Engineering Statistics
Professor Manjesh Hanawal
Industrial Engineering and Operations Research
Indian Institute of Technology, Bombay
Lecture 48
Interval Estimation

(Refer Slide Time: 00:20)

So, now some more definitions. Now, you have introduced this intervals interval estimators. So, we can define some terms related to this how good that interval is. Suppose if you have an interval estimator given by this lower function L of X and upper function by U of x for a particular parameter theta its coverage probability is a probability that it covers the true parameter.

You understand this. So basically mathematically we are saying what is the probability that the true parameter lies in this interval. So, this is what we are going to call it as coverage probability with what probability you are I mean you are covering. Now, similarly if I have an interval estimator given by L and U functions for a particular parameter theta its confidence coefficient now is the smallest value of its coverage probability across the parameter set.

So, we just said this is the coverage probability and my theta can comes from any value over my parameter space right that is something which is not in my control theta can be any value up.

Where theta like now I have to look at the worst case, so I will take the smallest value of this coverage probability over my parameter space as my confidence coefficient. So, couple of things to note here, so when I write a probability here what is random here is the interval random or theta is random?

It is the random quantity is this not the theta here. Theta is fixed you should be clear like earlier when we are studying the basic probability we said what is the probability that X belongs to let us say some interval A and B. There let us say if let us say X is uniform let us say 0, 1 and if I take A equals to 0.1 and B equals to let us say 0.5. So, A and B were given there and X is my random quantity and I was looking into and how I computed this probability I just computed this probability like one between 0.1 to 0.5.

For the uniform distribution this is how Phi computed. But here theta is fixed what is changing is L of X and U of X depending on your sample. And whenever I write this we actually mean that probability that my theta is L of X and U of X is the meaning. Often this interval estimators we have along with their conf along with their measure of confidence one confidence measure we have is the confidence coefficient.

Together with that they are called as confidence interval. So, you need to first give me confidence interval and you have to tell me the measure of the confidence how you are going to measure the confidence. Maybe, one I have defined one measure here I am interested in the worst case like the smallest value of my coverage probability. Maybe somebody has other way of measuring it. For example, I may just say that okay I am other possibilities I may say that theta is itself is coming from some distribution.

I do not know what distribution it is and then I may be interested in expected value of this probability. So, here you are making multiple things this interval is random and you will compute this quantity for a given theta but then you want to take expectation over all possible values of theta you get that could be another measure of your confidence.

But you will not get into that you will be only interested in the confidence coefficient here which is the worst-case coverage probability together when you have this con interval estimator with this confidence coefficient measure that has given to you, you are going to take it as a confidence interval.

Let us see that this confidence coefficients I have of course does this quantity depend on theta yes or no? Does this confidence coefficient depends on theta? How should it depend on theta? You have taken among all the thetas you are taking already smallest value of that right you are already taking infimum of them then where is the theta coming into picture here.

(Refer Slide Time: 06:53)

So, these are probabilities like see that probability is going to be between 0, 1 and you are looking into the smallest value of that probability. So, and you are taking the smallest value across all the set like you are already looked into all possible values of theta that theta set the parameter set.

So, it will not depend on any particular theta here. So, then let us look into the coverage probability itself how coverage probability should definitely depend on the theta right because this is for a given theta. Whereas, your confidence coefficient is or all possible values of theta. So, let us see how to how the theta affects our coverage probability and should it depend on how we define our interval estimators.

So, here is one example of calculating this thing on some estimators. Suppose let us say you have a random sample which is drawn from uniform distribution with parameter 0 and theta and theta is obviously unknown. So, what could be a potential estimator for theta? So, I have written here one natural estimator for theta I simply take the max value of that. Can that be a good estimator for theta? But that is still a point estimator.

Y here is a point estimator. So, now let us try to come up with some couple of interval estimator based on this. So, what I have done here basically let us say this is my X1 and I let us say I have already ordered them X(1), X(2), X(3) and let us say X(n). Actually, I should have put them in the in in terms of our order statistics notation it is the notation. X(1) is the first order statistics X to the second order statistic. And now what I am taking is I am taking this to be the value of theta in the first case and I am calling it as Y.

But instead of that what I will do is ok now tell me this Y will be most of the time can it exceed theta or it will be lesser than or equals to theta? It is going to be less than or equals to theta.

(Refer Slide Time: 09:52)

Because maybe my my range is this much 0 to theta and the samples I have observed here and maximum value Y I am going to get is somewhere here. So, thinking along those lines maybe I know that whatever the Y I get my actual value should be slightly larger than that. So, keeping that in mind maybe I can take my intervals to be something here AY. Where A is something greater than 1. To the right of this and maybe another region let which we call it as BY. I am now defining this to be my possible one estimator does this interval estimator make sense.

Right now, I am not telling what is A and B. But let us say I A is just simply greater than 1 so that I am scaling what are the Y I get by that number and look something on the other side also this is one my possible interval estimator. So, then white scale them you can offset them also so for that I will consider another possible estimator instead of scaling like this one I am just taking that to be Y plus c and maybe Y plus d I am just offsetting them.

And here naturally I want this C to be greater than or equals to 0 and D of course has to be greater than C but maybe one more constant we can put that as D has to be less than theta. D cannot be itself larger than theta. I mean we will be interested in D where it is way less than the value of theta.

(Refer Slide Time: 11:45)

Now, let us try to compute what is the coverage probability of these two estimators we have. So, this is estimator 1. So, probability that theta belongs to aY, bY interval. This I can rearrange it as sorry this should have been capital Y I have rearranged it by dividing them and I will get this. So, why is max of X1, X2 up to Xn. If X1, X2 are all uniformly distributed do we know the distribution of Y. Did you people compute it before? What is the PDF of Y? n into I just writing it here this is going to be by theta and.

(Refer Slide Time: 13:06)

If Y is between 0 and theta and a 0 otherwise. And then T is Y by theta. What will be its distribution? What will be the distribution of T? Just by scaling you can again check that I am just is going to be n minus T, if T is between 0 and 1 and 0 otherwise after by scaling this. Now, can I compute this probability T I already have its distribution this is going to be integration of T which is nt, n minus 1 between 1 by b and 1 by a dt.

This is going to be what 1 upon so you know how to integrate this right this is going to be what simply 2 to the power t to the power n and when I put T equals to 1 by a I am going to get this quantity n minus 1 by b which is n. Did I make it correct? Let me cross check lower limit is 1 by a. Is this can I do sanity check. B is this is always going to be positive right because B is greater than A. So, this quantity is going to be smaller than this quantity and so this is 5.

Now notice that this quantity here does it depend on theta? This quantity does not depend on theta. So, in this case the confidence coefficient in this case it does not depend on theta. Irrespective of what is theta this is always you are going to get. So, confidence coefficient here is coefficient for interval estimator 1 is simply this quantity 1 by a to the power n 1 by b to the power n. Now, let us look into estimator 2. Interval estimator 2. So, what is that that is going to be P of theta, theta belongings to Y plus c comma Y plus d.

So, this is nothing but probability that Y theta is less than or equals to Y plus d less than or equals to Y plus c. This further I want to simplify such that should I get it in terms of I think it is again try it better to get it in terms of T so that. So, let me divide it by T so this is going to be dividing here could be an issue what we can do is. No, we can do that just a minute okay let us not worry about time being theta what was our you said uniformly between theta what is the parameter space of theta?

Did we say anything about where is the theta? We just say theta is a para theta like range. So, theta is a positive quantity here right because we said uniform 0 theta right. So, Theta has to be naturally greater than 0. Then only it makes sense. So, let us find then let us it is no issue in dividing by theta throughout C by theta is equals to 1 is equals to Y by theta plus d by theta. This is fine no issue in dividing by theta. So, now let us get into this into Y by theta.

So, now if I want to write Y by theta here. So, now I know that Y by theta is upward bounded by 1 minus c by theta and lower bounded by 1 minus d by theta. And I already know what is the

distribution of Y by theta. It is this so this is nothing but now the lower limit is 1 upon d by theta 1 upon c by theta and the distribution I have is T into sorry n into 3 to the power n minus 1 dt. So, what this is going to give me?

This is going to give me 1 upon c by theta minus 1 upon d by theta to the power n. Now, you see that this quantity interval estimator 2 coverage depends on theta. Now, let us do couple of sanity checks. Now let us take estimator 1. If I let n go to infinity what should happen to my coverage probability of estimator 1? If I let n go to infinity, so ideally like if I have large number of samples I should be able to do a better estimate.

Large number of samples that I should be able to give you I should be able to exactly give an interval which covers and I should be more get more and more precise.

(Refer Slide Time: 19:31)

So, as n tends to infinity this term goes to 0 this term goes to 0 because both of them are less than 1. 1 by a is less than 1, 1 by b is also less than 1. So, coverage probability goes to 0 for as this is as n tends to infinity, as and for 2 for 2. Does this happen for the second estimator also? So, as n tends to infinity this is also going to happen because theta is anyway less than d it has to be less than c. So, both of them are less than 1 so this will also go as n goes to infinity.

But now this still does not matter what happens to theta but on the way if you look into this as theta goes to infinity, what is going to happen to the coverage probability of estimator 2? This is

also going to go to 0 right because if you fix c and d and let theta go to infinity, this get killed this get killed.

Then 1 minus 1 is going to be 0. So, second one irrespective of where is theta if you have enough number of samples it will also give you in good coverage probability. So, let me stop.