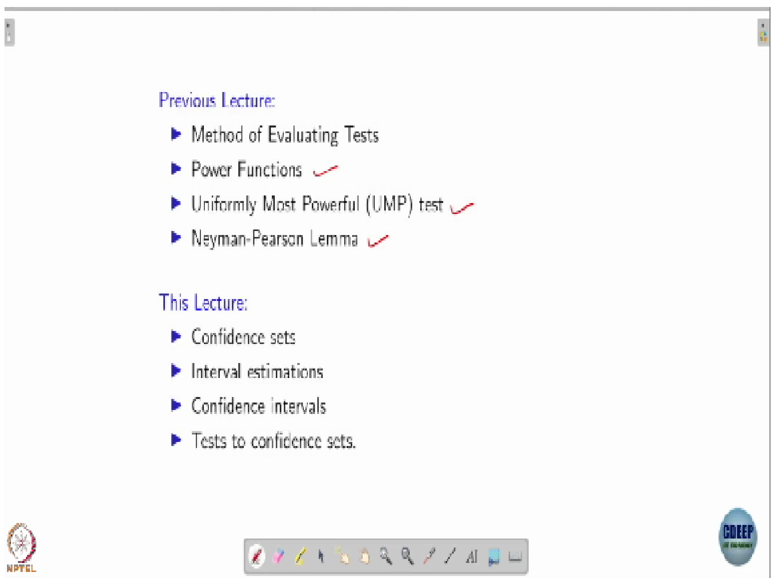
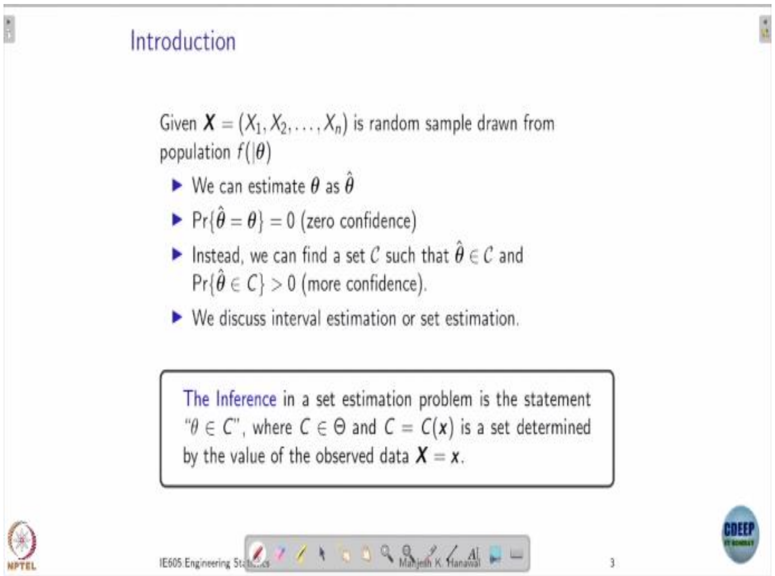
Engineering Statistics
Professor Manjesh Hanawal
Industrial Engineering and Operation Research
Indian Institute of Technology, Bombay
Lecture 47
Interval Estimation

So, in the last lecture we started discussing about little bit about the confidence sets. So, we will continue and continue that today. But before that we mostly talked about the hypothesis testing and there, we covered all these aspects of power functions uniformly most powerful function tests and then we talked about Neyman-Pearson Lemma.

(Refer Slide time: 00:44)



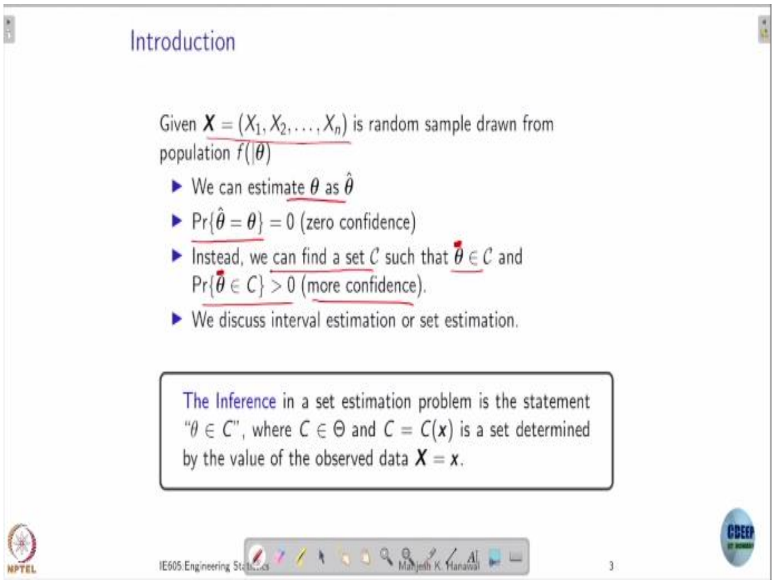


Now, so far what all the methods we discussed they were all like a point estimator. Like when we had a maximum likelihood estimator, we got one estimator based on the samples. One value for the estimator based on the sample and when we did method of moments again, we got one value for your estimator. And what are the other methods we discussed in the point estimation we talked about base estimators.

There also we get one value for your estimator. We did not talk about another method called EM method, Expectation and Maximization. There also I will try to find relatively one value. And in the hypothesis testing we mostly focused on how to come up whether instead of asking the question whether what is the parameter we just ask answer the question like whether this sample corresponds to this set of parameters.

Like for example we said whether my theta is equals to some theta not or not wanted us such questions there and where we come up with our rejection region. Now, in this confidence interval instead of looking for one value for your estimator we will say my estimator possibly lies in some set estimation my true parameter which I hope try to find out using my estimator estimated values we are going to say that lies in some region. So, in in that like earlier okay let us get started with this.

(Refer Slide Time: 03:04)

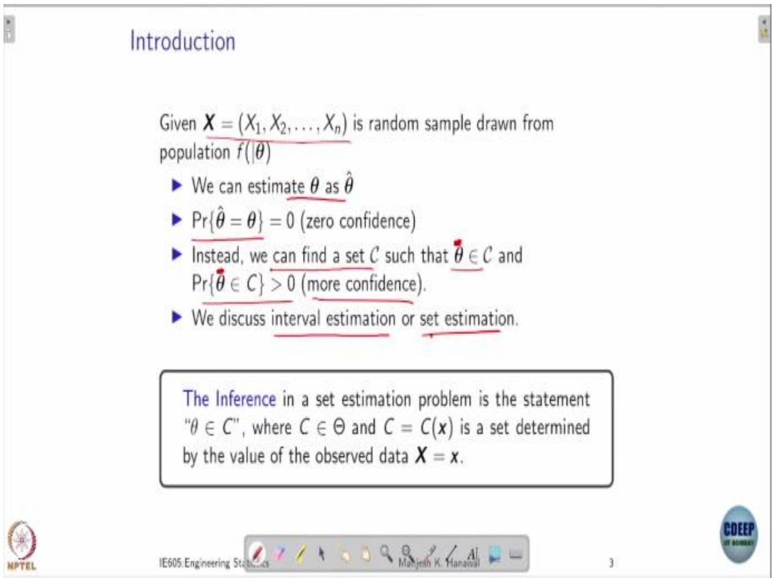


Suppose we have this random sample which are coming from some population with underlying parameter theta which is unknown to you. Now, we can estimate a point estimator for theta as theta hat and we have discussed many methods for them and whenever this theta happens to be continuous valued, we know that the probability that the estimated value is equals to the true value is going to be 0.

So, that means no way I can tell that my estimated value is actually representing the true value. So, my confidence is 0 in that, however, instead we can find a set C such that this theta hat belongs to C sorry maybe sorry this is there is no hat here. Such that the true value lies in that set and try to find out what is the probability that my true value theta lies in that confidence sets. And maybe we will be able to say that to happen with positive probability.

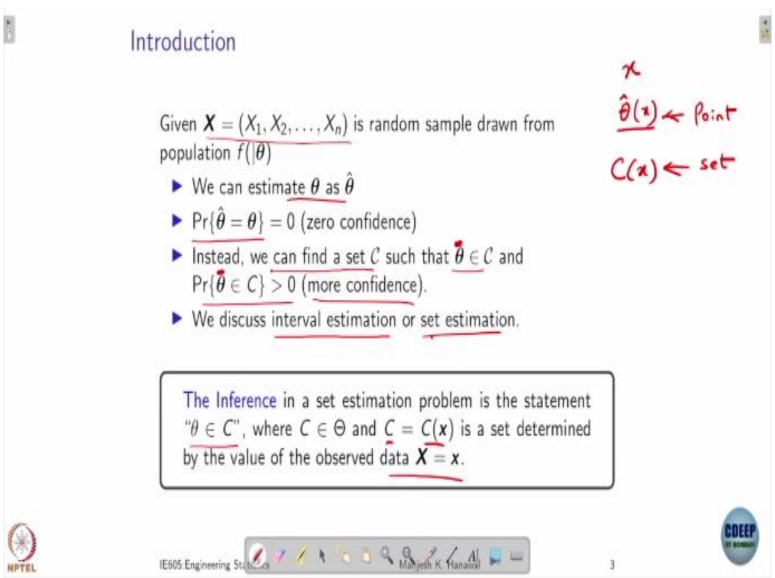
So, then maybe I can say that something more confidently. So, for example instead of saying that I will come to the class exactly at 2 pm you can say that maybe I will come to the class maybe between 2 to 2:05 pm. So, in that way maybe most of the times I will be between 2 to 2:05. So, that means like I am more confident exactly like I will be able to reach in that 5-minute window. So, that is not like I am giving you a window there.

(Refer Slide time: 05:03)



Now, we will discuss how to come up with such interval estimators which we also call it as a set estimation. So, in this problem our inference in a set estimation problem is a statement that theta belongs to C.

(Refer Slide Time: 05:22)

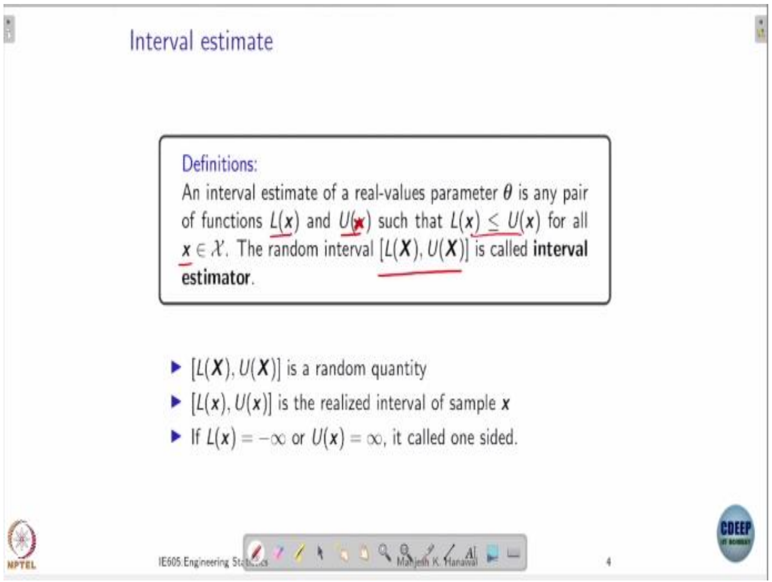


The set I want to come up where C is going to be obviously set of my parameters. I do not want to say anything outside my possible set of parameters that theta is my set of parameters. And this set C is itself is a function of your observed sample X is a set determined by or observed sample X.

So, what we want to now do is given a random sample X you can find theta hat x which of course will depend on X.

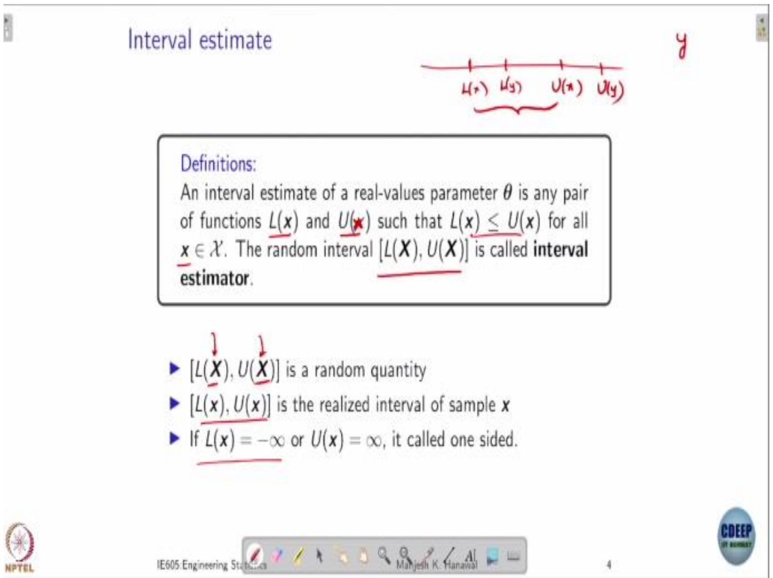
But now, instead of coming up with one value we are going to say that in this instead of this we are going to come up with a set now this is this is a point and now this is a set. Now, set how you are going to characterize this set.

(Refer Slide Time: 06:31)



An interval estimation of a real valued parameter theta is any pair of functions L of X and U of X, such that L of X is going to be smaller than U of X and this should happen for all possible samples. So, this should be capital X here when bold bold X and this random interval is called the interval estimator. So, notice that so this L of X is a real number U of X is a real number.

(Refer Slide Time: 07:05)

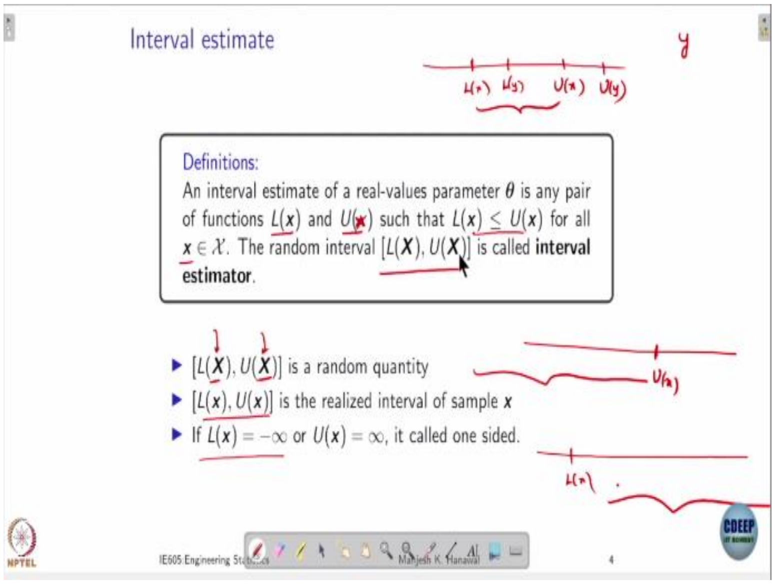


And what you are now doing is I am saying that ok this is your L of x and this is your U of X and now I am giving you this interval. It need not be always interval maybe in some situation we may end up with sets. Now, couple of things to note one is this L of X U of X that interval is a random quantity why is that because that depends on your random sample itself.

Like let us say for one sample X, I have this interval L of X and U of X and let us say I have got another sample may be from the same population itself for that I may have this quantity. So, this interval themselves are random quantities and when I have a particular realization of this random sample which we denote is X then that is for that X I have this realized interval.

So, that is what like I mean for different possible X I may get different different intervals that is why that random interval that interval itself is random and for X particular X I will get that realized value. However, in some cases we may not be interested in both L of X and U of X one side we can just say that okay L of X may I can take as minus infinity or U of X, I may take it as plus infinity so in this case I will be just getting one-sided confidence intervals.

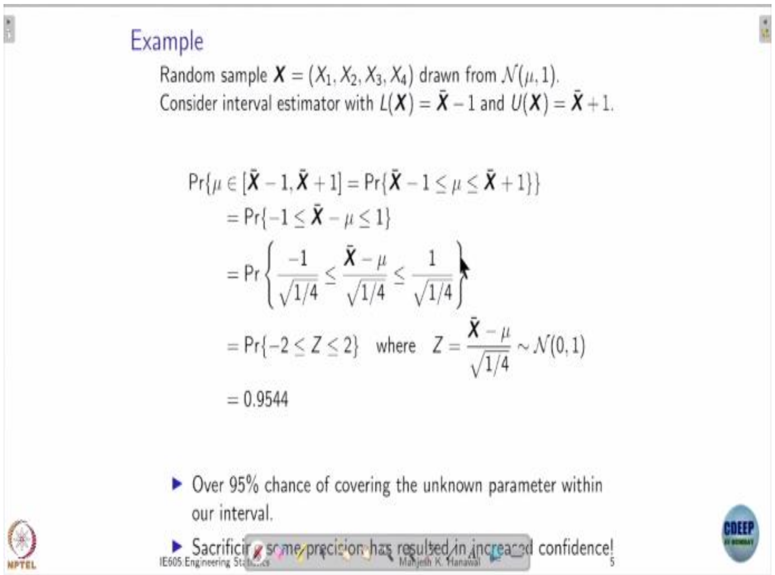
(Refer Slide Time: 09:00)



That is like here if your L of X is minus infinity like I will have only U of X. So, then entire thing from minus infinity U of X that is like my one-sided confidence interval and similarly if you set the other way around this and you just set L of X this is like another one sided. But mostly where we will be interested in something where both I have a lower bond and an upper bond that will govern my confidence interval.

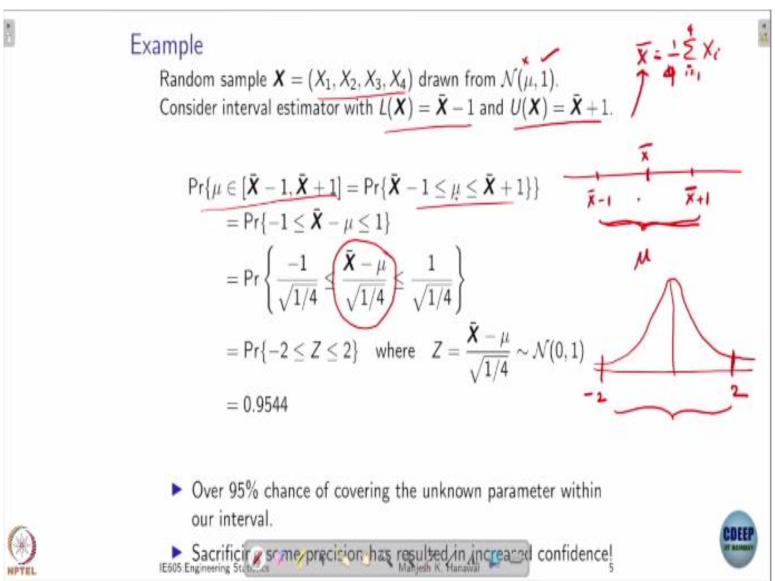
There are other variants also possible for example why you have to take close interval at both ends you may take open at one of them or a mix of both. It depends on how you want to define and what is your application.

(Refer Slide time: 09:45)



Actually, this example we discussed in the last class I am just repeating it here again. Suppose let us say I have random samples drawn from Gaussian distribution with unknown mean mu that mean is not known.

(Refer Slide Time: 10:08)



But, let us say variance is fixed at one. And for simplicity I am just working with four samples. Now, as usual my X bar is my sample mean in this case 4, but now instead of directly taking this as my point estimator I have defined my L of X as X bar minus 1 and X U Bar U of X as X bar plus one.

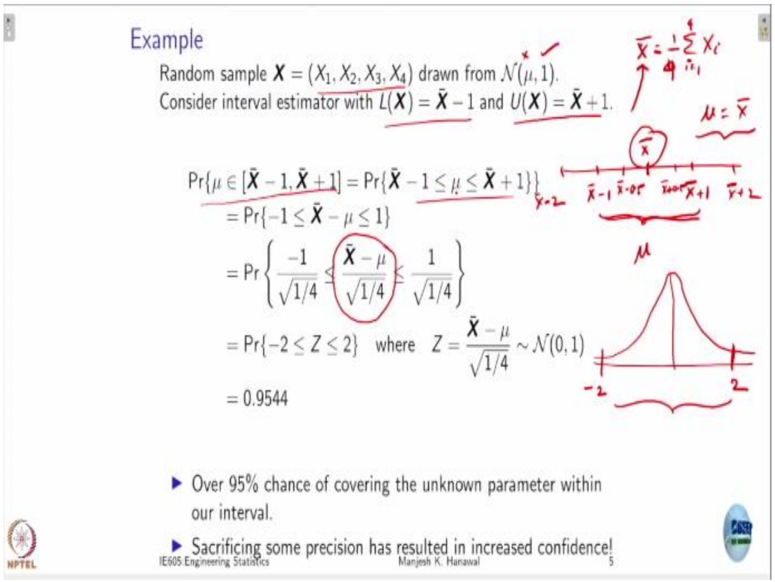
So, basically now instead of taking this X bar as my estimated value for my mu I am taking X plus 1 sorry X minus 1 and X Plus 1 as the value. Now let us compute what is the probability that my true value mu lies in this interval. Can we calculate that? So, the way to calculate that is yes, the probability that mu is going to be lie in the in and 12 X minus 1 X bar minus 1 and X bar plus 1 this question is same as asking. X lies between these two values that is that mu is something between these two extremes.

Now, I have simply manipulated this I have bought X bar minus mu in the middle and the left side now I am left with minus 1 and plus 1, that is just a simple reorganization I have done. Now, in the next step I have divided throughout by square root of 1 by 4. So, Y is specifically square root of 1 by 4 to make sure that this guy in the middle becomes a standard normal. Now, now that standard normal quantity I have represented by Z and now the left side is minus 2 to plus 2.

So, now I have the standard Gaussian we know most of its things about its tail behavior and they are all usually available in the tables are a simple program will immediately compute and queue. This value what you can figure out that if this quantity is let us say this is minus 2 and this is 2, the probability that it is going to be in this range minus 2 to plus 2 is going to be 0.9544.

Now, what you are saying is the true value is going to be between minus sorry a true value is going to be between X bar minus 1 and X bar plus 1, that you can say you can claim that with about 95 percent. What does this 95 percent means? There can somebody quantify it out of 100, like yeah on an average like if I get you give me a lot of samples, I can guarantee that 95 percent of the time I can guarantee you that the value I have provided is going to the interval I am going to provide here we are talking about interval the interval I am going to provide that is going to capture the true parameter.

(Refer Slide time: 14:31)



So, what we have done is instead of giving one point we have now given this interval by doing so what I have done. I have compromised on the precision like exactly I am not saying what is the value of unknown quantity mu I am saying okay instead of saying exact this is exactly it is I am saying no, it is in this interval but I am saying that is with more confidence had I said that okay your mu is exactly X bar, then I know that my confidence is very low I mean almost zero confidence I.

But if I say that mu is going to be between X bar and minus 1 and X bar plus 1, now 95 confident. So, a quick question instead of X minus 1 and X bar plus 1, had I made it X bar minus 0.5 and X bar plus 0.5. What would have happened to my confidence it would have increased our come down? It has decreased and if I made it like a X bar plus 2 and X bar minus 2? Increase so like see like I am kind of trading off here by being more and more loose and being more and more confident.

And if I want to be tighter and tighter, I am lesser confidence. So, you have to decide you want to be loose or tight. So, lose means more confident tight means less confident, so you have to trade off where to hit a good balance again some analogy there like, if tomorrow Meteorological Department says there is a chance of rain 30 percent and some other other one gives that okay there is a chance of rain that it is going to rain 10 percent.

Which one as per you is better I mean who may who is making a better prediction? The one guy is telling that it is going to rain 10 percent or the guy who is going to claim he is going to make it is going to 30 percent. That is the confidence right 10 percent, 30percent. Is it not confidence? They are going to say it is going to rain with likely 30 percent that means they have only 30 percent confident about rain or not train is our decision now.

10 percent or you are going to say that with 10 percent or 30 percent that is your confidence. Is that is enough for you to decide who is better in this case. What else is missing? Maybe we had to enlarge the details about this problem to discuss that let us not get into that now.