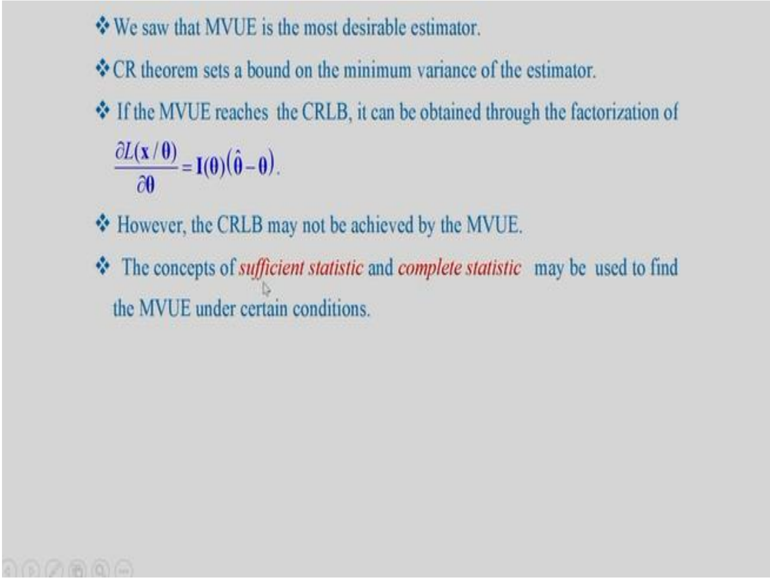
Statistical Signal Processing
Prof. Prabin Kumar Bora
Department of Electronics & Electrical Engineering
Indian Institute of Technology, Guwahati

Lecture No 11
MVUE through Sufficient Statistic

Hello students, welcome to lecture 12 and MVUE through sufficient statistic, miniMuem variance unbiased estimator through sufficient statistic.

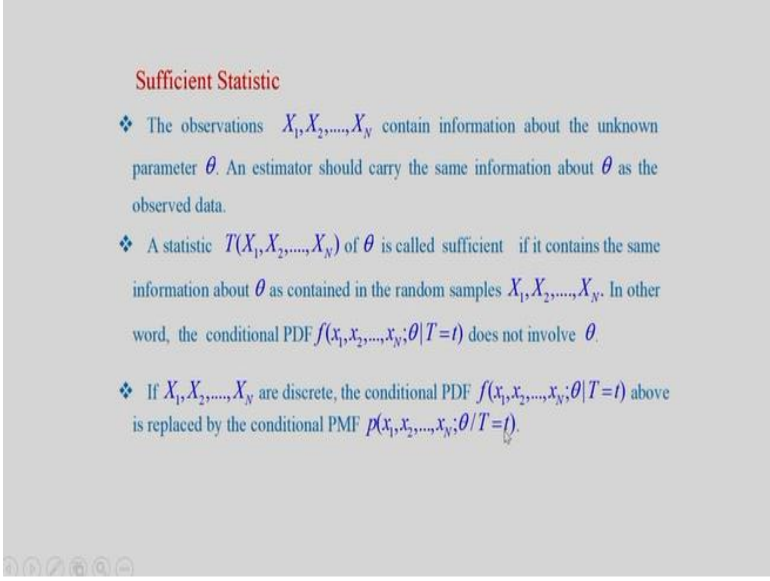
(Refer Slide Time: 00:48)



We saw that MVUE is the most desirable estimator; Cramer Rao theorem sets a bound on the minimum variance of the unbiased estimator. If the MVUE reaches the CRLB, it can be obtained through the factorization of this matrix log likelihood matrix del L del theta matrix that is equal to I theta into theta head minus theta, where theta is the parameter factor theta head is the unbiased estimator factor and I theta is the Fisher information statistics.

However CRLB may not be achieved by the MVUE. So in that situation the concepts of sufficient statistic and complete statistic may be used to find the MVUE under certain conditions. So we will introduce the concepts of sufficient statistic and complete statistic in this lecture.

(Refer Slide Time: 01:57)



We will start with sufficient statistic; the observations X1, X2 up to XN contain information about the unknown parameter theta. So because of that we will be able to estimate theta from the observations. An estimator should carry the same information about theta as the observed theta, whatever information is there for theta same should be carried by the estimator also. Now we will go to sufficient statistic, a statistics T X1, X2 up to XN of theta is called sufficient.

If it contains the same information about theta as containing the random samples X1, X2 up to XN. In other word, the conditional PDF f x1, x2 up to xN as a function of theta given T is equal to t does not involve theta. So this conditional PDF does not involve theta, then we will say that this statistics T X1, X2 up to XN is sufficient and now in the case of discrete random variables X1, X2 up to XN, the conditional PDF above is replaced by the conditional PMF.

So that means in the case of discrete random variable X1, X2 up to XN this statistic T X1, X2 up to XN of theta is sufficient. If this conditional PMF, conditional PMF of the theta given T is equal to small t, does not involve the parameter theta. So we have defined and this sufficient statistic

(Refer Slide Time: 03:58)

Example 1:

- Suppose $X_i \sim N(\mu,1), i=1,2$ and $T(X_1,X_2)=X_1+X_2$. Then,

$$f(x_1,x_2;\mu \mid T(X_1,X_2)=x_1+x_2)=\frac{f(x_1,x_2,x_1+x_2;\mu)}{f(x_1+x_2;\mu)}$$

$$=\frac{f(x_1,x_2;\mu)}{f(x_1+x_2;\mu)} \qquad X_1+X_2 \sim N(2\mu, 2)$$

$$=\frac{\frac{1}{2\pi}e^{-\frac{1}{2}\{(x_1-\mu)^2+(x_2-\mu)^2\}}}{\frac{1}{\sqrt{4\pi}}e^{-\frac{1}{4}(x_1+x_2-2\mu)^2}}$$

$$=\frac{1}{\sqrt{\pi}}e^{-\frac{1}{2}(x_1^2+x_2^2+2\mu^2-2\mu(x_1+x_2))+\frac{1}{4}(x_1^2+x_2^2+4\mu^2-4\mu(x_1+x_2)-2x_1x_2)}$$

$$=\frac{1}{\sqrt{\pi}}e^{-\frac{1}{4}(x_1^2+x_2^2+2x_1x_2)}$$

$$=\frac{1}{\sqrt{\pi}}e^{-\frac{1}{4}(x_1+x_2)^2}$$

And let us see an example; suppose XI is normally distributed with mean Mue and variance 1 for i is equal to 1 to 2 and T X1, X2 is equal to X1+ X2. We will examine if T X1, X2 is a sufficient statistics. Now the conditional PDF f x1, x2 as a function of Mue given T X1, X2 is equal to small x1+ x2 is equal to that is the joint PDF f x1, x2, x1+ x2; as a function of Mue divided by the marginal PDF that is f x1+ x2; as a function of Mue.

So this is the conditional PDF in terms of the joint PDF and marginal PDF. Now the same can be written as f x1, x2; Muee as a function of Muee divided by F x1 + x2; as a function of Mue, because x1, x2 and this x1+ x2 is a function of Muee. Therefore, the PDF will be same as the PDF of x1, x2. So that way this expression is equal to a joint PDF of x1, x2 as a function of Muee divided by PDF of x1 x2 as a function of Muee.

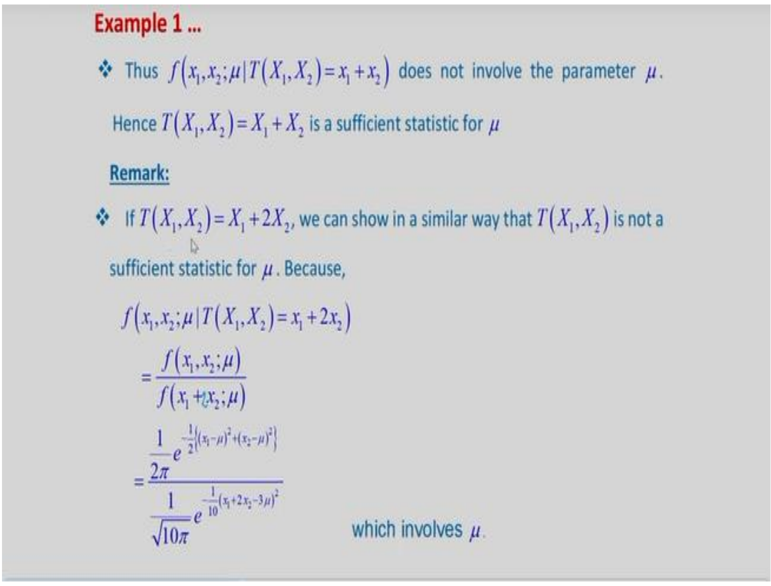
Now, since Xi is the iid Gaussian function; therefore joint PDF will be the product of the individual PDF. So that way f of x1 into f of x2; we can write like this, so 1 by 2 pi into e to the power minus 1/2 x1 minus Muee whole square plus x2 minus Muee whole square. So this is the joint PDF divided by marginal PDF that is the PDF of x1+x2. Now I know that x1 is normally distributed with mean Muee and variance 1 and x2 is also normally distributed with mean Mueeand variance 1.

Therefore, X 1 plus X 2 will be distributed as normal mean will be mean of X1 plus mean of X2, that is 2 Mue and variance of X1 plus variance of X2; that will be is equal to 2. So that way this will be normal with mean 2 Mue and variance 2. So therefore this marginal PDF of

X1 + X2 will be given by this; now if we simplify, I will get this expression so 1 by root Pi will be only there and bringing this to the numerator we will get like this.

And ultimately if we simplify this expression we will get the same equal to 1 by root pi into e to the power of ¼ into X1 square plus X2 square plus 2 X1, X2. So here we see that Mue is not involved, so this conditional PDF does not involve Mue therefore and these statistics X1+ X2 is a sufficient statistic.

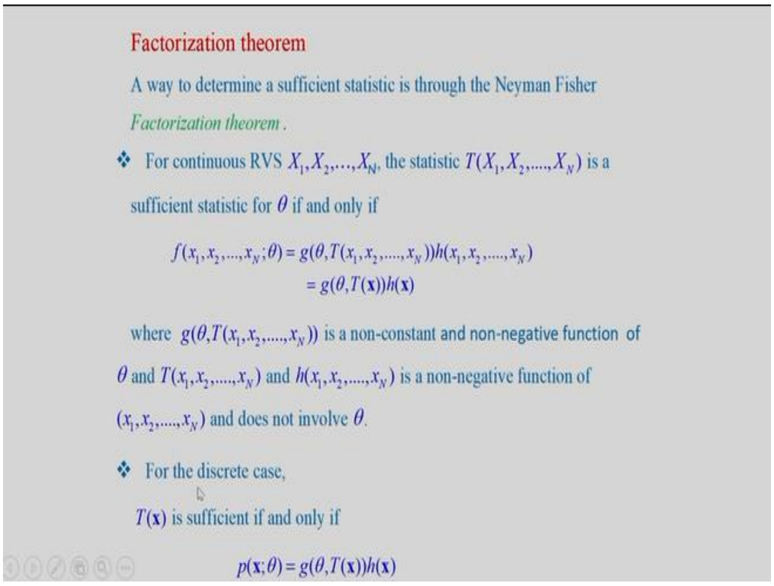
(Refer Slide Time: 07:38)



So we see that this conditional PDF does not involve any parameter Mue. Hence T X1, X2 is equal to X1+ X2 is a sufficient statistics for Mue. We will see one example for a statistic which is not sufficient; suppose we have this T X1, X2 is equal to X1+ 2 X 2. Earlier, we will consider X1+X2; now X1+ 2 X2, now this conditional PDF f x1,x2 as a function of Mue given that T X1, X2 is equal to small X1+ 2 X2.

That we can write as joint PDF of X1, X2 as a function of Mue divided by marginal PDF at point X1+ 2 X2 as a function of Mue. So if we write the Gaussian PDF for these two then we will get numerator as this and denominator as this. Here, we see that from these two expressions we cannot cancel out Mue; therefore this will involve Mue. So therefore this conditional PDF given this a statistic is will involve Mue therefore it is not a sufficient statistic.

(Refer Slide Time: 09:11)



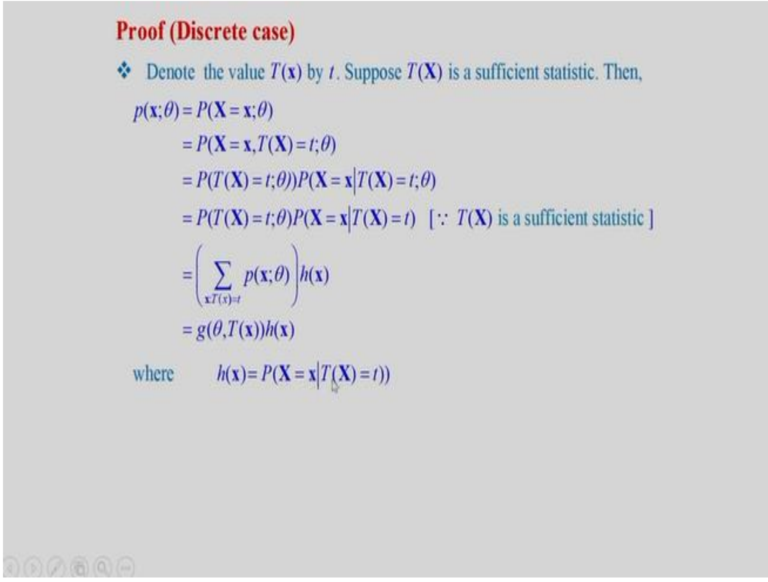
And now we will discuss one important theorem, what is known as the factorization theorem; that is Neyman Fisher factorization theorem. This is one of the ways to determine whether statistic is sufficient or not. For continuous random variables X1, X2 up to XN, this statistic T X1, X2 up to XN is the sufficient statistic for theta if and only if; this joint PDF as a function of theta is product of two functions g and h.

And g is a function of the parameter theta and the statistic TX and h is a function of data only, it is a function of X; is X1, X2 up to XN and it does not involve any theta. So that way we factorize the joint PDF in terms of two factor, where g theta TX is a non constant it should be a proper function of X1, X2 excetra and non-negative function of theta and T x1, x2 up to xN and h is x1, x2 up to xN is a non-negative function of X1, X2 up to xN and it does not involve in any parameter theta.

So then we say that, this statistic will be sufficient. So that way we have to see the joint PDF if it is a product of two factors; one factor is theta into T, other factor is a function of hX only then this statistic will be sufficient. For discrete case, we will be considering the condition in terms of joint PMF therefore TX is sufficient if and only if, this joint PMF px theta as a function of theta is product of g theta T X into hX.

Where g and h are defined as earlier, so we have introduced the factorization theorem we will try to prove this

(Refer Slide Time: 11:39)



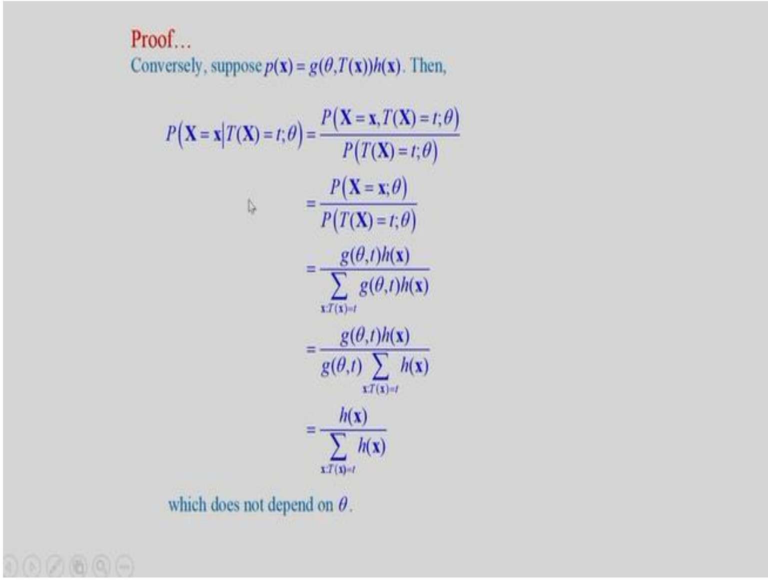
Considering the discrete random variables X1, X2 up to XN, let us denote the particular value of TX by a small t; that is small t is equal to T of small x1, small x2 up to small xn. Suppose T X is sufficient statistic. Then, PMF probability mass function of x as a function of theta that is equal to probability of X is equal to x as a function of theta. Now if we introduce any function T X then this joint probability will be also same.

Therefore this probability is same as probability of X is equal to small x and T X is equal to t as a function of theta. Now this I can write as probability of TX is equal to t of course as a function of theta and then into conditional probability. Probability of X is equal to small x given that TX is equal to t and this will also be a function of theta. So this is by applying the conditional probability result.

Now I know that, TX is a sufficient statistics; therefore this factor will not involve any theta. so I can write this term simply as probability of X is equal to small x given that TX is equal to t because it does not involve theta x TX is a sufficient statistic. So this term, now I can expand the in terms of the joint PMF summation of joint PMF for those value of x for which TX is equal to t into this part.

Now this is a function which does not involve any parameter or the statistics it is simply a function of X. So that way we have establish that this joint PMF is product of g and h, where g is a function of theta and TX and h is a function of X only and it is given by this probability of X is equal to small x, given that TX is equal to t, so we have proved this part.

(Refer Slide Time: 14:14)

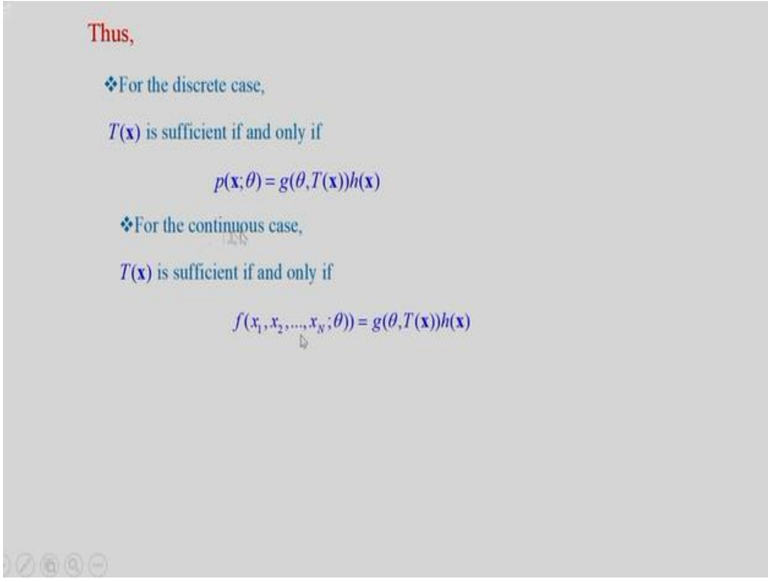


Now let us see the converse part, suppose px is equal to g of theta TX into hX. This is true and then we have to show that TX is a sufficient statistics. Now, this conditional PMF probability that X is equal to small x given that TX is equal to t as a function of theta. Now, you apply the definition of conditional probability; that is probability of X is equal to small x and this comma is for n TX is a function of theta divided by the marginal probability, probability that TX is equal to t as a function of theta.

Now since T X is a function of X this we can write as probability of X is equal to small x theta divided by probability of TX equal to t as a function of theta. Now we will apply this factorization result, so we will get g of theta t into hX divided by summation g of t into hX. X subset TX is equal to t. So this g of theta T will become 1 for all, so we are taking common here and therefore this denominator will be g of theta T into summation a h X, such that TX is equal to t.

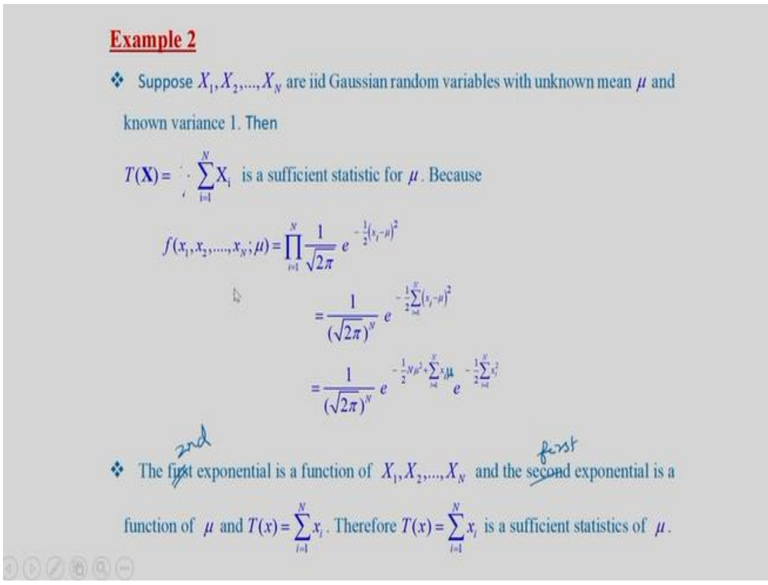
So this g of theta T and this g of theta T will get concerned therefore it will be simply hX divided by summation a h X, such that TX is equal to t and clearly this expression that is this conditional PMF does not involve any parameter theta, therefore TX is a sufficient statistic. Therefore we have proof for both necessary and sufficient conditions.

(Refer Slide Time: 16:22)



So for discrete case TX is sufficient if and only if this joint PMF is a product of two function g and h and similarly for the continuous case also we can establish that TX is sufficient if and only if the joint PDF is product of two factors; one is function of theta and TX, other is simply a function of X.

(Refer Slide Time: 16:50)

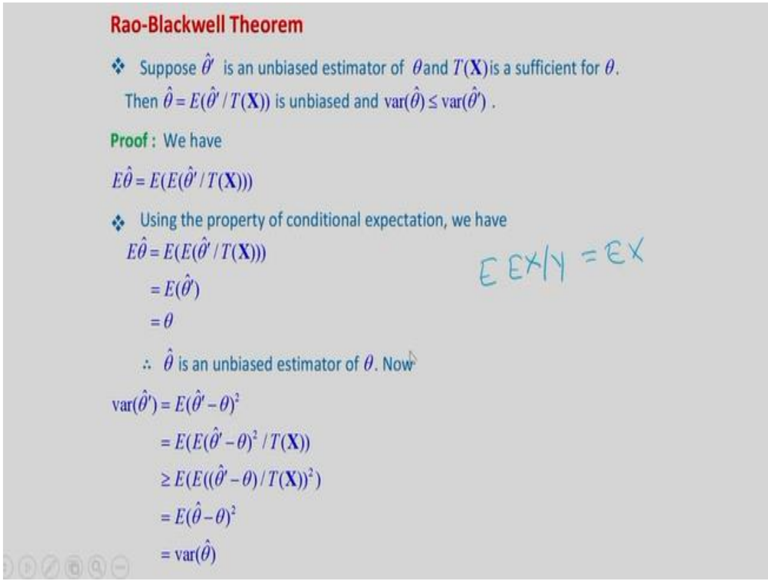


We will consider an example; suppose X1, X2 up to XN are iid Gaussian random variables with unknown mean Mue and known variance 1. Then, T X is equal to summation Xi, i going from 1 to N is a sufficient statistic for Mue because we will write this now joint PDF as a function of Mue, this is product of the individual PDF marginal PDFs and all are iid's; so same mean and same variance 1.

And this product we are writing as summation and this now we can expand Xi minus Mue whole square, so that we will get two factors here, now one factor is this and the other factor is this. We observed that, this second factor this is a function of Xi only and the first factor is function of both this statistic and Mue. Therefore, we conclude that TX is equal to summation Xi, i going from 1 to N is a sufficient statistic for Mue.

So through factorization theorem because this joint PDF is factorized into two terms; one term involves Xi only other term involves this statistic and the parameter.

(Refer Slide Time: 18:19)



Now we will establish one very important theorem, Rao Blackwell theorem; suppose theta head is an unbiased estimator of theta and TX is a sufficient statistic for theta. Then another estimator will get theta head, that is the conditional expectation of theta head dash given TX; this is the conditional expectation, this is unbiased, not only unbiased but its variance of theta head is less than or equal to the variance of theta hat dash.

So this is the Rao Blackwell theorem corresponding to any unbiased estimator theta head dfash, we have another unbiased estimator with lower or equal variance. We will prove this; E of theta head will be E of theta head because theta head is equal to this E of E of theta head is given TX. Now using the property of conditional expectation, so we have a property like E of X given Y so this will be same as E of X.

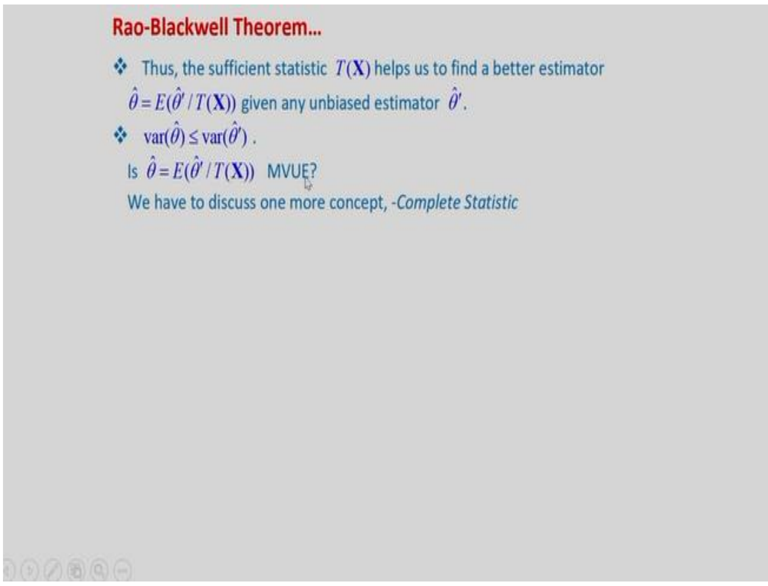
So this conditional expectation which is a random variable in terms of Y and if we take the conditional expectation in terms of Y then we will get E of X. Now we apply the property of conditional expectation E of theta head is equal to the expectation of this conditional expectation E of E of theta at this given T X and this will be same as E of theta head dash and this is unbiased therefore it will be equal to theta.

Therefore, E of theta head is equal to theta, therefore theta head is an unbiased estimator of theta. So we got the unbiased property of theta head. Now let us see the variance of theta head dash by definition it is equal to E of theta head minus theta whole squared because it is an unbiased estimator. Now I apply this conditional expectation property that will be E of E of theta head dash minus theta whole square given TX.

Now this mean square value theta head dash minus theta whole square is greater than equal to the square of the mean. So that way we can write in terms of square of the mean here, so now E of theta head given TX that will be equal to theta head and this is a constant quantity. So E of theta given TX will be theta itself. This quantity will be equal to E of theta head minus theta whole squared.

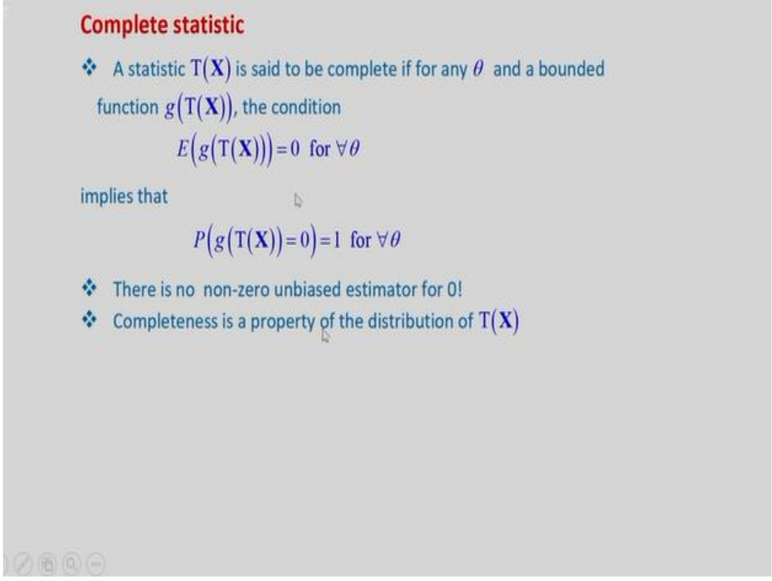
Therefore variance of theta head dash is greater than equal to variance of theta dash. That means variance of theta dash is less than or equal to variance of theta head , this that we established.

(Refer Slide Time: 21:38)



Thus, the sufficient statistic TX help us to find a better estimator theta head is equal to E of theta head dash given TX and variance of that estimator theta head is less than or equal to variance of theta head dash but still we have the question, is theta head that is equal to E of theta head dash given TX is their minimum variance unbiased estimator to answer this question we have to discuss one more concept that is the complete statistic.

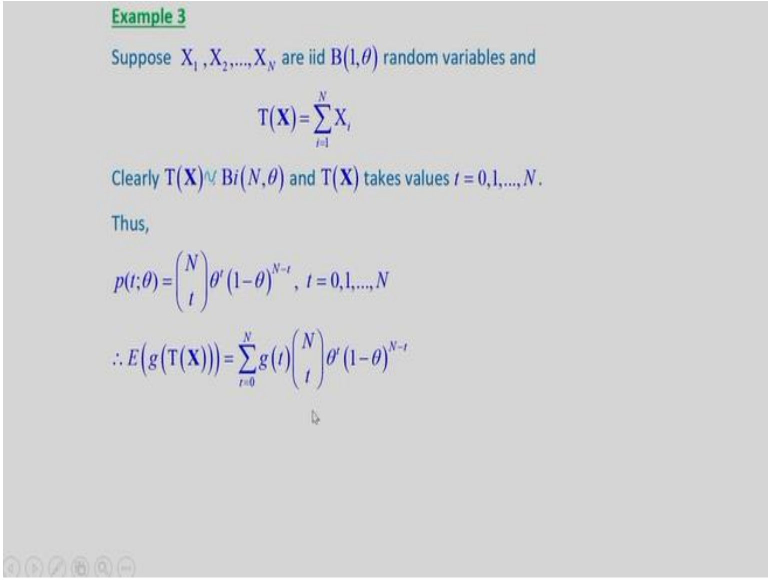
(Refer Slide Time: 22:17)



Complete statistic, a statistic TX is said to be complete if for any theta and a bounded function g of TX, the condition E of g TX is equal to 0 for all theta implies that probability of g TX is equal to 0 is equal to 1. That means g TX will be 0 with probability one, it also means that there is no nonzero unbiased estimator for 0. Suppose, zero is the quantity and we want to find out any estimator for zero such that E of g TX is equal to zero this is the definition for unbiased estimator.

But then this g TX must be equal to zero and there is no non-zero unbiased estimate or for zero. That way also we can interpret and completeness is a property of the distribution of TX. Same estimator TX may be complete for one distribution but may not be complete for another distribution.

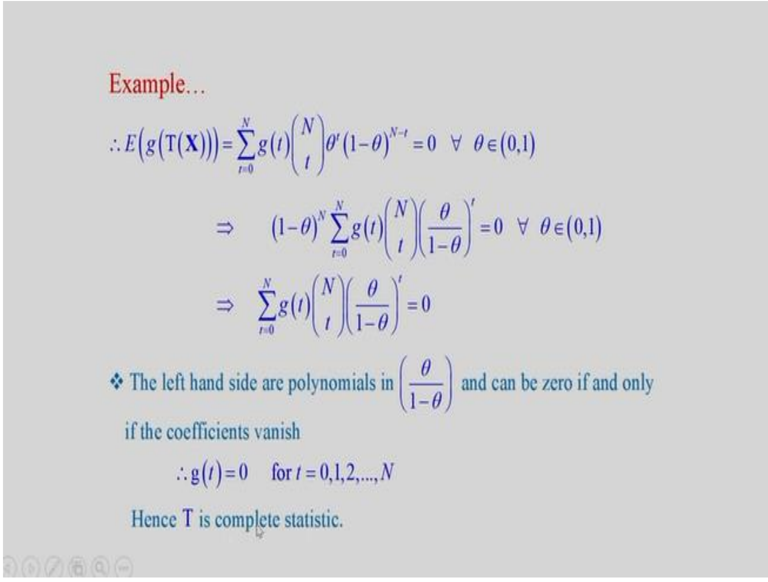
(Refer Slide Time: 23:31)



We will give an example; suppose Xi's are iid Bernoulli random variable, this is the symbol for Bernoulli random variable; it is symbolic like binomial with n is equal to 1. That is why we are writing B 1, theta and parameter theta is the probability parameter and TX is equal to summation Xi, i is equal to 1 to N. Clearly TX is a binomial distribution with parameter N and is the number of repetition and theta is the success parameter because TX is a sum of individual Bernoulli random variables.

So it is a binomial with parameter n and theta and TX takes value successfully 0, 1 up to N thus the probability mass function of at point T will be given by N c t theta to the power t into 1 minus theta to the power n minus t, t going from 0 to N and expected value of g TX, TX is a function of X. Now, so if I consider expected value of any bounded function g TX it will be given by this summation T going from 0 to N, GT into N c T into theta to the power t into 1 minus theta to the power N minus t. Now we will apply the condition for completeness.

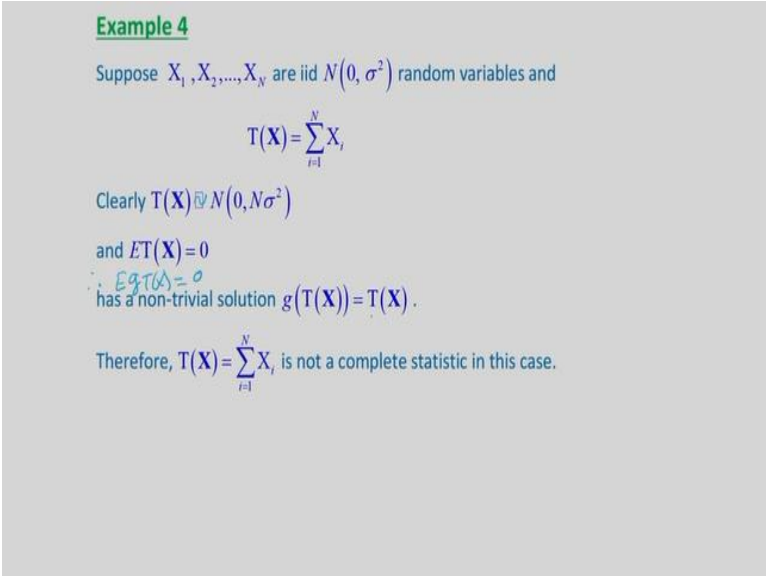
(Refer Slide Time: 25:02)



So, E of g TX will be this summation I showed earlier; now 1 minus theta to the power N is common, so that I can take out so it will be a same as 1 minus theta to the power N summation t going from 0 to N, gT into N c T into theta by 1 minus theta to the power t is equal to 0. So this equality can be written as this for all theta in the open interval 0 to 1 so theta can be anything between 0 & 1 excepting 0 & 1.

Therefore we get now one condition this part we can cancel out summation t going from 0 to N, gT into N c T into theta by 1 minus theta to the power T is equal to 0. Now this condition will be satisfied because it is a polynomial in theta by 1 minus theta. So this polynomial can be 0 if and only if all coefficients vanished, that means G T must be equal to 0 for t is equal to 0, 1, 2 up to N and hence T is a complete statistic. So in this case we prove that T is a complete statistic.

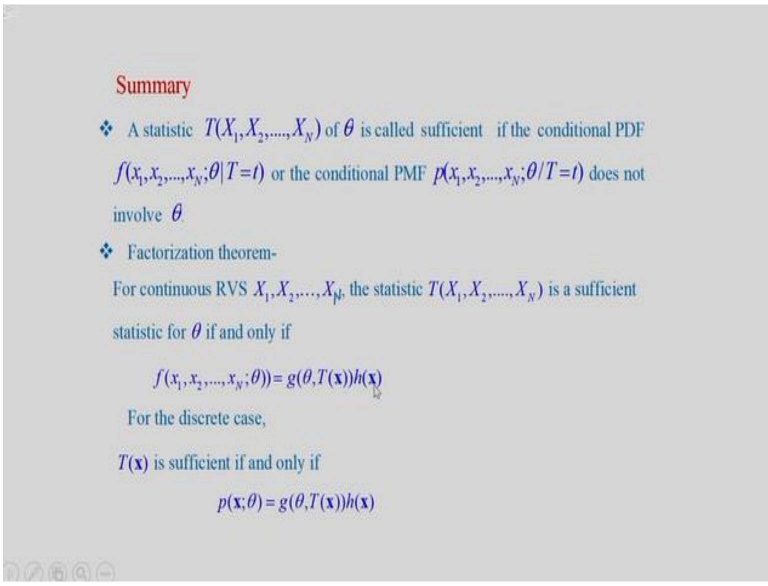
(Refer Slide Time: 26:24)



Now we will consider another example; suppose X1, X2 up to XN are iid normal zero mean variance Sigma square random variables. They are normally distributed with mean 0 and variance Sigma Square and TX is equal to summation Xi, i going from 1 to N. Clearly, because it is a sum and all X is 0 mean TX will be normal distribution and with mean 0 and variance and Sigma Square and we observe that E of TX is equal to 0.

So this is for non trivial X, therefore if we consider the equation E of g TX is equal to 0 it will have non-trivial solution g TX is equal to TX. If we substitute g of TX is equal to TX then this equation will be satisfied. Therefore TX that is equal to summation Xi, i going from 1 to N is not a complete statistic in this case.

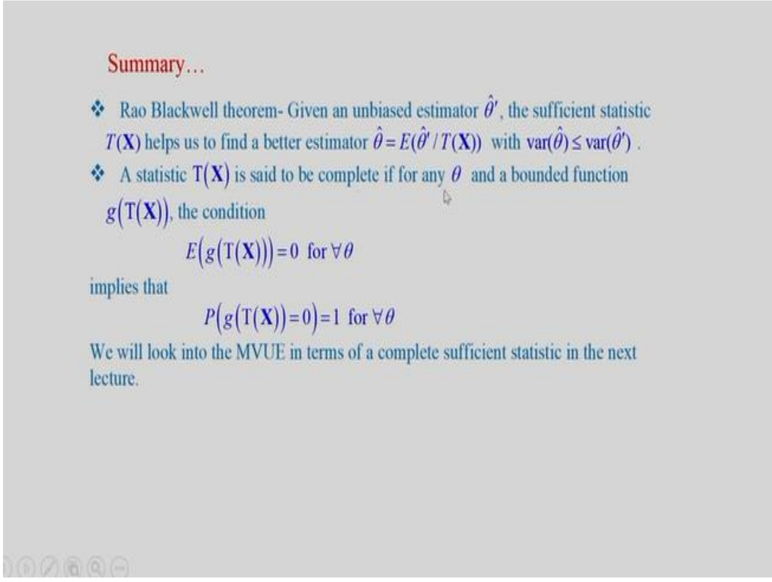
(Refer Slide Time: 27:36)



Let us summarize the class statistics T X1, X 2 up to XN of theta is called sufficient, if the conditional PDF that is conditional PDF of f x1, x2 up to xN as a function of theta given that T is equal to t or the conditional PMF p of x1, x2 up to xN as a function of theta given T is equal to t does not involve parameter theta. Conditional PDF or conditional PMF does not involve the parameter theta.

Then we discussed the factorization theorem; for continuous random variables X1, X2 up to XN, this statistic T X1, X2 up to XN is a sufficient statistic for theta if and only if. We can factorize the joint PDF as a two factors one factor g is a function of theta, TX and other one is a hX which is simply a function of X1, X2 up to XN. For discrete case, the TX is sufficient statistic if and only if joint PMF is product of these two factors.

(Refer Slide Time: 28:54)



Now Rao Blackwell theorem given an unbiased estimator theta head dash, the sufficient statistic TX helps us to find a better estimator theta head is equal to E of theta head dash given TX. This is the conditional expectation of theta head dash given T X. Now this estimator has the property that, it is not only unbiased but it is variance is lower than or equal to the variance of theta head dash.

Then we discuss about complete statistic, a statistic T X is say to be complete if for any theta and a bounded function g of TX, the condition E of g TX equal to 0 for all theta implies that P of g TX is equal to 0 with probability 1. So this is the definition of complete statistic. Now, we will look into the MVUE in terms of a complete sufficient statistic in the next lecture, thank you.