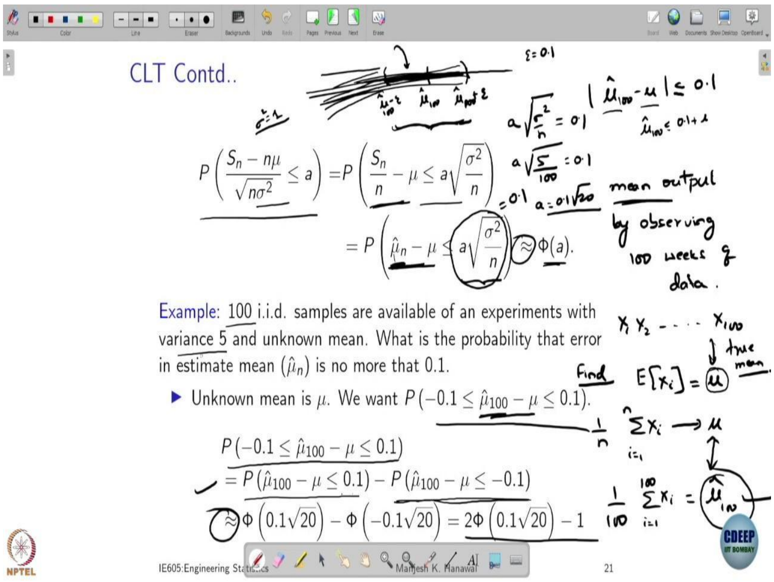
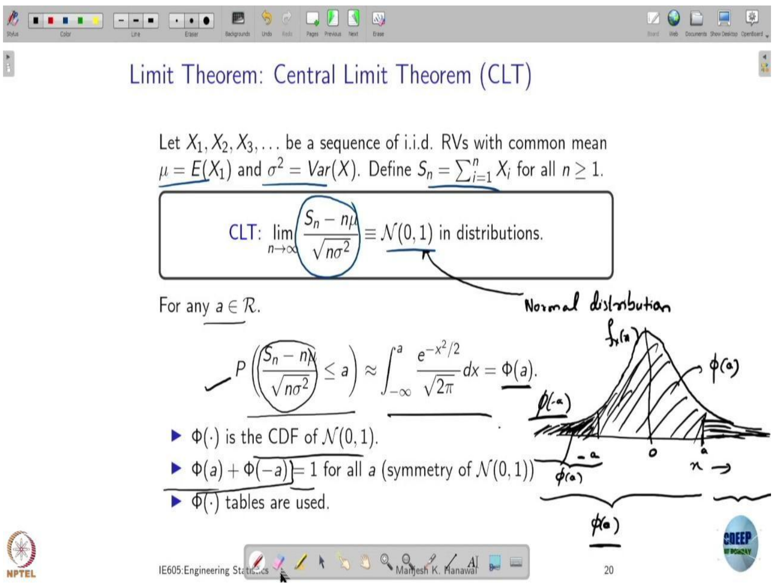
Engineering Statistics
Professor Manjesh Hanawal
Department of Industrial Engineering and Operations Research
Indian Institute of Technology Bombay
Lecture 18
Application of Central Limit Theorem - II

(Refer Slide Time: 0:15)





Notice that this mu hat 100, is this a random quantity? This mu hat 100? Is it a random quantity or not? Is a deterministic quantity? You people, I do not know if you are following what I am saying that mean value mu of a random variable that is a constant, that is a deterministic quantity. Now. If I am going to average 100 random variables, whatever the value I am going to get, will display a random quantity.

Student: Yes.

Professor: If I average 100 random variables, I will get another average random just like what last large numbers said is if you average infinitely many are countably many random variables, which are iid than that average is no more a random quantity that is deterministic, but it is not saying anything about what happens when you average only a finite number of random variables it is only saying when you average countably many it says what happens.

So, when you average only 100 it may be still a random quantity. So, that is why now you are trying to understand this random quantity whatever you opted after average 100 samples, how far it is from that true mean. And that is what notice that that this is the true mean and last large number said that if you average countably many random variables that average is going to be equal to the true mean. But now, when you average only 100 value, you got some quantity and that may not be the true mean.

And now what you are asking is what is the probability that the difference between what you obtained and the true mean is going to be less than 0.1. And because this mu 100 hat is a random variable random quantity you are going to ask the probability what is the probability that the value I obtained by averaging 100 sample that is going to be say equal to the true mean.

Now let us go. So, we are interested in this probability. Now this probability I want to simplify so, that I can write it in a format which I know about. So, I am going to write it as $\hat{\mu}_{100} - \mu$less than or equals to 0.1 minus $\hat{\mu}_{100} - \mu$ less than or equal to 0.01. Is this correct? So, what we are basically looking into is let us say this is you have $\hat{\mu}_{100}$. And now you are going to look into its difference. So, let us say this is like a μ.

So, let us say you whatever you got estimated after 100 rounds, you got a point here, which I am denoting as $\hat{\mu}_{100}$. And now, we want some μ to be somewhere within this interval itself. That is, it should not be less than $\mu_{100} - \epsilon$ and it should not be greater than μ_{100}. If this is the case you know that difference between your $\hat{\mu}_{100} - \mu$ is going to be less than so, this epsilon, in this case, we have taken this epsilon to be got 0.1.

So, if this is the case, then that you are a true mu is going to be lying going to be somewhere within this range. So, what we are so, to do this, what should I be doing, I am interested in my true value is going to be within this portion. So, to do that, what I can do is first thing, ensure that your value is below all this point and from that you subtracted this value from the other point, then you are going to get a value within this range.

So, take the one extreme and from that, you subtract the other extreme, then you will get exactly your value lying within this interval. And that is what we have done here. Now, let us say now, we are simpler.

Now, I want you to go back and use this definition that we have used here or whatever the definition we have now, notice that you have $\hat{\mu_n} - \mu$ here and now 0.1 is equals to a sigma square by n. So, now, but here this entire quantity here to me in this case is equals to 0.1. But from that I need to find out a what is a and then I have to use here. So, here the variance is told to be how much 5. So, let us plug in so, what we have is a from this a sigma square by n is 0.1 a sigma squared is 5 and how much is 100 sorry, how much is n? 100.

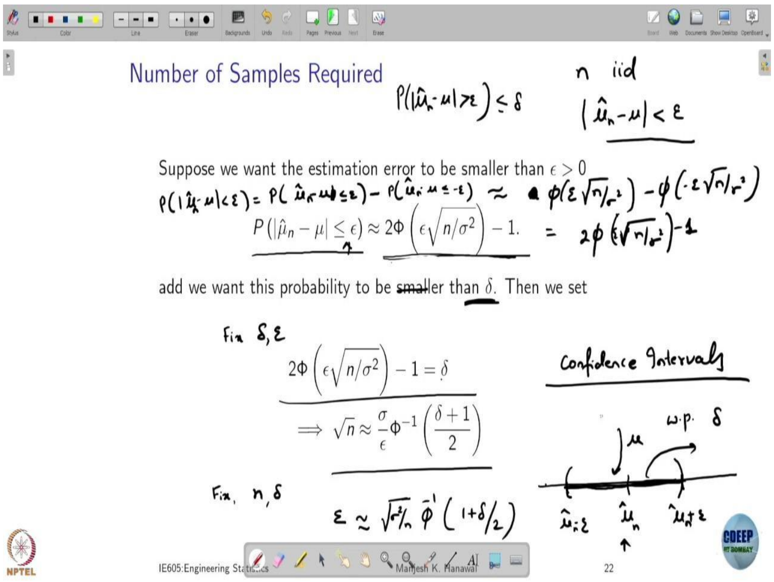
If, i do this, how much is the value of a? Now, this is going to be 0.1 root 20 and that is what I have written $\phi(a)$ is 0.1 root 20. And similarly, this quantity here similarly happens to be $\phi(0.1\sqrt{20})$. And now, you can further do manipulation we already talked that because of the symmetry this relation holds phi of like a $\phi(a)$ plus $\phi(a)$ equals to 1.

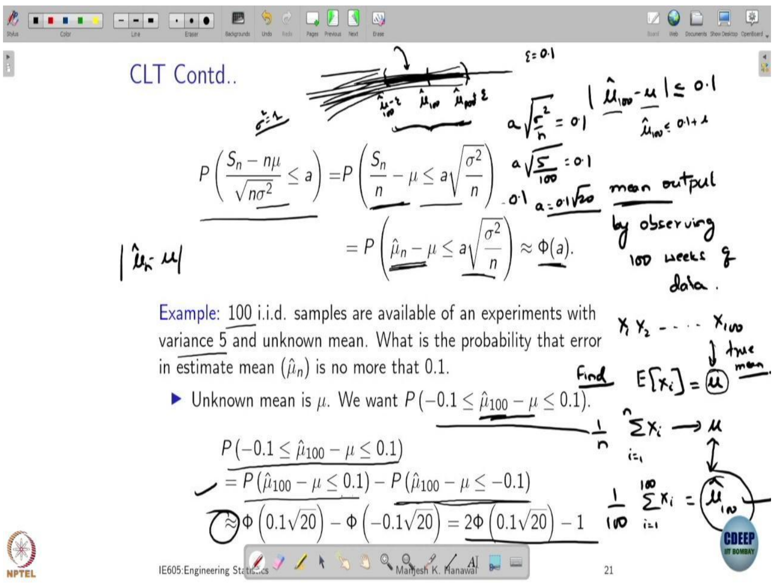
So, you go and substitute this you what I have done is phi of this quantity of subs substituted 1 - $\phi(a)$. So, then if you simplify this you will get this quantity and as I told you often this phi function is available in the form of tables you can look into its value at the point 0.1 square root 20 and you will get the numerical value for this you can plug in and get the value and that will give what is the probability that your error is absolute error is going to be less than 0.1.

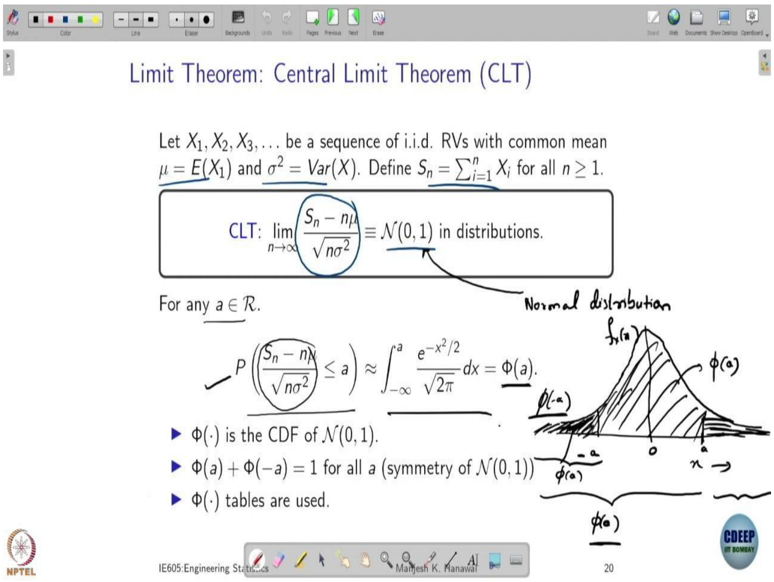
And you see that how now the central limit theorem is playing role and by the way this what it will do is approximate value that is why I put approximate here because this function phi once I say this right this is an approximation here and this approximation becoming more and more good as n becomes larger. And in general if your variance in the samples is small maybe within some 50 to 100 samples, it will be as good as taking as n tends to infinity.

And your approximations become very good because the convergence is actually happens to be very fast when n when your variance is going to be smaller. But of course if variance is larger than you may need more and more samples and approximation may not be good within 50 to 100 samples maybe you need to go for higher number of samples.

(Refer Slide Time: 9:52)







Now, what we know obtained this given a number of samples, what is the probability of error. So, what is this giving us? What does this give us? This basically said that if I have obtained this quantity after n samples the probability that this quantity will be erroneous by this much amount will be approximately $\phi(a)$ that is what we say. And this is basically bounding the error probability this is what we call it as error probability bounding and often this is what we call it as error mu and hat.

Because mu is that true value the mean value, but what we obtained is mu n as hat using n number of samples. Ideally, this should have been equals to mu the true value, but because of n is being not tending to infinity, this is not going to be the same. So, because of that there will be some error in this. What we try to bound is this error being smaller than this quantity we said that that is equals to this probability.

But in reality you may have be faced with the other question. Now, I want to compute something but you are going to specify what should be the error, we are going to specify that error should be this much. Then if you have to guarantee that much error of error, then you may be needing to decide how many samples I need to get. So, that if I average that many samples, my error is going to be, what you want. Error is going to be guaranteed to what you are expected.

Let us understand this, now let us go back to our same example. So, let us say instead of 400 samples, let us say now, I have been given n iid samples. And I have been asked to find the estimate, I am now calling this an estimate. And I have been told that that difference between this and this should be less than some epsilon, this has been told to you but can you guarantee that the error is going to be less than epsilon by any number of n.

So, we know that only when n tends to infinity, my $\hat{\mu}_n$ will be closer to μ. And when n tends to infinity, there is exactly 0, but when n is small, this error can be large. So, if I will give you more and more samples and then you can bring the error make the error smaller, but if not, you cannot guarantee any arbitrary play small amount of error. Now, let us see how to calculate this.

Suppose, somebody has been asked you, I want that my error should be less than epsilon. Now, if they want certain error to be less than epsilon, how many samples you should be asking, So, that you can guarantee what is the error they want? So, let us see how to calculate that value how many samples you should be asking, so, that you can guarantee the error that they are asking for.

We know that this error probability is approximately close to this quantity, I mean, this is I think, I skipped the step, but you can write it, probability that $\hat{\mu}_n - \mu$ less than epsilon, this is equals to probability that $\mu_n - \mu$ this is like less than epsilon minus probability that $\hat{\mu}_n - \mu$ this is less than or equal to epsilon.

This you have with we can write it and if you just manipulate this you will get that this is nothing but, so, the first quantity is clear the first quantity is going to be epsilon phi of epsilon square root n by sigma square. And the second quantity is minus epsilon square root of n sigma square. And if you again use this properties you will get this to be 2 times phi epsilon square root of n sigma square minus 1 sorry minus 1 is outside.

Now, suppose you have been told that the error should be less than epsilon with probability delta and that probability of that your error being smaller than epsilon should be smaller than delta and then what you can do in this case this probability you have you can equate it to delta and from that you can write revert it to get an expression of n and in this case if you manipulate it, you will get that so phi, I am going to take it as phi inverse here simply and you will get this expression.

So, what we have basically able to do in this case is we are able to come up with how many samples I should be getting so that the error that I am going to get is going to be smaller than delta. So, you will see more of this later as we went we will later revisit this topic something called Confidence Intervals.

Let us see how to interpret this delta. Now, I am setting this to be delta quantity this error should be less than or equals to epsilon i want this to be…

Student: (())(17:09)

Professor: See what is the delta for us delta is the probability what do you want this probability to be large or small?

Student: large.

Professor: This probability to be large that my true estimated value to be away from true value that error to be less than epsilon I want it to be very high probability. So, and we want this probability to be actually we should have said larger than, to be larger than this, that is fine. And if this delta is the probability that with which this error should be less than epsilon happen. And we want to set it like this delta and from that you can get this value.

So, usually like if you have encountered this earlier the way to bound is where like people will say that this error should not be larger than epsilon in that case, you want this to be less than smaller than delta, but here we are looking at the other quantity. So, we want it to be larger than delta. So, so, do not get confused by that whether we are asking for smaller or larger that depends on the context whether you will be interested in minima, you want that error probability to be smaller than certain value or not.

So, this is will lead us to something called confidence intervals which will actually revisit later, but now, just to quickly if you have to write this suppose, let us say now, so, here I have fixed delta and I have fixed epsilon and I try to find how many samples are required here fix delta and epsilon you have obtained this. In reality you may be have to face with fix n, the number of samples is already given to you and you have been asked to give me what is and you have been asked to calculate, give the guarantee that your error is going to be smaller than this delta.

Then you need to tell in what range you are some in your estimate lies. So, here, you can solve for epsilon like this. So, let us solve this for epsilon. So, what is the value of epsilon, if I do like this, this is going to be square root n times phi inverse of 1 plus delta by 2. Am I correct? I am just instead of solving for n here, I am solving for epsilon here, for a given n and delta. Now, what is going to happen now is support now let us say you have obtained $\hat{\mu}_n$ here.

Now what you have been able to achieve as you just called $\hat{\mu}_n + \epsilon$ at $\hat{\mu}_n - \epsilon$, you are now able to give an interval around this point mu n hat. And now you are going to say that the true value lie somewhere in this with probability delta. So, what now you are saying is you are giving an interval around the point $\hat{\mu}_n$, and you are going to say that your true value is going to lie within that interval. And this is often called the confidence interval of your point.

We will revisit this later, but as you see that this confidence interval can be computed using all the central limit theorem results, we are basically using this a phi of a function phi of a function arise because of this approximation, where we use the central limit theorem to make this approximation.

And later we will see, using this intuition that central limit theorem has given us one can go ahead and construct confidence intervals without requiring to make this approximation we can compute the exact values.