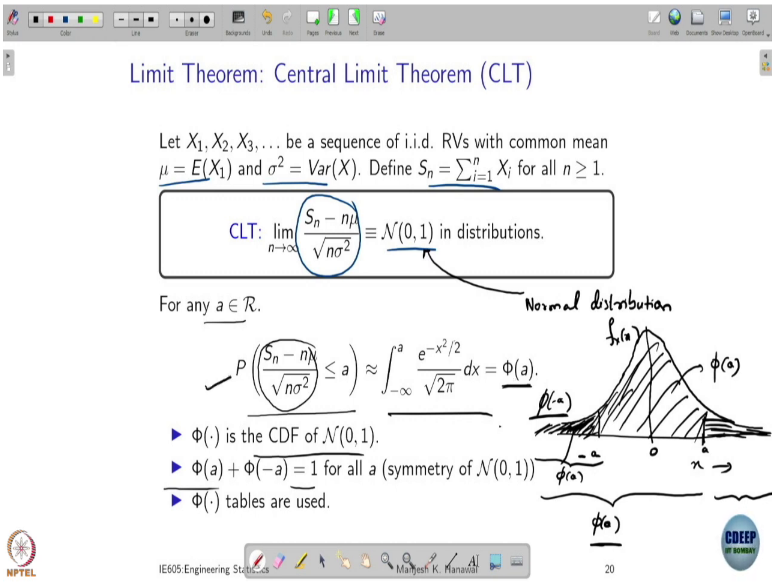
Engineering Statistics
Professor Manjesh Hanawala
Department of Industrial Engineering and Operation
Indian Institute of Technology, Bombay
Week 4
Lecture 17
Application of Central Limit Theorem - 1

(Refer Slide Time: 00:18)



Limit Theorem: Central Limit Theorem (CLT)

Let $X_1, X_2, X_3, \ldots$ be a sequence of i.i.d. RVs with common mean $\mu = E(X_1)$ and $\sigma^2 = Var(X)$. Define $S_n = \sum_{i=1}^{n} X_i$ for all $n \geq 1$.

CLT: $\lim_{n\to\infty} \frac{S_n - n\mu}{\sqrt{n\sigma^2}} \equiv \mathcal{N}(0,1)$ in distributions.

For any $a \in \mathcal{R}$.

$$P\left(\frac{S_n - n\mu}{\sqrt{n\sigma^2}} \leq a\right) \approx \int_{-\infty}^{a} \frac{e^{-x^2/2}}{\sqrt{2\pi}} dx = \Phi(a).$$

- $\Phi(\cdot)$ is the CDF of $\mathcal{N}(0,1)$.
- $\Phi(a) + \Phi(-a) = 1$ for all a (symmetry of $\mathcal{N}(0,1)$)
- $\Phi(\cdot)$ tables are used.

IE605:Engineering Statistics Manjesh K. Hanawal 20

In the last class, we discussed about the central limit theorem. So, we said that if you have a sequence of random variable, if you take their sum and do the centralization and a normalization and take the limit as n goes to infinity that limit is no more a constant, but that limit is actually a random variable which is going to have Gaussian distribution with mean 0 and variance 1.

And now, we need to see that how this limit theorem is going to be useful and as we will go later like you will see more usefulness of this but before that I will just highlight few of its usefulness. One thing is, suppose if you take some a real number and want to compute what is the probability that that $(S_n - n\mu)/\sigma^2$ divided by sigma square this is less than or equal to a. And here what is μ? μ is the mean of that random variables that is a common mean.

And now, how to compute this? One approximation we can do is when n is very large when n tends to infinity, this quantity here behaves like a Gaussian distribution with mean 0 and variance 1. By the way, we are going to call it this Gaussian distribution with mean 0 and variance 1 as normal distribution because this comes in use so many times this has been a special name called normal distribution.

We know that when n tends to infinity this quantity is a random variable with normal distribution. But when n is not tending to infinity when n is some finite quantity, we can assume that it is almost a normal distribution and because of that we can get some approximation. So, what is this? I know that this can be treated as integration of this PDF. What is this? This is the PDF of?

Student: Normal distribution.

Professor: Normal distribution. And I am just integrating it between minus infinity to a and that will give you this probability. And again, the function the integral of this form where I am going to integrate this quantity the PDF of normal distribution minus minus infinity to a this also appears many times to us we are going to encounter it many times. Because of this, this has also been given a special notation called $\phi(a)$.

For us $\phi(a)$ means, it is a integration of my CDF, sorry PDF of normal distribution in the interval minus infinity to a or I can say that phi of a is nothing but the area under the normal distribution till the point a. So, if you are going to have this, this is my, so my, is going to look like this. This is 0 and this is the mean value and the width of this PDF is going to decide its variance but we are going to say suppose let us say a is here.

This area till this point we are going to call it as $\phi(a)$. On the other hand, let us say if a is here, then we are going to consider all the area till this point, and in this case, this is going to be called phi of a. So, basically as you see $\phi(a)$ is nothing but this is basically CDF of your normal distribution, phi of a is the CDF of normal distribution computed at point a. Now, this function $\phi(a)$ has one nice property that if you are going to take a and minus a.

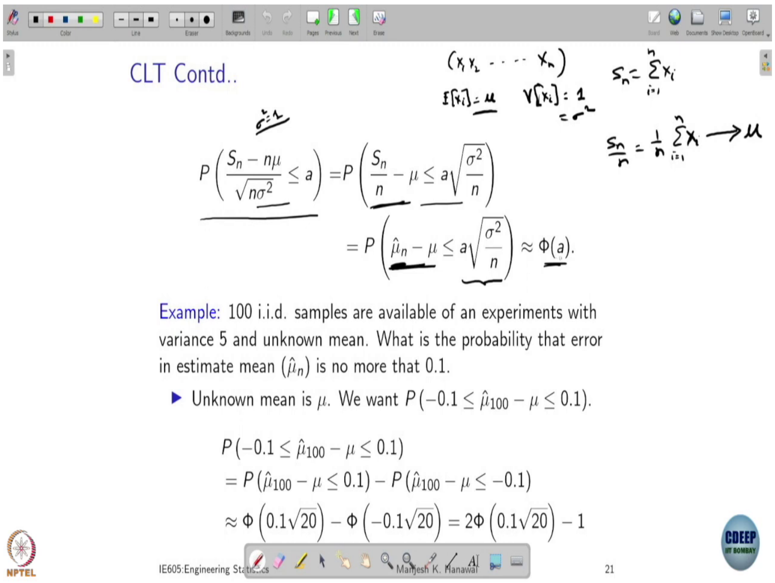
So, suppose let us say this is a and this is minus a, that is a symmetric around minus a. Now, the claim is if you add these 2 areas $\phi(a)$ and $\phi(-\ a)$ that is going to be 1. So, why is that?

Student: Symmetry.

Professor: Because of the symmetry. So, it is, because if you are going to add, if you take area till minus a that is as good as this area above a because they are symmetric. So, if I am going to add till this point $\phi(a)$ like this is my $\phi(a)$ and this $\phi(-\ a)$ is going to be same as all the remaining area. So, we know that area under PDF has to equals to 1. So, because of that if you are going to add this plus $\phi(-\ a)$ that is going to be 1.

Now, in often this comes, by the way these do not have any closed form expression. You cannot write this is like an integral just represented we do not have a closed form expression but you can compute it numerically. So, because of that often this function appears in the form of a tables like you have log tables, log tables for different values of a like that this $\phi(a)$ is given in terms of the tables and you can use these tables to compute probabilities.

(Refer Slide Time: 07:26)



CLT Contd..

$$P\left(\frac{S_n - n\mu}{\sqrt{n\sigma^2}} \le a\right) = P\left(\frac{S_n}{n} - \mu \le a\sqrt{\frac{\sigma^2}{n}}\right)$$
$$= P\left(\hat{\mu}_n - \mu \le a\sqrt{\frac{\sigma^2}{n}}\right) \approx \Phi(a).$$

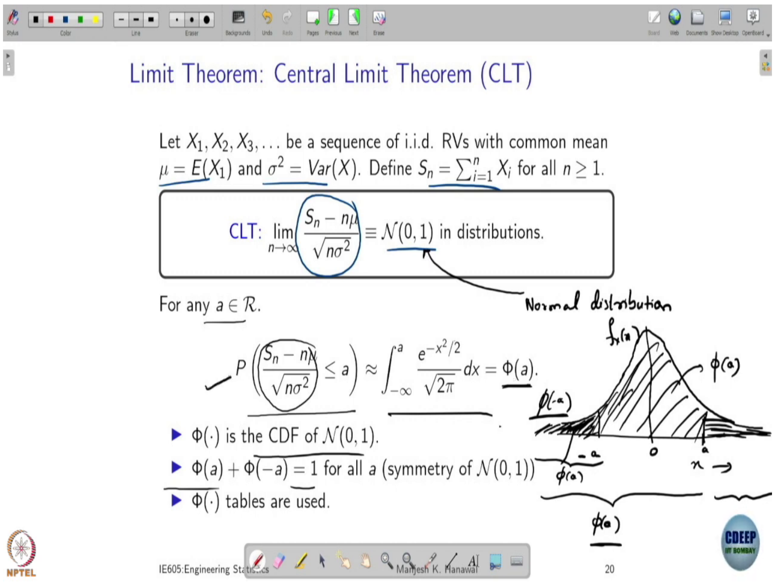
Example: 100 i.i.d. samples are available of an experiments with variance 5 and unknown mean. What is the probability that error in estimate mean $(\hat{\mu}_n)$ is no more that 0.1.

- Unknown mean is μ. We want $P(-0.1 \le \hat{\mu}_{100} - \mu \le 0.1)$.

$$P(-0.1 \le \hat{\mu}_{100} - \mu \le 0.1)$$
$$= P(\hat{\mu}_{100} - \mu \le 0.1) - P(\hat{\mu}_{100} - \mu \le -0.1)$$
$$\approx \Phi\left(0.1\sqrt{20}\right) - \Phi\left(-0.1\sqrt{20}\right) = 2\Phi\left(0.1\sqrt{20}\right) - 1$$

IE605:Engineering Statistics Manjesh K. Hanawal 21

$(X_1 X_2 \cdots X_n)$
$E[X_i] = \mu$
$V[X_i] = 1 = \sigma^2$
$S_n = \sum_{i=1}^{n} X_i$
$\frac{S_n}{n} = \frac{1}{n}\sum_{i=1}^{n} X_i \to \mu$
$\sigma^2 = 1$



Limit Theorem: Central Limit Theorem (CLT)

Let $X_1, X_2, X_3, \ldots$ be a sequence of i.i.d. RVs with common mean $\mu = E(X_1)$ and $\sigma^2 = Var(X)$. Define $S_n = \sum_{i=1}^{n} X_i$ for all $n \ge 1$.

CLT: $\lim_{n\to\infty} \frac{S_n - n\mu}{\sqrt{n\sigma^2}} \equiv \mathcal{N}(0,1)$ in distributions.

For any $a \in \mathcal{R}$.

$$P\left(\frac{S_n - n\mu}{\sqrt{n\sigma^2}} \le a\right) \approx \int_{-\infty}^{a} \frac{e^{-x^2/2}}{\sqrt{2\pi}} dx = \Phi(a).$$

- $\Phi(\cdot)$ is the CDF of $\mathcal{N}(0,1)$.
- $\Phi(a) + \Phi(-a) = 1$ for all a (symmetry of $\mathcal{N}(0,1)$)
- $\Phi(\cdot)$ tables are used.

IE605:Engineering Statistics Manjesh K. Hanawal 20

Now, try to get this function in some nice ways. I am interested in probability that this quantity is less than or equal to a. So, I did simple manipulation like S_n/n. And here we know that σ^2, I am taking σ^2 to be 1. So, right now assume that I have this x1, x1, x1, x2, I have these n random variables, all of them expectation of xi is μ and expectation of, sorry, variance of xi I am going to take it as 1 for simplicity. You can, this variance of xi's need not be 1.

Let us say variance is known, the only parameter I am interested is in the mean now. Now, what I will do. Now, Sn is taken to be sum of these xi's. Now, I will do simplification like I will, I want to now, not even like necessarily like let us take even this is σ^2. So, what I have done here is a basically did a manipulation of this. I have written in such a way that Sn minus n and minus mu comes on the left-hand side rest of the terms go on the other side. Can somebody check that this manipulation is correct here?

Student: Yes, sir. It is.

Professor: Now, this Sn by n. What is Sn by n? This is the Sn by n is average of the samples. I am going to call this as $\hat{\mu}_n$. The sum of this xi's divided by n I am going to take it as the estimate of this parameter μ and now I am able to write it as this $\hat{\mu}_n$ - μ. I know that when n goes to infinity, what this quantity goes to?

Student: Going to μ.

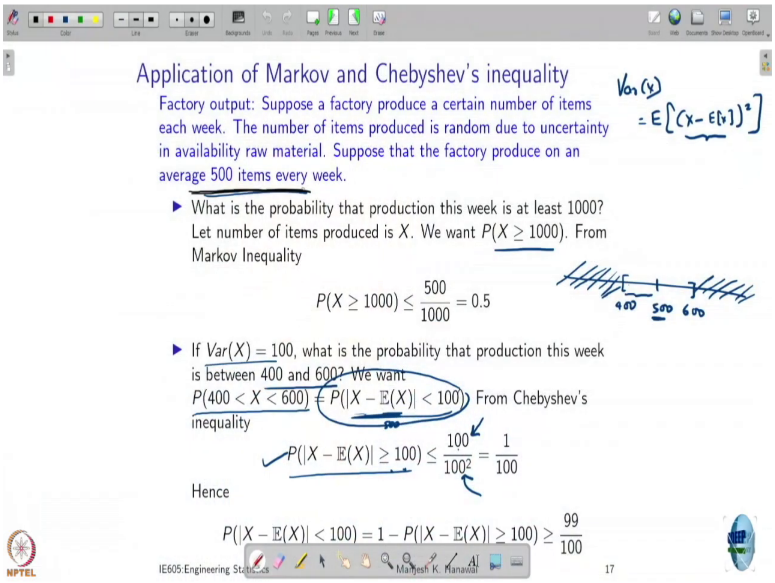
Professor: It is going to go to mu. Why is that?

Student: Law of large numbers.

Professor: Because of that law of large numbers. But now I am only dealing with finite number n, it is not going to infinity. So, that is what I am Instead of calling I cannot call it as mu I am going to call it as mu n hat. And now, I am trying to what, you see that now, I am looking at the difference of the finite average I got with its limiting value which is mu, how is this difference?

Now, we are saying that this difference being smaller than this quantity is approximately $\phi(a)$. And where is this $\phi(a)$ is coming? $\phi(a)$ is coming from our previous relation. So, by using this central limit theorem, I have now a way to characterize how far is mu n hat for a particular n compared to its limiting value? I know that this is going to be approximately $\phi(a)$. Now, let us see how this relation will be useful to us.

(Refer Slide Time: 11:38)



Application of Markov and Chebyshev's inequality

Factory output: Suppose a factory produce a certain number of items each week. The number of items produced is random due to uncertainty in availability raw material. Suppose that the factory produce on an average 500 items every week.

- What is the probability that production this week is at least 1000? Let number of items produced is X. We want $P(X \geq 1000)$. From Markov Inequality

$$P(X \geq 1000) \leq \frac{500}{1000} = 0.5$$

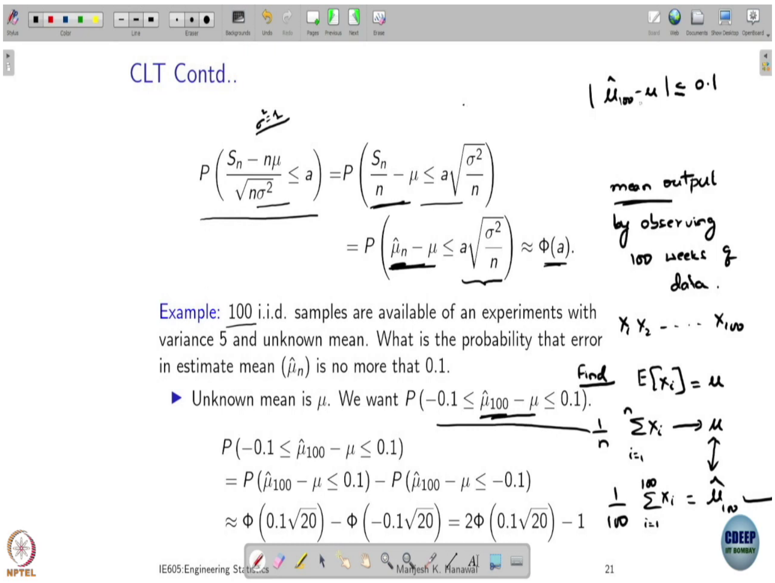
- If $Var(X) = 100$, what is the probability that production this week is between 400 and 600? We want $P(400 < X < 600) = P(|X - \mathbb{E}(X)| < 100)$. From Chebyshev's inequality

$$P(|X - \mathbb{E}(X)| \geq 100) \leq \frac{100}{100^2} = \frac{1}{100}$$

Hence

$$P(|X - \mathbb{E}(X)| < 100) = 1 - P(|X - \mathbb{E}(X)| \geq 100) \geq \frac{99}{100}$$

IE605:Engineering Statistics Manjesh K. Hanawal 17



CLT Contd..

$$P\left(\frac{S_n - n\mu}{\sqrt{n\sigma^2}} \leq a\right) = P\left(\frac{S_n}{n} - \mu \leq a\sqrt{\frac{\sigma^2}{n}}\right)$$

$$= P\left(\hat{\mu}_n - \mu \leq a\sqrt{\frac{\sigma^2}{n}}\right) \approx \Phi(a).$$

Example: 100 i.i.d. samples are available of an experiments with variance 5 and unknown mean. What is the probability that error in estimate mean $(\hat{\mu}_n)$ is no more that 0.1.

- Unknown mean is μ. We want $P(-0.1 \leq \hat{\mu}_{100} - \mu \leq 0.1)$.

$$P(-0.1 \leq \hat{\mu}_{100} - \mu \leq 0.1)$$
$$= P(\hat{\mu}_{100} - \mu \leq 0.1) - P(\hat{\mu}_{100} - \mu \leq -0.1)$$
$$\approx \Phi\left(0.1\sqrt{20}\right) - \Phi\left(-0.1\sqrt{20}\right) = 2\Phi\left(0.1\sqrt{20}\right) - 1$$

IE605:Engineering Statistics Manjesh K. Hanawal 21

$\sigma^2 = 1$
$|\hat{\mu}_{100} - \mu| \leq 0.1$
mean output
by observing
100 weeks of
data.
$X_1 X_2 \cdots X_{100}$
Find $E[X_i] = \mu$
$\frac{1}{n}\sum_{i=1}^{n} X_i \rightarrow \mu$
$\frac{1}{100}\sum_{i=1}^{100} X_i = \hat{\mu}_{100}$

Now, let us go back to the example we studied in our last class where we talked about this factory output. So, the factory is producing certain number of items, and its weekly average is about 500. Now, let us say I said earlier, that the average, I assume that average of 500 items every week, that is the mean value. But mean value, suppose let us say you do not know and now you want to find out what is the mean value, mean number of items produced from my factory.

How you can do it? You can get every time you observe. Today, how many are generated, how many outputs came out from factory like that every day, you get 100 samples. Let us say let us say these are all 100. You observed it let us say what we call this as a weekly, so we are talking about weekly output of the factory.

Let us say you observed for 100 weeks, what is the output from the factory? And now, you are going to call them as 100 samples and assume that, that is the same factory. The factory is going to generate these outputs according to the same underline distribution. So, they are all identically distributed. And let us assume that they are also identical. Every week, a factory, factory like started a fresh so that the effect of the output from one week is not influencing the effect what is that is produced in the next week.

Now, from this 100 samples, you want to generate, you want to identify the mean value that is output by this factory every week. So, what we basically are trying to do is here, trying to find out the mean output, mean output by observing 100 weeks of data. 100 weeks I monitor what is output which I am denoting it as let us say let us call them x1, x2 up to x100. Now, I want to find out what is that mean value?

That is, I want to find out find what is the exact value of xi. This is some value μ, which I do not know I want to find out. One thing I know is if I take this value i equals to 1 to n divided by n, this goes to mu but unfortunately, I do not have n going to infinity I can only take i equals to 1 to 100 xi and divided by 100. I will get this let us call this value as $\hat{\mu}_{100}$. Now, I want to know that how much is this from the true value mu.

So, that is what I am trying to find out $\hat{\mu}_{100}$ - μ will this be, this difference will be below 0.1 and greater than minus 0.1. I want to now calculate this probability. Whatever I estimated, what is the probability that the difference between mu 100, $\hat{\mu}_{100}$ - μ, this is going to be less than or equals to 0.1.

So, alternately what I am asking the question that the mean value I obtained by averaging these 100 values will it be within 0.1 error of the true value? I want to ask this question. Now, and you want to ask what is the probability that the error is going to remain within 0.1.