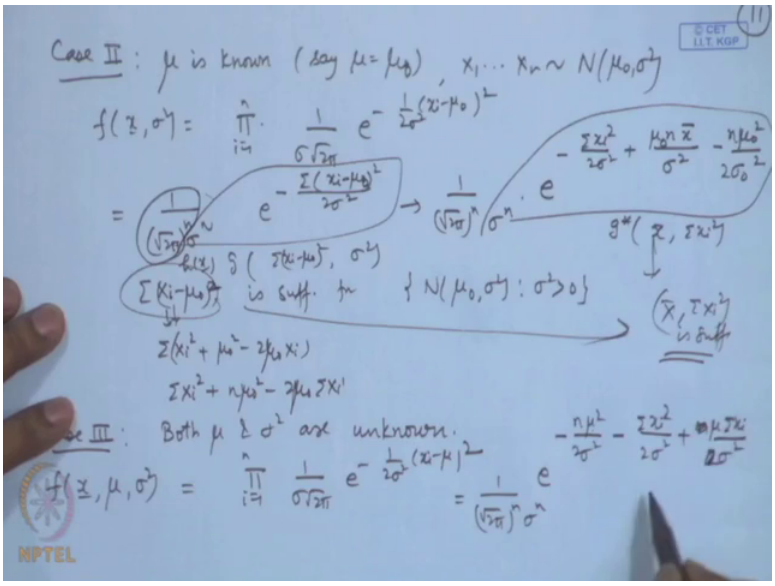
Let me give some examples here. So, let X 1, X 2, X n follow say normal distribution with mean mu and variance sigma square I will consider different cases.As I mentioned to you that the sufficiency is a property of the family of distribution it is not aproperty of a variable or a property of the parameter,it 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 these special cases suppose I say sigma square is known say sigma 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 ofjointdistribution of X 1, X 2, X n. So, that is equal to product 1 by root 2 pi e to the power minus 1 by 2, X i minus mu is square I is equal to 1to n. So, this if you see 1 by root 2 pi 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 pi 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 here if I take the cross product term that is twice mu X i with aminus sign. So, two will cancel out minus minus will become plus. So, we get mu sigma X i that I write as n mu X bar.

Now, if you write this function as 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 the other term which is dependent upon the parameter depends on the variable only through X bar. So, by factorization theoremgives 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: 49:14)





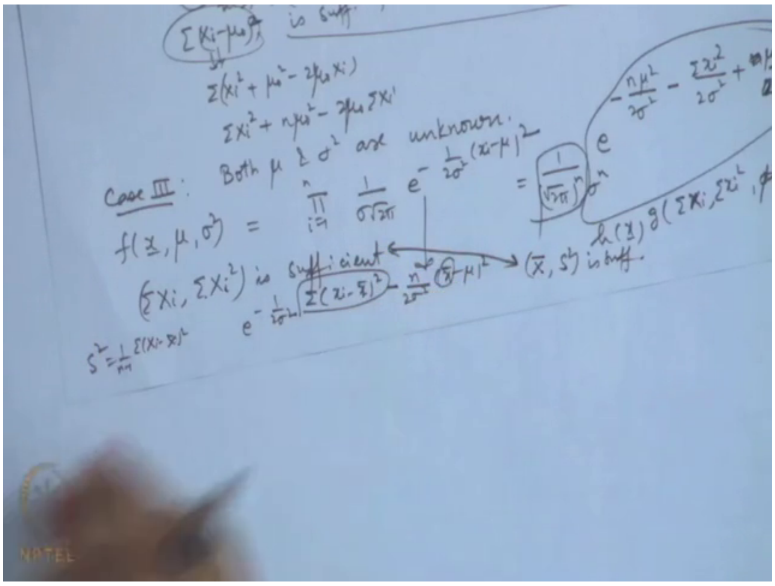
Let us take the second case, I take mu is known if mu is known say mu is equal to mu naught in that case the distribution of X 1, X 2, X n is normal mu naught sigma square. So, the joint distribution of X 1, X 2, X n will become 1 by sigma root 2 pi, e to the power minus 1 by 2 X i minus mu naught sigmasquare that is equal to 1 by root 2, pi to the power n sigma to the power n, e to the power minus now, this term we write as sigma X i minus mu naught square by twice sigma square. Here you see I cannot separate outx i's like in the case of sigma known case. So, we can say here that sigma X I, this term I write as say h x and remaining term I write as g of sigma X i, minus mu naught square and sigma square. So, sigma X i minus mu naught square is sufficient for the family of normaldistributions.

Now, we may do this factorization in different way also we may write here 1 by root 2, pi to the power n, sigma to the power n and as before let us expand this. So, we get minus sigma X i square by 2 sigma 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 break up then I can consider this as a function of X bar and sigma X i square. So, we can also conclude that X bar and sigma X i square is sufficient which is of course, true here, but if you see this 1 this is a larger reduction than 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 X i and sigma X i square, because if I expand this I get sigma X i square plus mu naught square minus 2 mu naught x i.

So, this is equal to sigma X i square plus n mu naught square minus 2mu naught sigma x i. 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 bothmu and sigma square as the parameter. 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 square X i minus mu square.

You expand this is equal to 1 by root 2, pi to the power n, sigma to the power n, e to the power minus n, mu square by 2 sigma square minus sigma X i square by twice sigma square plus n mu or you can say mu sigma X i by sigma square.

(Refer Slide Time: 49:14)



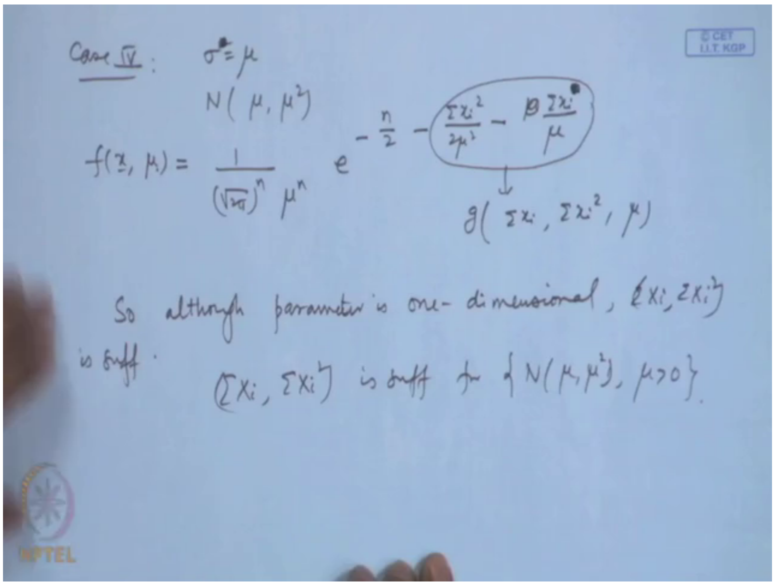
So, this term you can see, it is a function of this term is a function of sigma X I, sigma X i square and mu and sigma squareand this term you can call h x. So, we conclude that sigma X i, sigma X i square is sufficient. Here you can see that a further reduction is not possible; however, we can consider it in a slightly different way as follows, we may write this as e to the power, minus 1 by twice sigma square sigma X i 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 X i minus X bar whole square this is X bar. So, you can conclude that X bar and s square, where we have used earlier the notation s square for 1 by n minus 1 sigma, X i minus X bar Whole Squarethat is the sample variance. So, this is sufficient. Now, you see there is no discrepancy in this statement, if I consider x sigma X i and sigma X i square then this is one-to-one function of X bar and s square, because from here I can obtain this and from here I can obtain this. So, we consider that when both the 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 X bar is sufficient for mu and s square is sufficient for sigma square 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: 56:00)



For example, I takeanother case say sigma is equal to sigma square 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 break up here that joint density although,it is a function of x and mu alone, because sigma is vanished here this is equal to 1 by root 2, pi to the power n, mu to the power n, e to the power minus n by 2 minus sigma X i square by twice mu square minus mu sigma X i square pi mu square that is mu. So, here you see this is a function of sigma X i, sigma X i square and mu although the parameter is 1 dimensional the sufficient statistics is 2 dimensional.

So, although parameter is 1 dimensional sigma X i, sigma X i square is sufficient. So, again the statement is again the same that is sigma X i, sigma X i square is sufficient for this family normal mu mu square of course, you may put mu 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 are, how to apply the factorization theorem.To derive thevarious 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 we will be considering these concepts.