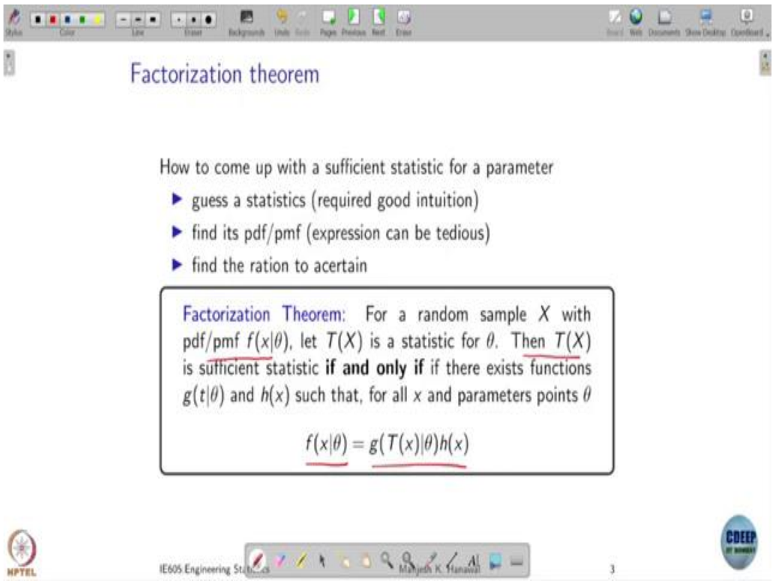
Engineering Statistics
Professor Manjesh Hanawal
Industrial Engineering and Operations Research
Indian Institute of Technology Bombay
Lecture 34
Example of Factorization Theorem, Minimal Sufficient Statistics

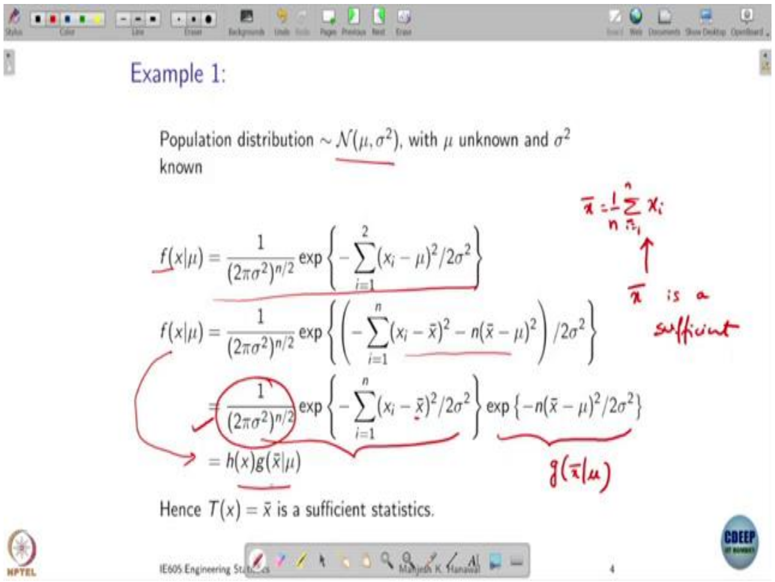
(Refer Slide Time: 00:25)



Let us revisit again this factorization theorem. So, we said that factorization theorem helps us identifying a sufficient statistics because this is a complete characterization weather of a, weather whether statistics is going to be sufficient statistics of a particular parameter on a particular population distribution.

What is the factorization theorem? We said that if I have a random sample with its population f and T is statistic for theta, then we are said that T(X) is a sufficient statistics if and only if I am able to find function g and h such that the density function splits in this format where h is a function of only on your random sample and g is a function of your random sample only through the statistics, the statistic t.

(Refer Slide Time: 01:44)



We discussed this example and if we discussed this example in the last class. But anyway, we already came across this example of having a population having samples from a population following Gaussian with parameter mu and sigma square, where let us say mu is unknown, but sigma square is known.

Now, let us say I want to start with a problem of identifying first a sufficient statistic. Right now, let us assume that you do not know all you know is factorization theorem. Factorization Theorem has said that if you are able to find a g and h function in some format, and this f can be written as a product of those two functions, then by looking into them, maybe you will find out what is the sufficient statistics. Let us try to find out this function f, I am able to write as a product of two functions h of h and g h only depends on x and g depends on x only through T.

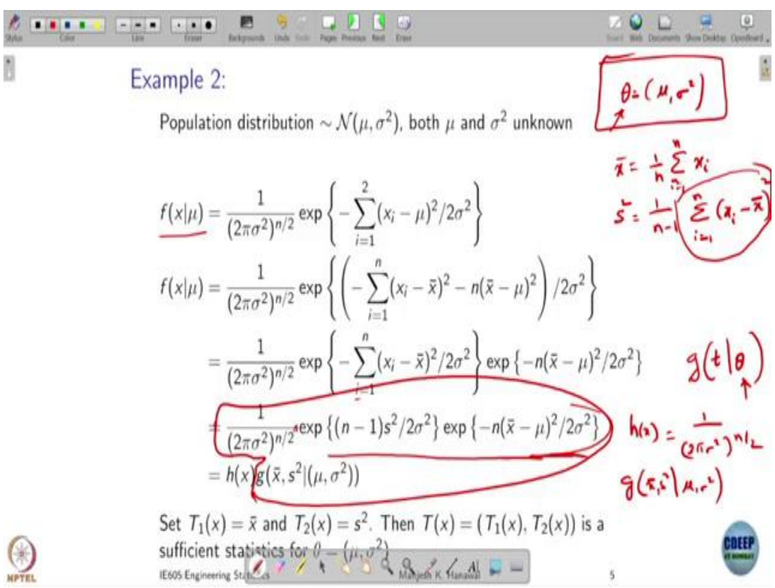
So, I am only interested in the parameter mu, because sigma squared is given to me only mu is unknown. Now, what I have done here is this is by the way, this is the joint Gaussian distribution and a little bit manipulated this exponent part here by adding and subtracting x bar, what is x bar here? It is the mean n samples here are taken mean.

Now, after adding and subtracting this you have to do a little bit algebraic manipulation after that, you will end up with this equation. Now, if you look into this sigma square is known mu is the only unknown thing this function here this portion here, can I treat it to dependent on only x bar given mu, can I treat this function as g of x bar given mu sigma squared is anyway known I do not need to worry only mu if I know this I can treat it as g x bar given mu and the rest of the part can I treat it as h(x)? Anyway the first part is constant x i x bar is the but x bar is again depends on all xi's.

So, I can treat it as h(x). So, now what is since we are able to factorize our population density in this format what it is saying what x bar is sufficient, but what you showed here is you just establish the fact that f can be factorized in this form you are able to get a h and g function where g only depends on x bar.

And you know now, this x bar is a sufficient condition why is this because you know that if this x bar if I am able to split f into this bar whatever this x bar it has to satisfy the definition of sufficient statistic because, that was the characterization of my factorization theorem I mean, that was the characterization of factorization theorem said you will be able to find such a factorization that h and g only if there x bar happens to be a sufficient statistics otherwise you possibly would have not been able to find this such a factorization.

(Refer Slide Time: 06:20)



Now, consider the other case where I have population distribution again Gaussian with parameter mu and sigma square, but here both are unknown mu is unknown sigma squared is unknown.

Now, again I want to see that now, my parameter is what my parameter is this mu and sigma square so, it is 2 dimensional law.

Now, I need to find out this parameter I need to find a sufficient I want to check whether for this parameter theta there is a sufficient statistic. So, what I will start looking into is see that if this function f I will be able to factorize into form g into h. So, again start do the standard manipulation that we did I added x bar here and then I am also trying to bring in a square so, x bar is your sample mean and we had also one more quantity called S square? What was that and what was the, now what I will do is? I will try to write this into S bar.

Now, you will see that now this quantity I have here so, I will cross multiply S squared with n minus 1 so, I will get this and this other factor I will just keep it like this. Now, let us see if I have the required factorization here. Now, if I give you mu, conditioned if I tell you mu and sigma square can I think of this form only depends on x square sorry x bar and the sigma square sorry x bar and S square.

Student: (())(08:41)

Professor Manjesh Hanawal: 1 by?

Student: We need to take (())(08:47) sigma square…

Professor Manjesh Hanawal: Yeah, we will do that. But first let us focus on the function g function the g function has to be such that this only depends on the statistics given parameter theta. So, now let us say if I assume that theta is given, can I say I have to basically this is up to me like I how to come up with a g and h function.

So, now let us try all the combination let us start with what you are saying. I have these three products here and how to grip them and see which I can take it as h and which I can take it as g, for according to you which can be h here?

Student: 1 by 2, 2 by 4.

Professor Manjesh Hanawal: That is a constant, that can be h(x) can be constant, that is fine. But if I wanted to define this to get that if, let us say if I want to take wanted to take this together as h, is there a problem in that?

Student: (())(10:02) sigma square.

Professor Manjesh Hanawal: It depends on the sigma square. So, that is a problem, because this h is not supposed to depend on the parameter. And similarly, if I want to take this let us say this first and the last one this one and this together, we ran into the same problem, because that depends on sigma square as well as mu also. So, the only way for me to now get out of this way is consider this possibility like if I tell mu and sigma square, does this only depend on some quantity yes, if I tell mu and sigma square this product depends on what?

Student: (())(10:58)

Professor Manjesh Hanawal: So, if I take if I tell you mu and sigma square this depends on, now this additional quantity x bar, S squared. Now, 1 possibility is if I took them as statistic, then I could treat this entire quantity as g T by mu sigma square or like yeah, they are now instead of t, I have to I have to take that there is now not 1 dimension there are 2 dimension. So, the first component is a square and other is x bar and S squared given mu sigma square now, and this h(x) could be simply thus constant. Now, what is the sufficient statistic?

Student: (())(11:48)

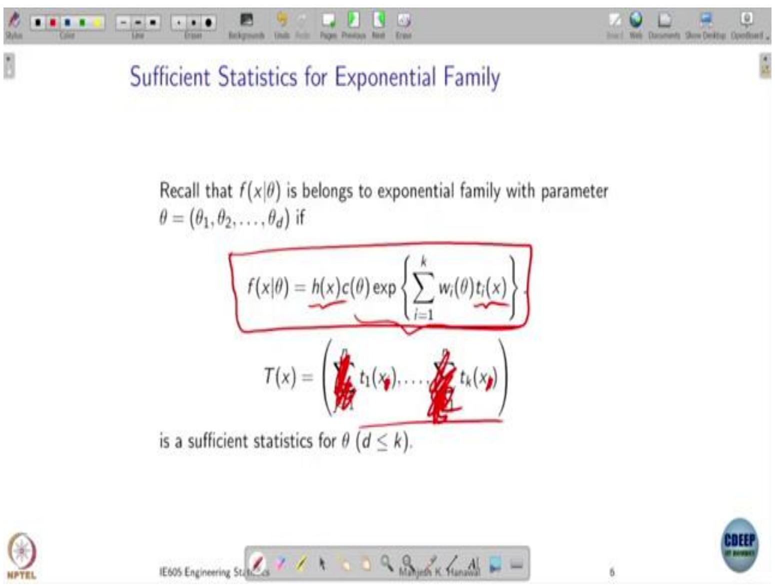
Professor Manjesh Hanawal: Everybody agree now x bar and S squared together is a sufficient statistics according to this definition. Even though it is I mean, I, even though we know that x bar is a proxy for mu, and a square is a proxy for a sigma square, but here, I am interested in this entire parameter, which consists of 2, so that is what I need to specify completely for this both the one. So, that is why I need to consider them together x bar and S square. Now, if you interpret like this, now, this is going to be sufficient statistics for theta. Now, I said now, I said in this case h(x) equals to in this case h(x) equals to 1 upon 2 pi sigma square n by 2 and g of x bar S square given mu sigma square is this entire thing this entire exponent form.

But instead of that, so, I have now given g to be this quantity sorry h to be this quantity and g to be this entire quantity. Instead, I wanted to consider it differently, I wanted to define this g to be g to be this entire thing, for this if I add a constant it is still a function of x bar and S square only then what is h(x) in that case? That is going to be a constant 1 there are different ways in which you can think of this h and g functions here. It is need not that h and g need to be unique that theorem is not saying that there exists a unique h and g that such that this happen all the theorem said is?

Student: (())(13:54)

Professor Manjesh Hanawal: There exists functions g and h all that you just need to show that there exists some g and h function this factorization happens. You need not I mean necessarily there exist in a unique way.

(Refer Slide Time: 14:15)



Now, let us look into the exponential family. So, where does let us one looks into something and see that we can by just looking into the structure, will it help us to find what is a sufficient statistic so, we know that exponential family many many distributions belongs to exponential family. By the way, did you come across anything which did not fall in exponential family? T distribution did not fall in.

Student: (())(14:49)

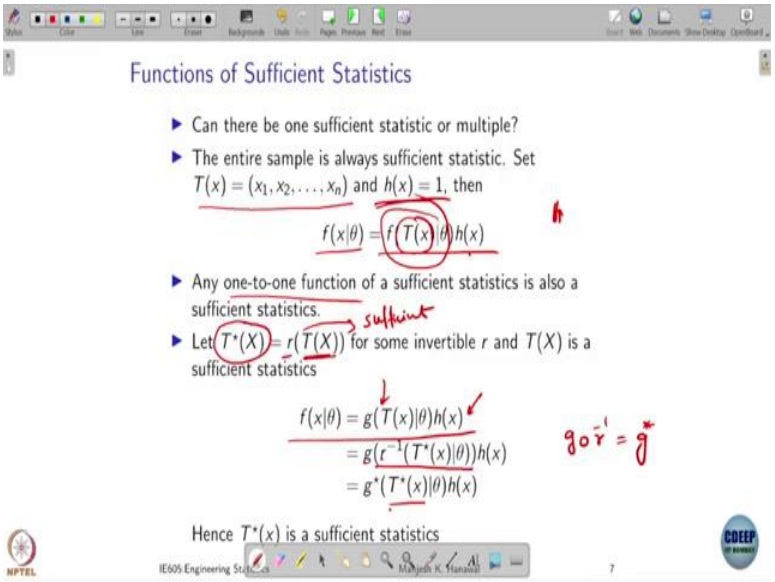
Professor Manjesh Hanawal: F also did not fall. T for any b and what about the uniform distribution did uniform distribution fall in exponential family? Okay fine. Let us say we know that many distributions fall in exponential family and this exponential family has this nice structure. We said that our distribution falls into an exponential family if it can be written in terms of h C wi's and Ti's. Now, you see that already this this PDF is already in a way factorize into h function C functions and within the exponent W and Ti.

Now, this is already in some factorize form, can I apply this factorization theorem and see if this exponential family has a sufficient statistic, what is that? So, first part h is already given for you for free in the exponential h is already there, you only need to worry about g function and c theta is also harmless because it is not involving any x.

So, given theta it goes away, what matters is only the things inside but the things inside are also nicely beheld theta and Xs are separated. Once you give me theta, what matters is only Ti's now and up because wi is our constant now, what matters is only tis, then all these Ti(x) are I do not know why I added this summation here a typo.

So, there is all this t1 to tk on that x are is going to be I do not know again this there is so many typos here so, this vector of t1 of x, t2 f x all the way up to tk f x is right away is a sufficient statistic for you. You do not need to do all the circles of seeing that whether the ratio becomes independent of theta like one of the characterization of sufficient statistics we said it is if you take the ratio of its unconditional probability and conditional probability of the statistics given the parameter that is independent of theta that was one test for us. But, if you are going to use a factorization theorem, you do not need to do all those circles you have right away, not only statistics, but you also know that this is going to be a sufficient statistic.

(Refer Slide Time: 17:47)



I just said that in the factorization theorem, the factors g and h need not be unique. But are the sufficient status themselves are unique or there could be multiple sufficient statistics for a parameter there could be multiple. Then if I tell you one sufficient statistics, can you generate another, how multiplying the constant multiply what? Multiplying by constant is trivial thing you can do, what else smart something more there because multiplying the by of constant you know constant if you multiply and you will divide and give it back to you. There is no addition any more information in that.

Student: Sir, t which is of the information of the given data, then instead of p we also (())(18:55)

Professor Manjesh Hanawal: What kind of function?

Student: (())(19:00)

Professor Manjesh Hanawal: So, that mapping has to be there, so, that is what we will say now. So, if t is doing the data reduction and having all the information and you do another transformation of that in a unique way that should still retain that property that is what we are going to. Another that is what will come, but you will see that if I have a random sample consisting of n points, I can simply take my statistics to be that sample itself.

When you are trying to extract information about that sample I am giving you that entire sample itself is your value. That sounds so natural statistic sufficient statistic. I mean, you are not even you are not reduced the data you are just given me everything. You do whatever you want. So, that is a in one natural sufficient statistic sorry statistic and it also happens to be a sufficient statistic because you can always write fx given T in term like this, where h(X) is just 1 you can write always f(x) given theta is f of T of x theta into h of x h f is 1 and this T of t of x is an identity function.

So, I think a change and our factorization theorem is saying that this T is indeed a sufficient statistics, because h is already there, and this I can treat it as my g function. Now, as you some of you noticed, if you take a sufficient statistic and take any 1 to 1 function on that, that continues to remain a sufficient statistic.

So, how is that? I mean, you can see that, again using our factorization theorem, suppose let us say T of X is a sufficient statistic and you are going to do a transformation of that using some function r and we want the transformation to be 1 to 1, if it is 1 to 1, this function r is invertible. So, that our function r is invertible for me.

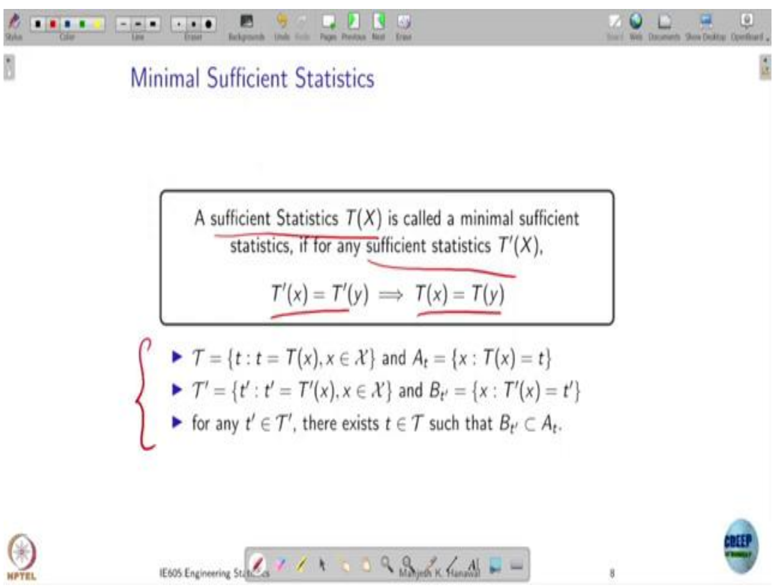
Now, let us see how I can use my factorization theorem to conclude, if T(x) is also T star, the new sufficient statistics I have is also a sufficient statistic sorry, the new statistics T star I have is a sufficient statistic. If I started with T which is a sufficient statistic, I know that my factorization theorem tells me I will have g and h function such that this factorization holds, this is simply through a factorization theorem. So, I hope all of you in sync with me I am saying that T is a sufficient statistics here, by doing the transformation r I got a new statistic and I want to check whether that is a sufficient or not.

So, now T(X) so, T(X) I can invert this relation here, then T(X) is going to be simply r inverse of t square x, now I can treat now, I have three function g function R inverse function and T's function I can treat this g composition r star as another function g star, then I will have g star of t star of x and now, this g star function only depends on x through T star x all of you see this if this is the case can I now claim through my factorization theorem T star is a sufficient statistic.

So, if you give me a statistic, you do not need to just multiply some constant you take any 1 to 1 function apply it on that and you will get a new sufficient statistic and by taking different 1 to 1 functions, you may generate as many sufficient statistics you want. Now, we just discussed that there could be many sufficient statistics.

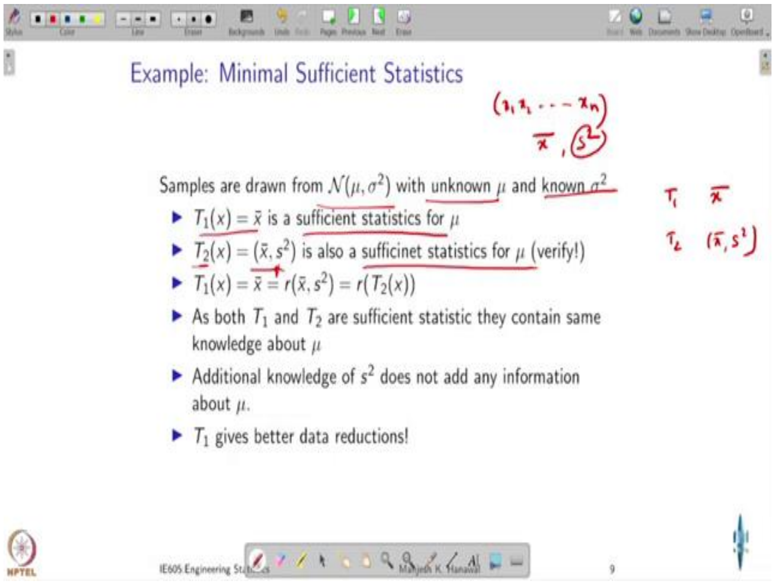
So, first thing is we started with the statistic, statistic is something which is giving you data reduction but what we wanted is the guy who best reduces data in some sense that is where we will look for sufficient statistics, we say that sufficient statistics once you reduce that set you do not need any additional information. But now, we want to see like now that there are so many so many sufficient statistics we are not happy. We want to see that among the sufficient statistics which are good and that is where we use the notion of minimal sufficient statistics.

(Refer Slide Time: 25:12)



Now, what is this minimal sufficient statistics? It says it says that we are going to call a sufficient statistic T to be minimal if we if there is any other sufficient statistics t prime, this condition implies this condition what is this come condition it says according to this sufficient statistics X and Y points are the same, that is what we said if x and y points are giving me the same statistics, they are kind of same for me if they are same under this t prime them they are also same under T. So, if these sufficient statistics is saying that these two points are same the one which is better should also say that these 2 points are also same.

(Refer Slide Time: 26:19)



We will not go into that instead of that let us look into directly example, what is the minimum sufficient statistics where it arises, let us consider a case of a sample coming from this Gaussian with parameter mu and sigma squared where I am now assuming only mu is unknown, but sigma square is known.

Now, if I have samples let us say x1 x2 which are coming from this population I can compute x bar from the same sample I can also compute sigma square sorry S square and I know that if I take x bar as a statistic, this is a sufficient statistics for my parameter mu, since I can compute S square also I can take another statistic which is x bar and S square I know that this is also sufficient statistics for mu why you can simply ignore the square and then it is but it is more your and nobody is stopping you from computing S square from the samples.

But now I can obtain T1 from this T2, I can I obtain T1 by T2, T2 you are basically giving me x bar and S square I can just ignore a square and I got x bar which is basically T1 whether you give me x bar the T1 which is containing x bar and T2 which is containing x bar as well as S square if I think about mu information about the parameter mu both are having same amount of information.

But the additional information provided by the statistics S square is not adding any more information about the parameter mu than what has been already provided to me by my parameter mu. So, if you think in terms of data reduction, both T1 and T2 are good in terms of getting information about my parameter mu both are sufficient statistics.

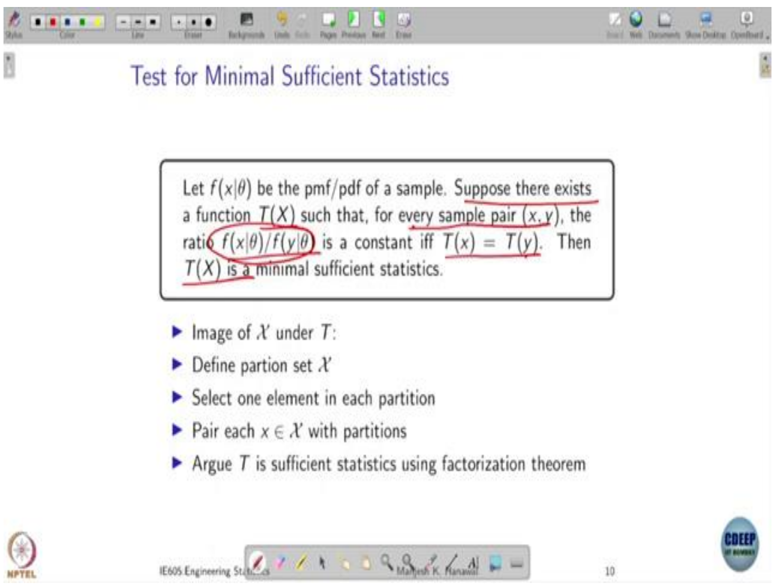
But from the data reduction point of view, which one is good? T1 is good, because T2 is unnecessarily storing S square which is not providing any more information about mu once I know x bar so, not only I am interested in how well information about my parameter is contained, I also want to provide any more redundant information to me.

So, this guy is providing this about mu this this is a redundant information here S square. So, that is why in terms of T1 and T2, I would we can say here T1 is better here about the parameter mu because it is storing less information I mean storing it is reducing in terms of data reduction this is better. So, which is minimal here into in about the parameter mu from T1 and T2 if you want if you compare T1 and T2 which is minimal?

Student: T1

Professor Manjesh Hanawal: T1 is relatively minimal compared to T2 here from the data reduction point of view.

(Refer Slide Time: 30:25)



I will just leave say this definition and leave its proof because at least you people can continue looking into the problems, which uses this definition of minimal statistics and other things we will do later. And this is involved definition let us focus let f of be a probability density function suppose that there exists a statistic T and if every pair of samples x comma y this ratio of this PDF is a constant if and only if they provide the same information through the statistics about the parameter mu, that means they are they are same means they are kind of indistinguishable to me. If that is the case, then I am going to call such that T is going to be minimal sufficient statistic.

So, just understand how to check a sufficient statistics is going to be minimal or not. One way to do is take a sufficient statistic take two points on which x and y x or x and y on which it is giving me the same value and now if I take the ratio of this PDF on these two points x and y that should be a constant. If that is the case, then that is a minimal sufficient statistic.

On the other hand, these are minimum sufficient statistic and if this ratio happens to be a constant, this is going to happen only this is going to happen. So, we will not get into the proof of that those who are interested take a look into that because it needs us to is proof is not simple. But yeah, we will we will skip this those who are interested look into that, I will stop here and next class we will revisit this and discuss the examples related to the minimal sufficient statistics.