Applied Time-Series Analysis

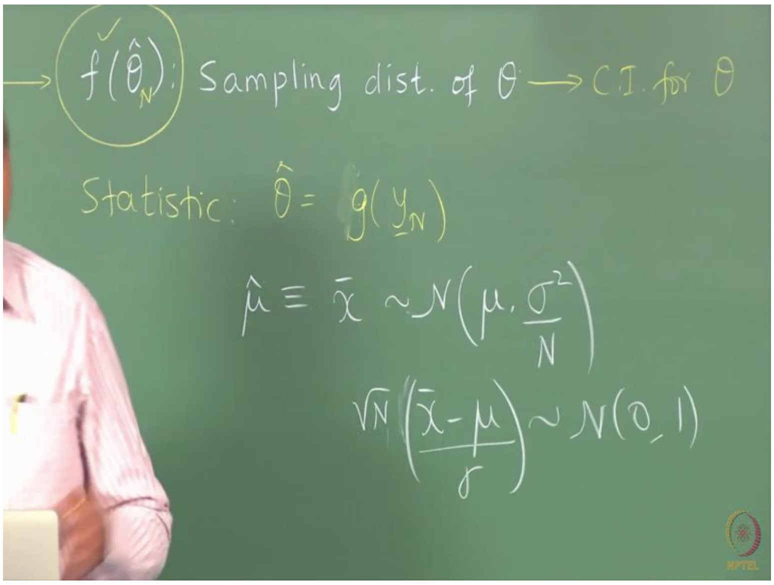
Prof. Arun. K. Tangirala
Department of Chemical Engineering
Indian Institute of Technology, Madras

Lecture – 92
Lecture 40A - Goodness of Estimators 2 -7

Very good morning, we are going to continue our discussion on the properties of the estimators of course, kind of talked about most of the aspects and I said the most one of the most important aspects is to make confidence statement about the truth and to that effect we talked about distributions of parameters estimates. What I told you previously was that in order to construct confidence regions for the truth, we always construct confidence region for the truth; we do not talk about confidence statements for the estimates there is nothing uncertain in the sense that unless you made a calculation error for a given date set and for a given estimation method you can only get 1 value.

What we are not sure about is what the truth is and we want to provide a region in which we believe the truth falls in and today we will see how the distribution of the parameter estimate place a significant role and a critical role in construct in a confidence region and we talked about central limit theorem as a classic example of a result in deriving the distributions of parameter estimates, but the central limit theorem is useful for linear estimators more or less. And you have non-linear estimators then one has to use other methods such as either Monte Carlo simulations or bootstrapping methods. These distributions of parameter estimates that we talked about are also known as sampling distributions.

(Refer Slide Time: 01:58)



It is said to the sampling distribution of the theta, now often in statistics you will come across the term known as statistic, which is the name of the subject itself and this statistic is actually defined as some function of the observations. Any statistic is simply a function of the observations in that sense it is no different from an estimator, we also we have observed many a times at an estimator is also some function of the observations that produces a value which is relevant to a parameter.

So, mathematically there is no difference between a statistic and an estimator or an estimate, but there is a very important and settled different which is that a statistic is any mathematical function that you may construct for your own joy because you have lot of time on your own hand, you want to do something with your observations. You perform some mathematical operation on it, you will get a number but on the other hand when you look at an estimator or an estimate, it is a mathematical operation that you perform on the observations specifically with the objective of estimating a parameter.

So there is a clear cut objective in constructing this function g, when it comes to estimation where as in constructing a statistic it could be any function; obviously, you will not construct any function, but you should remember you will run into these kinds of questions frequently asked on forums as to what is a difference between a statistic an then estimator and so on; that is the prime difference, otherwise both are functions of observations. So, this terms statistic; obviously, you would have also heard in the lectures on hypothesis testing and whatever we are going to discuss today is just a very quick round up of the detailed exposition that we had in the lectures on hypothesis testing.

So, first let us begin with confidence regions, in the lectures on hypothesis testing I would have gone the other way round. I would have actually explain to you what is hypothesis testing and what are the different approaches and then proceed at towards constructing confidence intervals, but we will flip the sequence here so that you see that the confidence regions here have a larger role to play because the starting point for us has been estimation; where as in those lectures the focus from lecture 1 has been on hypothesis testing. Here we are coming from an estimation view point therefore we want to conclude that estimation excise with a confidence statement.

Essentially constructing confidence region is very easy provided you know this f of theta hat. If you know this then it is just a 2 step process for any estimating almost I mean does not change much. So, the first step is given f of theta hat in fact, specifically we talk about f of theta hat subscript n indicating that we have obtained this estimate from n observations. We construct a probabilistic interval and what we mean by probabilistic interval is we would like to for example, specify region in which we believe 95 percent of the times estimates would show up right and it could be 95 percent, it could be 99 percent you specify that and generally when you want to construct for example, a 95 percent confidence interval, then first you would begin with a 95 percent probabilistic interval for theta hat from which you would derive the 95 percent confidence region for theta.

I will just show you this standard example on the mean, but before we do that an important point to keep in mind in constructing a confidence region is, there are two aspects one is degree of confidence and the other is the width of a confidence interval. Normally we would like to make a highly confidence data, ideally we would like to make a 100 percent confidence statement about the truth of course, but the unfortunate part is when we attempt to raise the degree of confidence about the region in which we believe the truth is; the interval also widens.

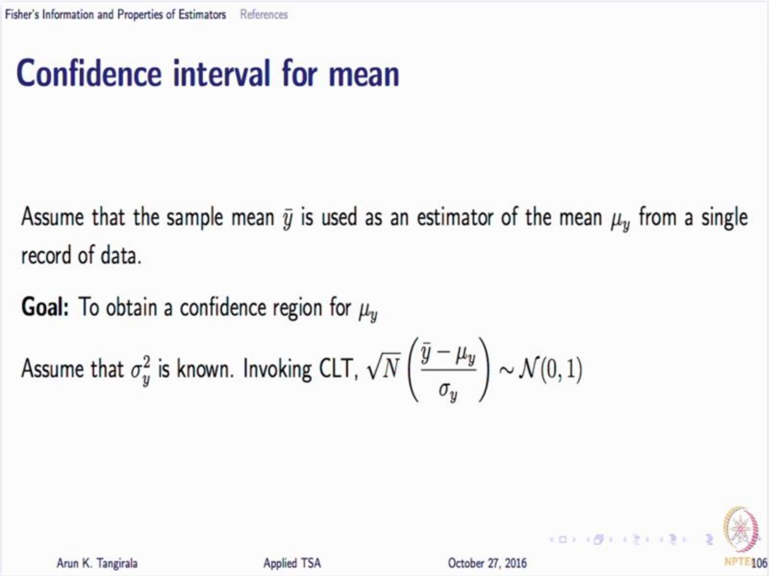
We would ideally like to make a highly confidence statement and obtain a very narrow interval like I really would like to pin point, but that is not possible, but we would like to have as narrow interval as possible, so that the truth is you know the truth has been captured in as fine way as possible. Unfortunately these are confecting requirements as I keep raising the demands of the degree of confidence, the width of the interval also takes a biting.

So, there are many examples that can be given as I always say suppose you have to ask what about the students were absent today, you pick 1 of the students and you ask well where is the student at present. Let us say you happen to be his friends, you would say somewhere; now the location is the parameter that you are trying estimate and statement that you would make is well he is somewhere on campus, you know assuming it is a he and well that is a confidence statement that you are making and; obviously, you cannot state that with 100 percent certainty. So, there is a degree of confidence associated with that statement; if you where to answer; make a 100 percent confidence statement about the location of your friend then the best answer that is you can assure it is 100 percent is somewhere in the universe.

In fact, you would go on and say either in the form of mass or energy so that you are completely safe and then you prove you are self to be an engineer worthy of graduation, but obviously, that is a very practically useless statement; there is no meaning to that it. So, obvious right he has to be in the form of mass or energy in the universe somewhere, but what we are looking at is practically useful confidence statements.

Now you face the heat when you are required to make 99 percent or a 100 percent confidence statement and I would like you just confine this region to a finite interval that; obviously, is not possible. So, you would back off and then say well 95 percent; I am 95 confident that he is somewhere here, but what does that confidence 95 percent confidence mean and we will just go through an example and try to understand what is that 95 percent confidence mean.

(Refer Slide Time: 09:35)

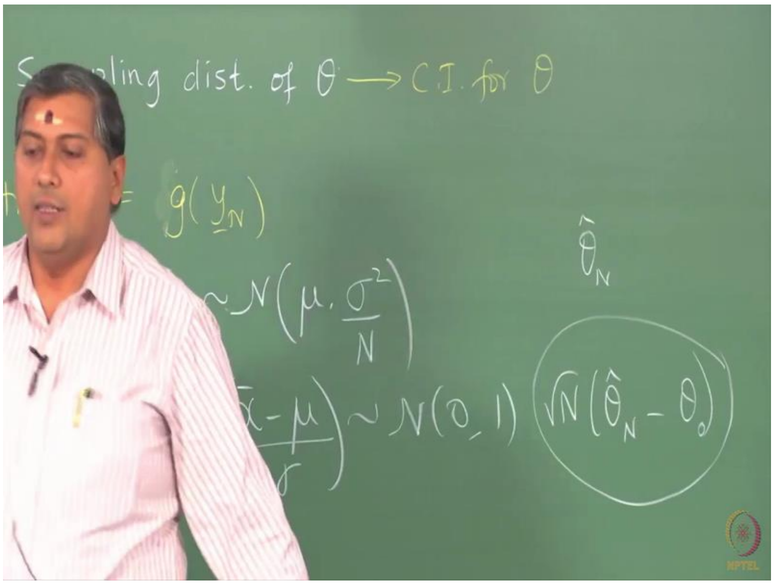


So, let us go through the simple example of constructing the confidence intervals for mean. When you are using sample mean as an estimator - remember the confidence region depends on f of theta hat which in turn depends on theta hat itself. So, this you are staring point from where you would contract f of theta hat and from where you would actually construct confidence intervals for theta. So, we know that the sample mean follows a Gaussian distribution well for large sample case regardless of the joint distribution of the observations and let us assume that we have looking at a white noise process.

So, we know that the sample mean follows a Gaussian distribution, in other words we would say that mu hat or which is let us say x bar or v v bar whatever. So, mu hat which is nothing, but your x bar follows a Gaussian distribution with mean mu because it is an unbiased estimator and variance sigma square over n very good. Now generally this is not how the distribution statements are made, as I said yesterday a lot of times the distribution results are available for large sample sizes. In fact, this is true in general for large sample sizes there is regardless of the distributions that the observation follow.

Therefore, very often you would find the statement that I have shown you on the screen where we rewrite this and say when the root n times x bar minus mu over sigma right follows a standard Gaussian distribution, sigma here is the variability of the process and that is a very important point to observe that is normally statements on distributions are in terms of root n times.

(Refer Slide Time: 11:54)

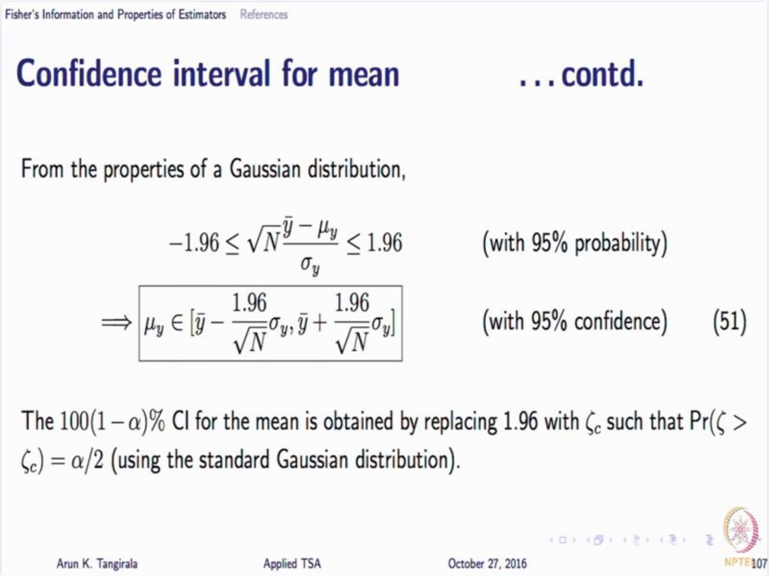


For example you would find in general root n, theta n; theta hat n minus theta naught or theta. Distribution statements are made for this quantity rather than theta hat, although we keep seeing that we want f of theta hat, but this is the case and the reason is if you look at this statement here; obviously, as n goes to 0, we have already seen that sample mean is a consistent estimator or its anyway unbiased its variance strings to 0. When n goes to infinity and the random variable acquires variance of 0; it see as to be a random variable.

So, it becomes a deterministic quantity therefore, it does not make sense to talk about its distribution. Distributions statements are made for variables that are random, not for deterministic variables. On the other hand, if you look at this parameter here that is root n times x bar minus mu by sigma; its variance does not died on to 0 as n goes to infinity, it remains a random variables. So, that is a difference between giving distribution statements for theta hat alone and a standardised theta hat and in this case standardization is this. In some other case, the standardization would be with respect to something else, but always you would see the standardization being made with respect to the truth and with a dominator that has got to do with the variability of the theta hat.

So, to summarize distributional statements are always made in terms of root n times theta hat minus theta n, theta naught by something if theta hat as a Gaussian distribution; if it as some other distributions; some other standardization, but the root n appears so that the random the standardised random parameter estimate remains a random variable; clear any question on it, you are ok fine. If you have any question, you should actually ask it. Now in confidence region construction and in hypothesis testing we work usually with standardised statistic rather than the plain statistic or estimates.

(Refer Slide Time: 14:26)



So, let us now go ahead and construct the confidence region for mean. When we use sample mean as an estimator, in other words if you use median, a sample median or some other estimator or weighted mean and so on, you should come back to this step one where you come back to f of theta hat n so that your probability interval is properly setup. So, we know that for a standard Gaussian distribution; the 95 percent probability region for that random variable is plus or minus 1.96 and that is the statement, there is nothing new there.

But what we are doing is we are now constructing a confidence region from this probability region. So, we are turning the tables around; first we say that this is a confidence region for the standardised statistic and now we say the focus is on the truth theta naught which is in this case mu y of course, I have used x on the board, but it is ok you should be able to follow. So, I have here; my interest is on saying something about mu y, I have already computed y bar; I am assuming that I know sigma y. So, the only unknown is mu y and clearly this says that I cannot pinpoint mu y from y bar and as a result I construct this confidence statement.

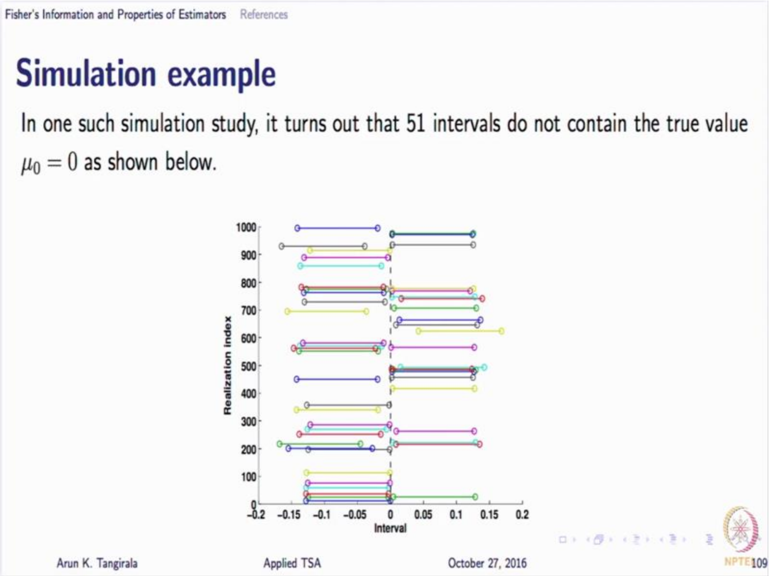
So, must ask yourself and must I have talked about this in video lecture as well; as to why we now change the term from probabilistic interval to confidence region, why is that? Let us see if you are really listened to video and why do we not use the term; same term as we move on to mu yes.

Student: (Refer Time: 16:18).

Good, it is a deterministic quantity therefore, it does not make sense to say with 95 percent probability, mu is between this although my knowledge has uncertainty that has got nothing to do with the mu; that is the philosophy in classical estimation. Where as in Bayesian estimation I say that my knowledge is uncertain about the truth therefore, I will begin my assuming the truth itself is random that it difference because the true parameter that is theta naught is the deterministic quantity, we use the term confident and also there is another reason which is related of course - suppose you where to follow this procedure in practice of course, you will end up with that interval y bar plus or minus 1.96 times sigma by root N. That interval is a guess for your mu and by saying that your 95 percent confident, now comes a interpretation what your implying is if I were to run for example, 1000 simulations; if I had 1000 realizations of the data; each realization would give me y bar and from each y bar, I would construct one interval, this kind of an interval.

Out of these 1000 intervals 950 intervals would have captured the truth and 50 would have missed out the truth and if your rashi of the day is not so great then you may run into that realization which gives you that interval; that does not contain a truth. But it says the probability of missing out on the truth in your interval is 5 percent, it does not say that the probability the truth does not reside in the interval, it does not say that; it say the probability of you missing out on that truth in that interval is 5 percent. So, that is the interpretation of 95 percent confident interval.

(Refer Slide Time: 18:39)



Let me just show you here exactly what I simulated what I just said exactly, I generated 1000 different realizations; each realization containing 1000 observations and then I have y bar constructed from each of those 1000 realizations. Corresponding to each y bar, I have an interval using that formula assuming that n o sigma is a very simple case, but it is ok; it serves to illustrate the just serves the purpose.

Now, what I am showing you on the screen are those intervals which have failed to capture the truth because this is simulation I know the truth. The truth is 0 and you can see I am showing the end points of the interval and 0 is in middle here. You can see that this intervals are either completely to left or to the right which means they have missed out capturing the truth that count if you were to count such intervals, they turn out to be 51; obviously, because I have only simulated 1000 realizations; this number this is now 0.051 right, but if you were to now extend the simulations to many more realizations, it gets better and better and goes to 0.05 these are 95 percent confidence intervals.

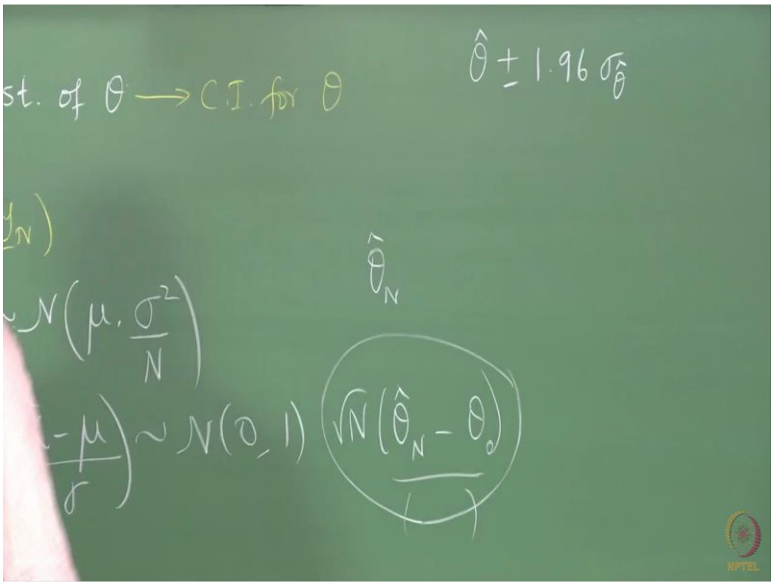
So, now, hopefully you understand what is the meaning of 95 percent confidence. 99 percent confidence interval would mean that you would have how many such intervals that miss out the truth in 1000 realizations.

Student: (Refer Time: 20:10)

10 and you should do that, you should go back yes it is a very simple exercise; if you just go and do it for yourself so that you are convince that you understand this concept very well right. But of course what would happen is when you go to the 99 percent confidence problem, the intervals becomes wider, So obviously, as a interval getting wider, the probability of missing out on the truth; it is like your casting a net like a fisher man and the degree of confidents is rising, the width of the net; the size of the net. So obviously, you would be able to, the probability that you would able to capture the truth this increasing and as you go up and up it will, the probability of missing out on the truth would come down, so this is what is confidence region for you.

So, therefore when you estimate any parameter there is a certain procedure; you choose an estimator and you should know the distribution of that estimate, from where you should able to construct the confidence region. You should not just report point estimate and say I am done you ask me for a point estimate; I gave you that is only half job done. You should report the confidence region typically what one does is first calculates the variance that is your sigma theta hat; not sigma y, sigma theta hat and reports the standard error that is what you would see in many of your ARIMA models that you have estimated, the estimates are reported along with the standard errors. Those standard errors are nothing but sigma theta hat.

(Refer Slide Time: 21:57)

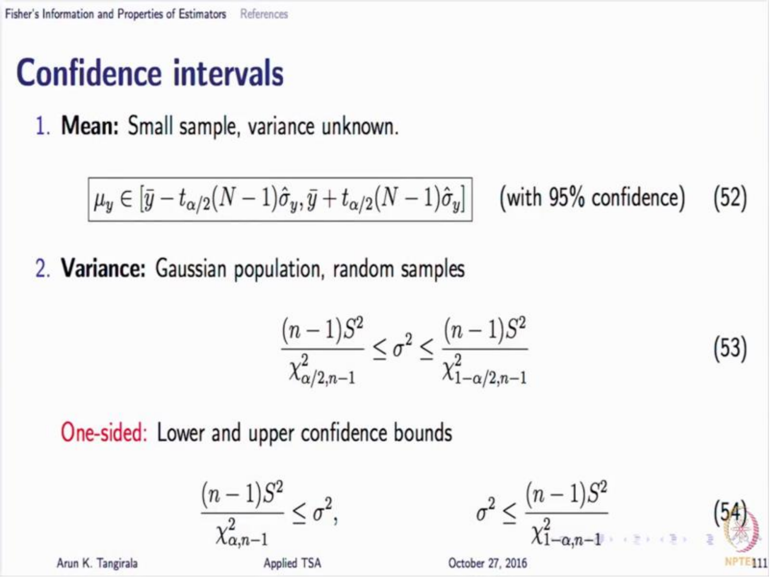


From the sigma theta hat if you assume theta hat follows a Gaussian distribution then the confidence region would be 95 percent confidence region would be simply theta hat plus 1 plus or minus 1.9 sigma theta hat, you may wonder where is a root n gone, but that has gone into your sigma theta hat. The root n appeared in this expression because you have sigma y.

So, if you know that theta hat follows a Gaussian distributions; most of the times the model parameters that your obtaining that is whether it is of AR or MA or ARMA and so on all of them asymptotically follow Gaussian distribution and you should construct this confident region. What is the purpose of constructing this confident region? It is 2 fold; one you are telling me, you are giving me an interval in which the truth is and if you done a good job then it should have actually capture the truth, but there is chance I know 5 percent chance that you would have missed out on the truth right. But I would have done my homework to make sure that I do not give you that kind of realization. So that you know I do not realize it later on, I just make sure that you have a realization which will give you an interval capture the true.

The second purpose of confident region is hypothesis testing and which I will show you very quickly. So, next 5 minute or 7 minute will review hypothesis testing and then we will move on.

(Refer Slide Time: 23:28)



These are some of the intervals for some of the other estimation exercises for example, just now we assumed that the sample size is large and sigma is known. If that is not the case that is you have small sample size, let us say the observations of following out of a Gaussian distribution, but the variance is unknown; then as you know from statistic somewhere you must have heard something called t distribution.

In that case the confidence interval changes slightly where you replace this 1.96 that number for 95 percent. Although it says with 95 percent confidence I have use a generic symbol alpha, we should set alpha to be 0.05 there in this statement. With t alpha by 2 in bracket n minus 1; denotes the quantile in the t chart or in the t distribution of n minus 1 degrees of freedom and alpha by 2 correspond to the quantile there that is your probability there you are looking at the left side probability of it. The t distribution looks pretty much of a Gaussian except that, it has; it does not have as flat tail as a Gaussian.

In fact, as n grows large, you would actually use basically it should not read sigma hat y there; it should read sigma hat theta there is a mistake there I would say. So, as n grows very large t distribution tends to Gaussian distribution, so you should not be even worried of t distribution when n is very large even though the variance is unknown.

So, these are some standard results from statistics very well known for decades and one can construct what are also known as one sided confident regions that is you can ask and this may be required in hypothesis testing where you give a lower bound and bound depending on what you interested in. In equation 53, I have given the confidence region for the variance; if you use the unbiased estimator of the variance, you have to be careful here; we are assuming that we are using an unbiased estimator of the variance. Of course, as n grows very large then it does not matter whether you are using unbiased or biased estimator. Whereas, in 54 we give what are known as one sided confident regions and those are useful in conducting hypothesis test which are called one sided or lower tail or upper tail hypothesis test.