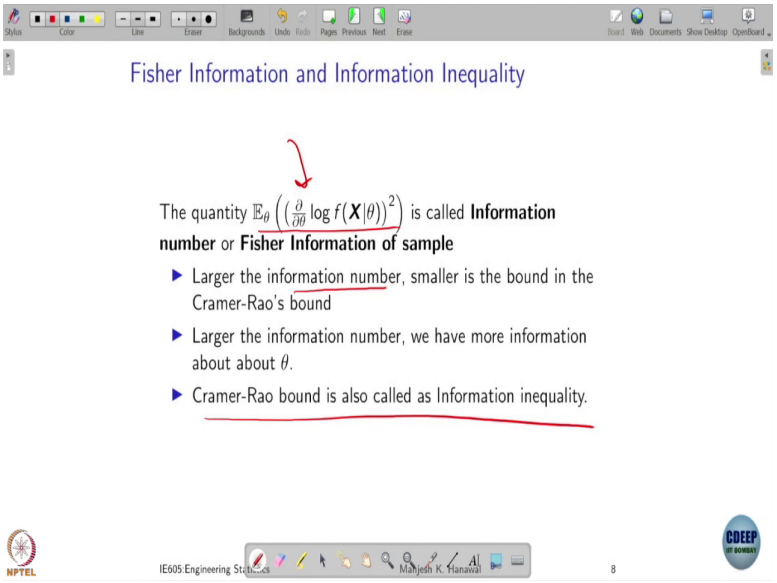
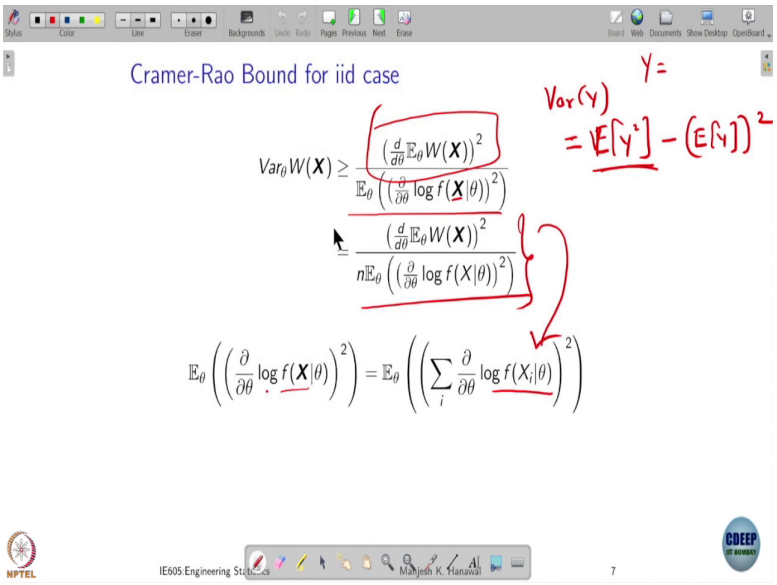
Engineering Statistics
Professor Manjesh Hanawal
Department of Industrial Engineering and Operations Research
Indian Institute of Technology, Bombay
Lecture 41
Hypothesis Testing, Likelihood Ratio Test

(Refer Slide Time: 00:22)





So, in the Cramer-Rao bound, we ended up saying this term. The partial derivative of your log of your PDF function. And now you see that, this actually governs this denominator actually governs how the small your lower bound can be. So, because of that it has been given a special name and it has also special interpretation the, it is called information number

or it is called Fisher information of your sample. And naturally, if this information number is going to be larger, the lower bound in your Cramer-Rao bound is going to be small.

And it in a way also says that if your variance is going to be small. So, if the estimator is such that its variance is going to be small. What does that mean? Louder, data is, it is the best estimator that is fine. But in what sense? It is able to essentially capture information in the data about your parameter well.

Student: Data is spread it out.

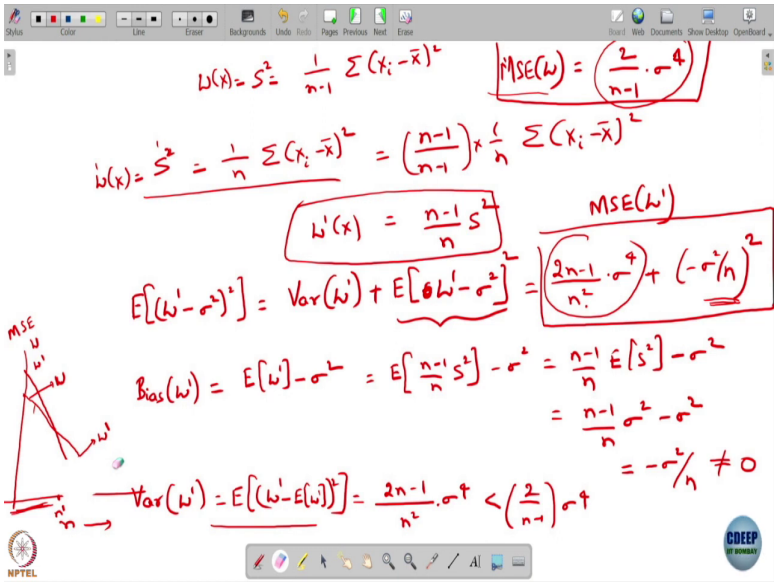
Professor: Data is? It could be, yeah, let us spread it off for because of which PDF itself if the data is less spread off our data is not going to be spread out too much. That is fine it will concentrate about its variance part. But now, if your sum, if you are, by the way notice that this quantity here it does not depend on the estimator, it is only about your PDF function. So, then in a way what you people are right like, it is not about estimator irrespective of what is your estimator? What matters is how good my data is spread out.

Like how far my data is spread out, so, it is a property of only your PDF function. And if your PDF is such that, the data is spread out too much will this be larger or smaller? If your data is spread out, what do you expect your any estimator is going to do a good job or bad job? Bad job. In that case, why do you expect this quantity to be? Lower, because the Cramer bound says that it is coming in the denominator that means it is going to say any estimator has to incur a large variance.

But on the other hand if your data is not so spread out, it is easier to infer the parameter maybe then it may be the case that this quantity is going to be larger for that. That is why sometimes because of the appearance of this quantity in the lower bound in Cramer's Rao bound, it is also called Cramer Rao bound is also called information inequality.

So, in the tutorial, we will see like how to compute this lower bound for various PDF functions and all. So specific examples we will see the tutorial, but any question about the general steps one has to follow in computing main square estimator or Cramer or this Fisher information or the Cramer or lower bound, if you have any question you should ask now.

(Refer Slide Time: 04:29)



Student: Audio not clear.

Professor: This one.

Student: Audio not clear.

Professor: Why not? We could use, right. So here I forgot to plot this. I mean, I did not plot this but if you people read the book, you can plot. Let us say you have this, mean square estimators as a function of n, n is on the X axis. And let us say mean squared error on the Y axis and this could be for both I am going to w and w prime I will plot, it may happen that by the way, both of them the mean squared error are going to fall with as n increases. So let us say for w this is the top one is W, let us say it is going to fall like this.

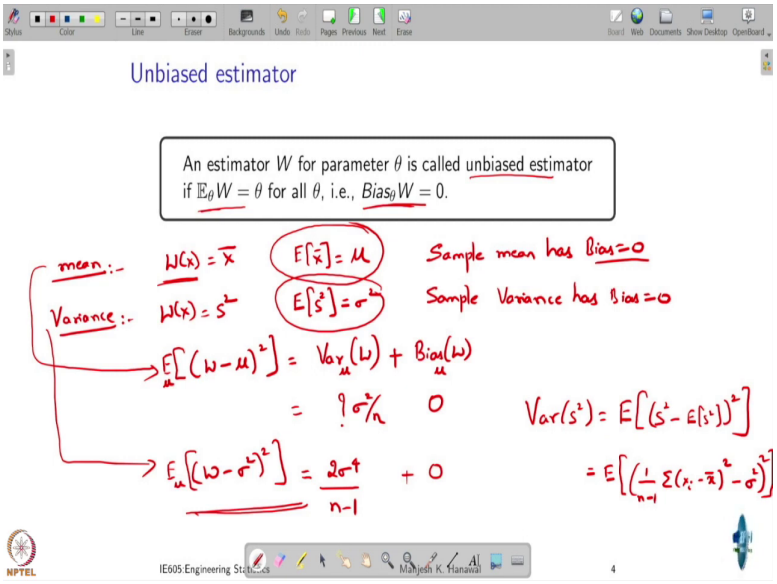
And for w prime, it may happen that initially it may be larger, but as n goes, it may start falling faster than this at some point, let us call that point as n prime. So, for n small n smaller than n prime, which one you feel is better? The second one, sorry, this one, let us call this as w and let us call this as a w prime. So, you see that for n in this region, when n is small, your w is better actually. And it may happen that as n increases beyond n prime, your w prime may be better. That depending on how many samples you have, you can decide whether the biased one is good or unbiased is going to be better.

So of course, this is the mean squared error is not going to fall down to 0, but this is just like, I mean, this is just for representative purposes, but I hope you got the picture and this is

where the analysis is important when you have to you have data. And depending on your samples, you need to decide which variants is going to work out better for me, maybe you need to compute all this mean squared errors of your various estimators you can think of, and it may end up that you may want to you may end up using biased estimators because for that many samples, maybe a biased estimator will work out better.

So, it is not necessary that all the time, unbiased estimators are going to do a good job, maybe biased estimator can also do a good job but that depends, yes, that is why it is important to compute all these expression maybe for some toy examples. Like first you need to see, your data is discrete maybe if it is a discrete and you feel that it looks too close to binomial compute all these things for a binomial. And if you feel that your data is looking more gaussian and then compute all this for gaussian.

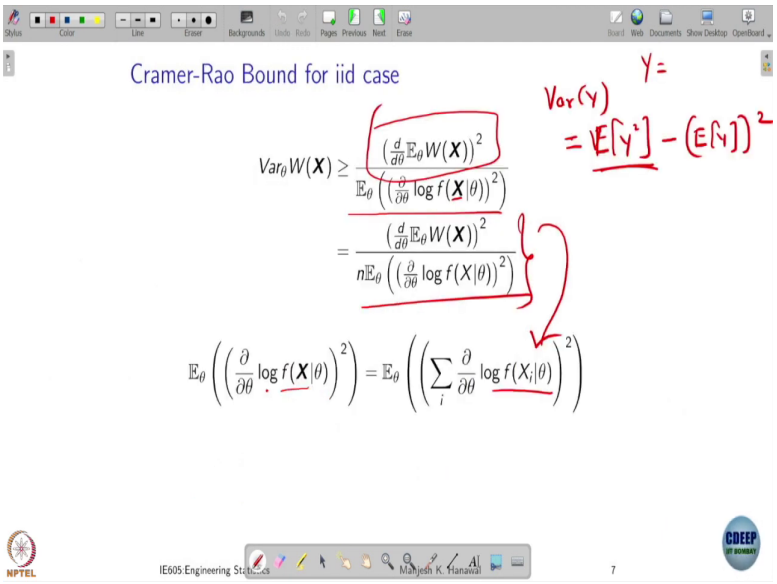
(Refer Slide Time: 08:02)



And by the way, notice that this is you can compute this expression is true, this expression is true independent of what is your underlying distribution, agree. This is also true irrespective of what is your underlying distribution maybe you need some property here, to calculate the variance of your variance samples, variance estimator. So, that is where you need to see which distributions you should use to make these computations. So, maybe what I have written here this is, this holds for the Gaussian distribution not necessarily for everything. This this holds for every distribution, but this to compute variance of their variance estimator that is not easy.

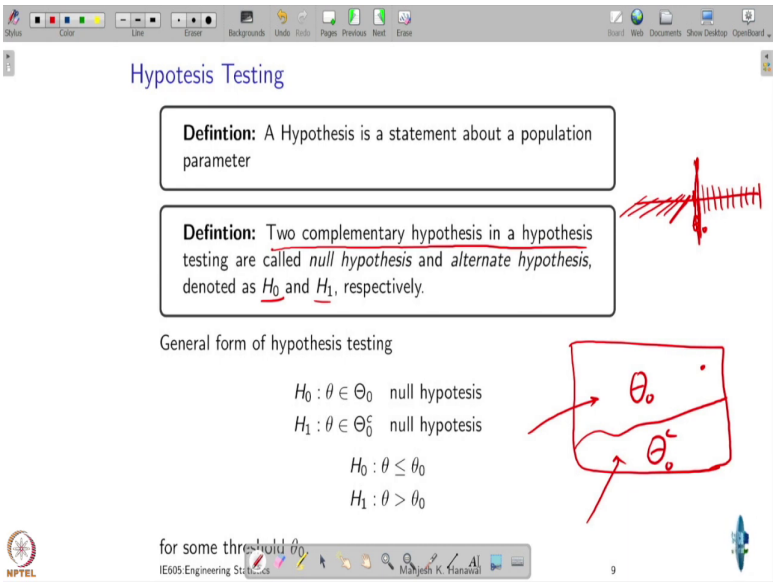
That maybe you will be able to compute only for specific distributions like gaussian or maybe simpler ones. So, depending on your data is better representation your data is closer to the gaussian or better represented by exponential maybe you should use that particular distribution, put it plug and see what is a better works out for you.

(Refer Slide Time: 09:21)



Any other question on this? If not, you will see the examples on this in the tutorial. So let us now move on to our next topic called hypothesis testing.

(Refer Slide Time: 09:46)



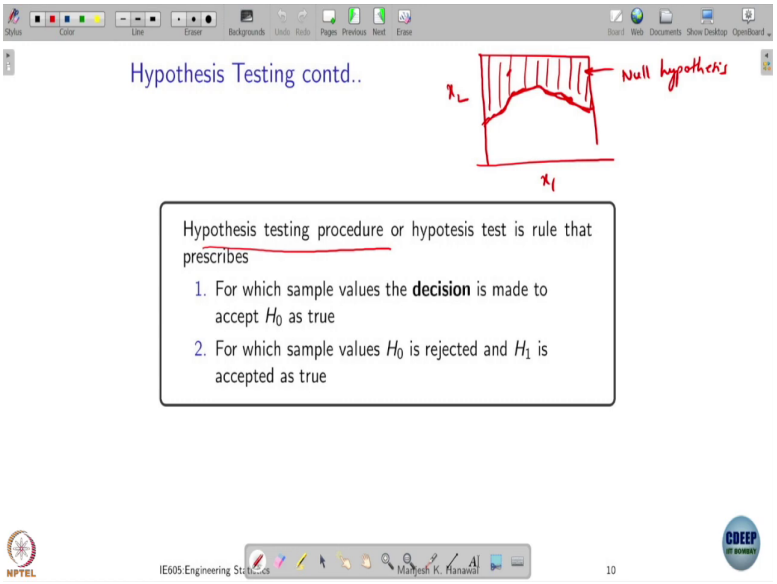
So how many of you heard this topic before? Yes or no? Now, the same question we are talking about asking about the parameters we are estimating the parameter we could ask in a different way. We can ask whether the parameter the data that I am seeing, whether it is going to represent this population parameter or not? I can ask kinda yes no questions, earlier I was exactly trying to find out what is the parameter, but here I could just say maybe yes or no, or maybe it belongs to their here or there, maybe these kinds of things. And so, I will make hypothesis like that, yes hypothesis is this, no hypothesis is this, now it boils down to checking those hypothesis.

So, the definition is a hypothesis is a statement about a population parameter. Me if the hypothesis can be okay the parameter lies in this range or that range, this is like hypothesis. Now, you decide whether it lies in this range or that range come up with a criteria to evaluate it. Most of the time, we go with two hypothesis and hypothesis testing, which are complementary to each other. And the two complementary hypothesis and a hypothesis testing are called null hypothesis and alternative hypothesis and they are often denoted as H0 and H1 respectively.

And here the general form of the hypothesis testing is you will assume that maybe you will make a let us say this is your parameters theta space, you have partitioned into 2 parts. Let us call the upper part as theta 0 and the lower part as theta 0 compliment, you do not care which is a particular point it belongs to, what you care is whether my theta belongs to this or this, there are only 2 hypotheses here, my theta belongs to this region or this region and you need to find out and come up with a method to evaluate it.

And if you take a real line, my hypothesis could be as simple as maybe you put some threshold here, some this is your known threshold and you ask the question whether my parameter lies in this region or in this region, this is like a boundary.

(Refer Slide Time: 12:57)

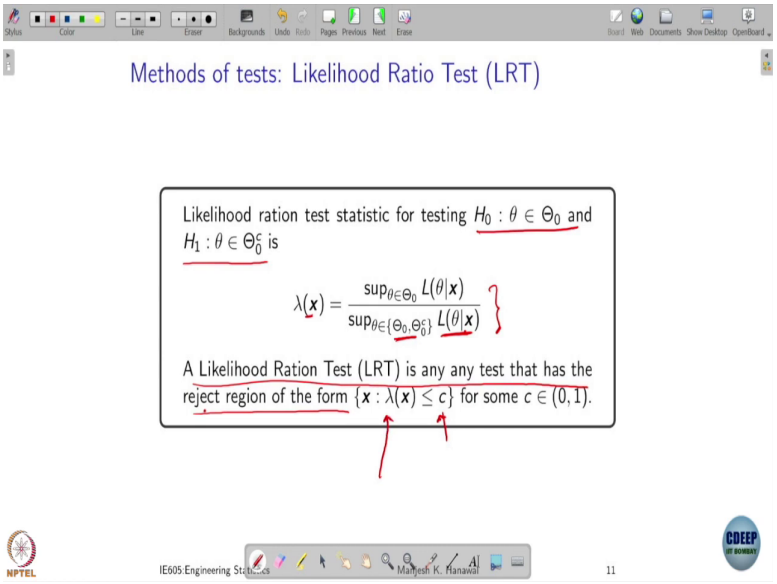


Now, hypothesis need to be tested. You have two hypothesis I said null hypothesis and alternative hypothesis in which you are they need to be tested. And now, that hypothesis testing procedure or a hypothesis test is a rule that prescribes for which sample values the decision is made to accept your null hypothesis to be true. And for with sample values, here H naught is rejected and H1 is accepted as true. Naturally in this case, since there are only two hypotheses to be tested, when I reject hypothesis, null hypothesis that indicate that I am accepting alternate hypothesis.

So now, the previous diagram, I showed one as a parameter space. Now let us say this is my sample space and I have only two samples and this and maybe there is some partition here. I need to come up with a decision rule like this. Maybe something which says that all the points which are here, they corresponds to let us say a null hypothesis.

I have come up with a decision boundary here saying that if my points belongs to in this region, I am going to accept it as null hypothesis and if it is coming from any of this space, I am going to accept alternate hypothesis or basically reject my null hypothesis. Now the question is how to come up with a decision boundary. How who is going to give me the decision boundary? So for that, we need to come up with a method.

(Refer Slide Time: 15:08)



So, the method to come up for this decision boundary we are going to use something called likelihood ratio test. So, all of you already know what is the likelihood function. Now, we are going to use it to define likelihood ratio test and using this likelihood ratio test we will define our decision rules. Now, if you have given, two hypothesis H naught which says my parameter belongs to the space theta 0 and H1 which says my parameter belongs to the complement of that set.

Now, for any random sample x you are going to take ratio of these two quantities where the numerator is going to compute the likelihood and maximize the likelihood function over your space of null hypothesis and the denominator is maximizing it over all possible parameters this is both our null and alternative space this is all possible parameters. So, now, let us go back and recall what we said about our likelihood functions.

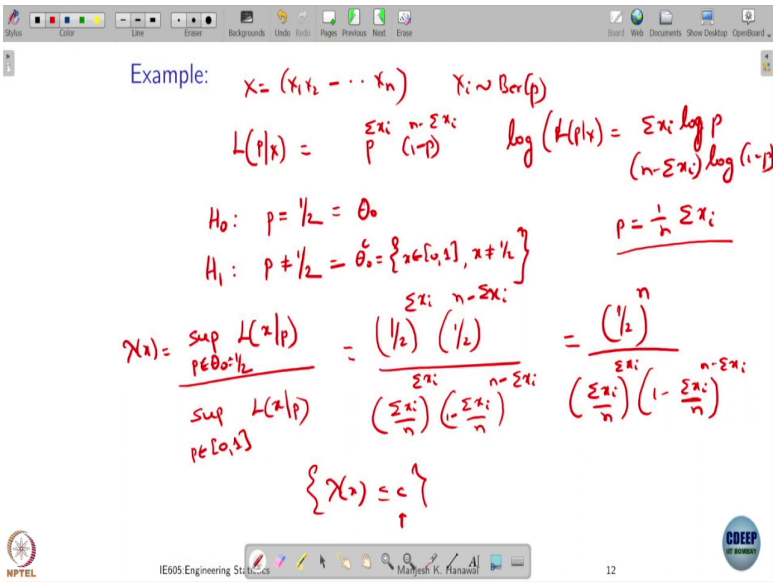
So, likelihood function is capturing how likely that value theta is for my observed samples. So, the numerator is trying to compute the best theta that is explained in my observed sample x and the denominator is called capturing all possible theta that is among all possible theta which is explaining my x best.

Now, just think intuitively if this lambda x happens to be large, what do you expect? There that means, that numerator is dominating if lambda x is larger that means, sometime in the theta in my null hypothesis is better explaining what is x and if this quantity happens to be small that means denominator is dominating that means, a parameter which is not in my

hypothesis is better explaining my data. So, hypothetically let us say if lambda x is going to be large you want to accept it has a null hypothesis or alternate hypothesis, null hypothesis.

But then the question comes what is this large? So, for that we are going to define a parameter. So, then a likelihood ratio test is any test that has a rejection regions of the form x lambda x is less than or equal to c. So, c is some parameter that you are going to decide. So, what is this is going to do is all the points for which this lambda of x is going to be less than or equals to x it says reject. That means, they are not coming from my null hypothesis they are coming from an alternate hypothesis. And this is my rejection region.

(Refer Slide Time: 19:17)



Example I forgot, can you think of an example which we can work now? Let us take simple case now let us take Bernoulli. My X is Bernoulli, x1 x2 to xn and xi are Bernoulli with parameter p. What is the likelihood function for Bernoulli? p to the power 1 minus, p to the power, let us take the log likelihood function that is simpler, let us take the log of it we know that nothing changes. Because we are optimizing. That is the changes, we are not looking for the argument we are looking for actual values here.

But for computing the optimal value that is fine, let us take the log of this and what are my, now let us define first let me finish this xi minus n (mi), that is what log p and n minus summation xi log 1 minus p. Now let us propose me a hypothesis, what should be our hypothesis, I am going to check hypothesis whether this tosses are coming from a fair coin or

not. So, what should be my high p high, my null hypothesis? p equals to half and what is H1 is going to be? p not equals to half.

Now, in this is my theta set, my theta naught set, here is this just to have one point. And what is my theta 0 complement has, it is like all x belonging to 0 1 and that x is not equals to half, it has every other point in it. Now, let us do the optimization. First, let us consider the numerator, where so, log likelihood lambda of x, is what sup over p equals to p belongs to theta 0, let me write it like this, L of x given p divided by sup over p belonging to entire thing 0 1 now, that includes half and the entire thing.

And L x given p, now let us compute let us I know that the maximum value it does not change with respect to, so what is the maximum value here? First numerator theta is what half, I am going to compute it at p equals to half, there is nothing to optimize, only one point is there. So, the numerator is going to be what, let me directly put half of summation xi, half of summation xi. And what is the optimizer here? I know that if I have to optimize it, my p is going to be 1 by n summation xi that we already noticed, everybody agree with this the optimal value of this.

So, the denominator is going to be summation xi by n summation xi, p I am replacing by this quantity and summation xi 1 minus n divided by n minus summation xi, so, what is the numerator? Numerator is going to be simply 1 upon n and what is denominator, is there anything I can simplify in the denominator? Nothing, we just keep it like this, xi and 1 minus summation xi by n and n minus xi. Now, what you are going to do is now, if you want to if a point x is given to you, a point x is given to you.

Now, you are going to come up with your rejection region lambda of x is this. Now, you are going to see if this is going to be less than or equals to c, for a given c if this is the case, you are going to reject it, you are going to say this is not coming from your fair point and you are going to say other way. On the other hand, if lambda x happens to be larger than you are going to accept it to be coming from a fair. So, all depends on this c, how you are going to set, so we will continue discussing this in the next class.