

Hardware Security
Prof. Debdeep Mukhopadhyay
Department of Computer Science and Engineering
Indian Institute of Technology, Kharagpur

Lecture – 28
Power Analysis- IV

So, welcome to this class on Hardware Security. So, we shall be continuing our discussions on hardware security and in particular we shall be discussing about Power Analysis.

(Refer Slide Time: 00:25)



So, I will be trying to talk about in particular noise as we were discussing in the previous class and we will try to see how we can statistically analyze the component of noise.


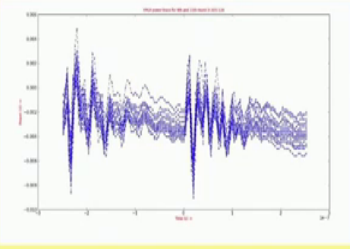
(Refer Slide Time: 00:37)

Quality of Measurement and Noise

- High quality of power traces captured is central to the accuracy of DPA.

AES-128 Encryption with same plaintext and key, but resulting in different power traces.

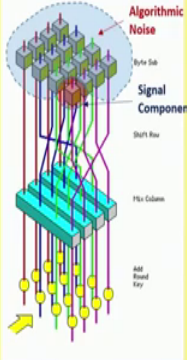
These fluctuations are due to **electrical noise**, caused by noise due to power supply, clock generator, conduction and radiation emissions from the components connected to the device under attack.



And try to estimate the success rate of a difference of mean attack. So, in particular right as I was discussing in the last class, we essentially have this phenomenon of electrical noise in circuits which is often due to the power supply, the clock generator, the conduction and also the radiation emissions from components which are connected to the device and their attack. And the idea is that because of these presence of noise although you are processing the same operation your ok, that is you are suppose in the context of encryption you are basically taking the same plane takes the same key, but still you will get slight variations in the power consumption.

(Refer Slide Time: 01:21)

Algorithmic Noise



Algorithmic Noise


Algorithmic or switching noise occurs because of contributions of logic cells to the power consumption, which are not under attack.

The power trace corresponds to the total power consumption of the circuit.

However, in the attack we target only a small part (see red circle in fig.) to reveal a portion of the key.

The power consumption from all the other parts (see blue shaded portion in fig.) form the algorithmic noise.

This would be more in a parallel implementation, compared to a serialized implementation.



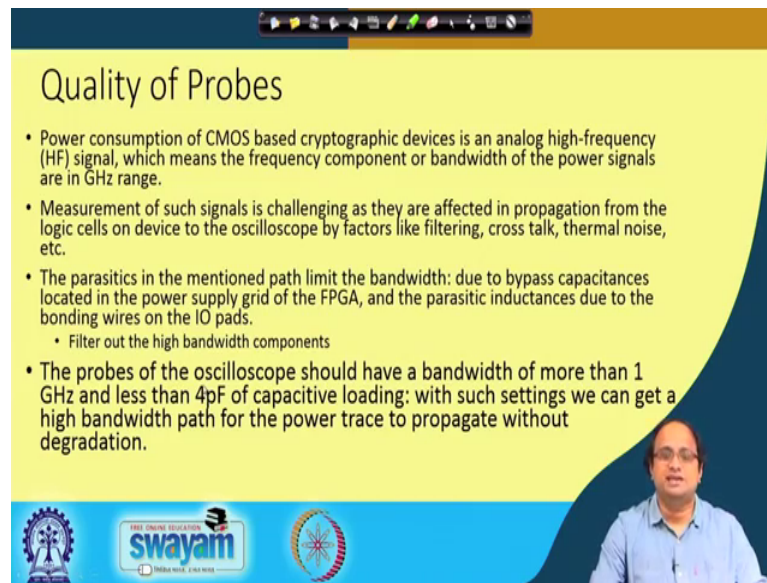
And that is a bother for a cycle and attacker because a cycle and attacker you want to get as much noiseless in a power trace as possible ok. Along with electrical noise, you also have another component of noise which is called as the Algorithmic Noise. So, what is an algorithmic noise? So, let us take this picture of Rijndael or AES. So, this you can easily generalize it to any other target. So, what we do is as we know as we have discussed is that in a power attack setting, we basically try to do a divide and conquer attack right. So, we basically try to kind of concentrate on say a red portion here as shown here and this essentially is a component of your for example, or a part of the state.

So, suppose we are doing a power attack from the plaintext, we can easily do it or just change in the discussion and do it from the cipher text also. So, here the algorithmic noise or the switching noise occurs because of the contribution of logic cells to the power consumption, we just which are not under attack ok. The power trace corresponds to the total power consumption of the circuit. So that means, like although you are say targeting a specific region or portion of your state, but at the same time the other logic component of the logic cells are also toggling, like they are also consuming power and therefore, the power consumption right essentially it takes care I mean has got both these components.

So, it has got a component which I called as a signal whereas, the other components; that means, which are not a part of this attack or we are not you know like considering them in this attack. They are serving to generate noise. This component of noise is called as algorithmic noise because you know like it basically has got something to do with the algorithm itself ok. So however, in I say that however, in the attack we target only a small part see the red circle in the figure to reveal a portion of the key ok.

Now the power consumption from all other parts; so, this blue component for example, essentially forms what is called as the algorithmic noise. So, therefore, you can easily kind of rationalize that this would be more in a parallel implementation compared to a serialized implementation because in a serialize implementation maybe you will just do one s box at a time and therefore, you will not have the other s boxes generating algorithmic noise. So, therefore, right I mean and I will come back to this point later in our discussion.

(Refer Slide Time: 03:23)



The slide is titled "Quality of Probes" and contains the following text:

- Power consumption of CMOS based cryptographic devices is an analog high-frequency (HF) signal, which means the frequency component or bandwidth of the power signals are in GHz range.
- Measurement of such signals is challenging as they are affected in propagation from the logic cells on device to the oscilloscope by factors like filtering, cross talk, thermal noise, etc.
- The parasitics in the mentioned path limit the bandwidth: due to bypass capacitances located in the power supply grid of the FPGA, and the parasitic inductances due to the bonding wires on the IO pads.
 - Filter out the high bandwidth components
- The probes of the oscilloscope should have a bandwidth of more than 1 GHz and less than 4pF of capacitive loading: with such settings we can get a high bandwidth path for the power trace to propagate without degradation.

The slide also features a video feed of a presenter in the bottom right corner and logos for Swamyam and other organizations at the bottom.

So, therefore right as I said at the beginning that for your set up the quality of probes is extremely important and they need to be of good of you know of satisfactory quality. So, therefore, like I mean, so the observe that power as I say the power consumption of the CMOS based cryptographic devices is an analog high-frequency signal which means that the frequency components are of the power signals are quite high or could be in you know like GIGA hertz range.

Therefore, when you are measuring these signals, it could be challenging because you can immediately understand that when you are transferring the power ok; power signal from your target to the oscilloscope, there are there are a lot of things right through which the power signal is travelling and because of that right those there could be components which are lost; there could be high frequency components which are lost.

In particular right, I mean there could be you know like when you are propagating from the logic cell on the device to the oscilloscope. There could be factors like filtering because of these channels. There could be cross talks. There could be thermal noise and the parasitics in this mentioned path essentially it serves as a filter. They basically limit the bandwidth ok. So, for example, you can have the bypass capacitances or you can have you know the inductances, all these things will play role do you know will play a role to basically you know like serve as you know like sort of like a band pass filter.

They will they will basically you know like kind of remove some of the high frequency components.

So, therefore, the probes of the oscilloscope should have a bandwidth of more than 1 Gigahertz as a thumb rule and typically less than 4 pico Farad of capacitive loading ok. With such settings, you can get a high bandwidth part of the power traces to propagate without any degradation. So, this is an important thing and that we need to take care of when we develop our side channel setup.

(Refer Slide Time: 05:13)

The slide is titled "Statistical Analysis of Power Traces" and features a 3D Trace Plot showing a series of overlapping bell-shaped curves. The plot has three axes: a vertical axis ranging from 0.00 to 0.05, and two horizontal axes ranging from 0 to 4000. The curves represent the statistical distribution of power traces. At the bottom of the slide, there is a video feed of a presenter and logos for "swayam" and "THE UNION EDUCATION".

- Electrical noise of a power trace can be characterized by a normal distribution.
- A normal distribution is defined by mean, μ and variance σ^2 .
- Best estimators of these are the average \bar{x} , and the empirical variance, s^2 .
- For some fixed value and operation, the power consumption is fixed, except for the variance that is introduced by electronic noise.

So, again like I said that; so, let us try to look again at the power traces, but now from the point of you know like trying to do a develop a statistical analysis. So, electrical noise as I said of a power trace can be characterized right typically ah. So, it is a very important component like when you are considering the power consumption your power is kind of superimposed right with a component from the noise or the electrical noise.

Now, this component of noise which is the electrical noise typically can be modeled by a Gaussian distribution ok. So, therefore, right you can see that you have an you can basically kind of define it by a mean which is mu and a variance which is sigma square and as you know that from statistical theory that you try to develop estimates for these parameters. So, for example, right average or \bar{x} is you know like is essentially a an estimator which is used for beam and likewise the empirical variance or s square is used as an estimator for variance; your sigma square.

Now for some fixed value and operation right as I said the power consumption is fixed except for the variance that is introduced by the electrical noise. So, therefore, right when I am keeping my plain text and my input key for example, same; then, I would expect that the component which is essentially that is the real signal. That means, the power consumed due to the actual processing of information that is a constant ok; whereas, the other components essentially or you know like the electrical the noise missing essentially tries to kind of develop the variation. I mean the variation is coming because of the noise ok.

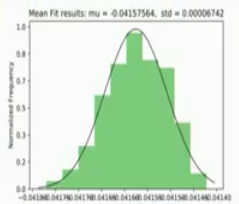
And therefore, and therefore, you will see that if you plot at a specific time instance if you basically try to plot the power consumption, although you are keeping you know like keeping the plane takes same and the key same, you will find out there is a variation and you will see that the variation looks like a Gaussian distribution or a Normal distribution.

So, this is this diagram shows you know like again that you are basically trying to obtain the power traces of AES and you can see that all of them looks roughly same, but there are small differences ok. So, therefore, the question is why is that difference and the answer is because of electrical noise ok. You have got slight differences. Note that this component right is not due to algorithmic noise because the in algorithm noise should not defer because you are keeping the data same in both in all these cases.

(Refer Slide Time: 07:39)

Components of Power Traces

- $P_t = P_{data} + N_t$
 - P_t : Total power at time t.
 - P_{data} : Power consumption at time t due to the data-dependent component.
 - N_t : This is the noise component.
- For the DPA to work, we would be trying to find the point-of-interest (POI) where the distinguishing of wrong key from correct one can be done with more confidence.
- The adjacent histogram shows the distribution of total power at a given time instance for 1000 encryptions with the same input and key.
- The normal distribution shows the distribution of electrical noise.







Mean Fit results: $\mu = -0.04137564$, $\text{std} = 0.00006742$

Points in the power trace follow a normal distribution (also called Gaussian distribution). We write: $X \sim \mathcal{N}(\mu, \sigma)$

The density function describing the normal distribution depends on the parameters μ and σ , where $-\infty < \mu < \infty$, and $\sigma > 0$. We write: $X \sim \mathcal{N}(\mu, \sigma)$, where:

$$f(x) = \frac{1}{\sqrt{2 \cdot \pi} \cdot \sigma} \cdot \exp\left(-\frac{1}{2} \left(\frac{x - \mu}{\sigma}\right)^2\right)$$

If $\mu=0$, and $\sigma=1$, we call this standard normal distribution. The cumulative distribution function of the standard normal distribution is denoted as $\Phi(x)$

So, therefore, right how does it look like? So, therefore, the total power at a time t typically has got two important components; one is speed data which is the power consumed at time t due to the data dependent component and the other part is a noise component. So, I am assuming that you are doing the same operation ok. So, in that case right if there is no operation dependent component.

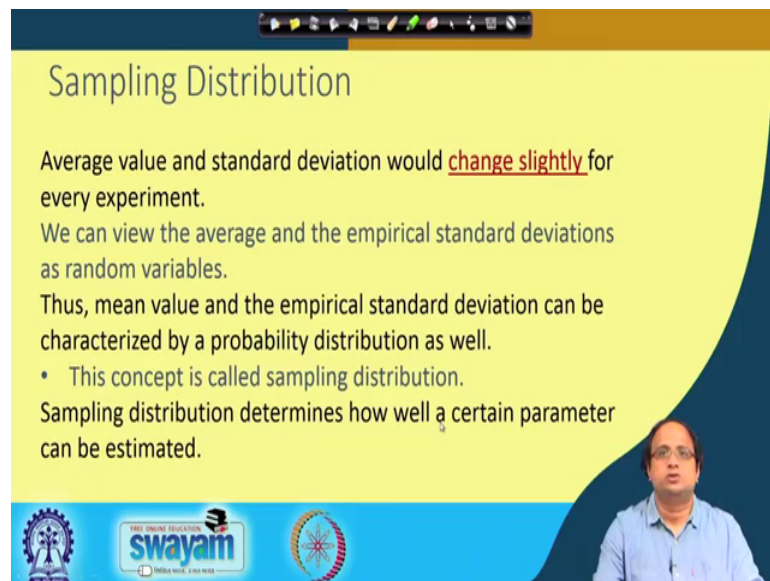
But rather there is a data dependent component because the data is changing although you are doing the same encryption again and again. So, and the other thing is the noise. So, this noise essentially you can see that say you know like essentially has got a Gaussian distribution and because of which right the power consumption also takes the Gaussian shape. So, for example, here we plot the point the power trace at a specific instant of time. So, essentially we see that you know like ah. So, in this case for example, this adjacent histogram shows the distribution of total power at a given instance of time for say 1000 encryptions with the same input and key ok.

And you can observe that the normal distribution is quite evident right you see that there is a nice normal distribution which where you can fit your histogram ok. So, we this is the histogram the green bars are the histograms and we have fit in a normal distribution to it. So, this normal distribution shows the distribution of electrical noise and just to recapitulate this is how a normal distribution looks like. This is a density function for the normal distribution. Now, the density function describes the describing the normal distribution depends upon the parameters μ and σ ; the mean and the standard deviation. Of course, the mean can be anything between minus infinity to plus infinity and a σ is essentially greater than 0 ok.

And you can write the usual way of saying that x follows a normal distribution is by using this notation and n and the parameters are μ and σ and if you say so, then it implies that your density function f_x right is essentially given by this equation ok. That is $\frac{1}{\sqrt{2\pi}\sigma}$ in the denominator exponentiation; e to the power of minus $\frac{1}{2} \left(\frac{x-\mu}{\sigma}\right)^2$ ok. If μ is 0 and σ is 1; then we call this as a standard normal distribution and the cumulative distribution function of the standard normal distribution is denoted as $\Phi(x)$; that means, like if I integrate for example, and I try to find out you know like the entire say suppose I draw a specific I draw a straight line here which is for a specific value of x and I try to find out basically add up all of them you know like till that value x right.

Then essentially I get the cumulative value and that distribution is often denoted as by the notation $\phi(x)$. So, the points in the power trace follows a normal distribution and the normal distribution is also called as the Gaussian distribution ok. So, we can also use these words interchangeably. We can say sometimes Normal distribution and we can sometimes say Gaussian distribution. So, now, what is the Sampling distribution?

(Refer Slide Time: 10:41)



Sampling Distribution

Average value and standard deviation would change slightly for every experiment.

We can view the average and the empirical standard deviations as random variables.

Thus, mean value and the empirical standard deviation can be characterized by a probability distribution as well.

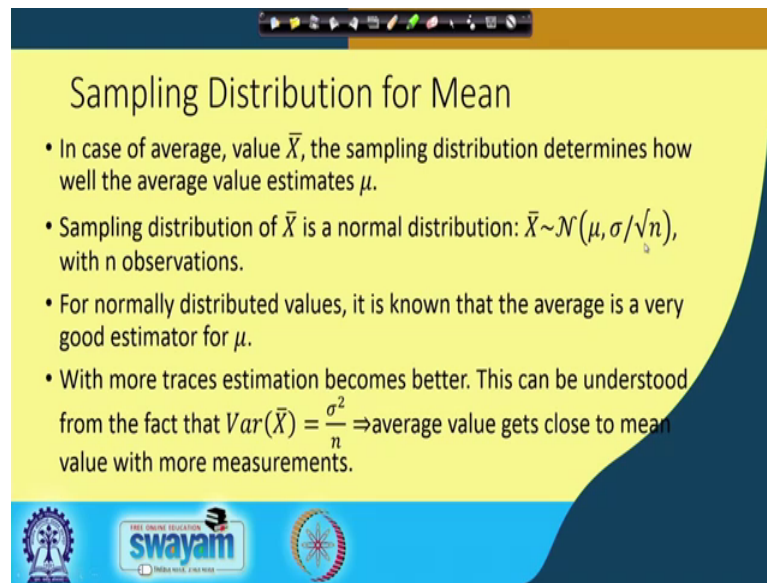
- This concept is called sampling distribution.

Sampling distribution determines how well a certain parameter can be estimated.

So, the average value and the standard deviation, suppose I am trying to estimate them ok; that means, I am basically say you know like getting 1 sample, 2 sample, 3 samples and I am trying to kind of estimate the corresponding average value and just and also the standard deviation. Now, this would change every time right, you are doing the experiment or redoing the experiment, these values will differ.

So, you can view them the average and the empirical standard deviations; therefore, as random variables. And the mean value and the empirical standard deviation can be characterized again as a probability distribution ok. So, that means and this is this concept is what is called as a sampling distribution. So, the sampling distribution determines often how well a certain parameter can be estimated. So, we try to develop sampling distributions for say mean.

(Refer Slide Time: 11:33)



The slide is titled "Sampling Distribution for Mean" and contains the following text:

- In case of average, value \bar{X} , the sampling distribution determines how well the average value estimates μ .
- Sampling distribution of \bar{X} is a normal distribution: $\bar{X} \sim \mathcal{N}(\mu, \sigma/\sqrt{n})$, with n observations.
- For normally distributed values, it is known that the average is a very good estimator for μ .
- With more traces estimation becomes better. This can be understood from the fact that $Var(\bar{X}) = \frac{\sigma^2}{n} \Rightarrow$ average value gets close to mean value with more measurements.

At the bottom of the slide, there are three logos: the Indian Institute of Space Science and Technology (IISST), the Swayam logo (Free Online Education), and the All India Institute of Medical Sciences (AIIMS).

And then, write the sampling distribution for mean would probably look it would look like this like suppose in a case of average the value \bar{X} which is a sampling distribution in this case it determines how well the average value estimates μ ok. So, we know that if I take large number of samples, then by the law of large numbers this will tend to μ . But of course, with finite values, it will not be exactly μ . But it will start approaching μ . So, the sampling distribution of \bar{X} is a normal distribution and it is denoted as \bar{X} follows normal distribution with the mean same as μ .

But the variance becomes σ^2/n . These also shows that when you are increasing the value of n ; then σ/\sqrt{n} is getting smaller and smaller; that means, you are basically approaching the average value is getting close to the mean value with more and more measurements; with more and more observations ok. So, therefore, right for normal distributed values, we know that the average is a very good estimator for μ and with more traces they are more trace estimations becomes better and the average value is tending more towards the object you know towards the value of μ ok.

(Refer Slide Time: 12:45)

The slide features a yellow background with a dark blue curved border on the right. At the top, there is a navigation bar with various icons. The title 'Confidence Intervals and Hypothesis Testing' is centered at the top. Below the title is a list of six bullet points. The first point defines confidence intervals. The second point explains a 0.99 confidence interval for μ . The third point states that hypothesis testing is a statistical tool to define the confidence interval. The fourth point discusses testing a hypothesis about a parameter μ . The fifth point lists two hypotheses: Null Hypothesis ($H_0: \mu = \mu_0$) and Alternative Hypothesis ($\mu \neq \mu_0$). The sixth point notes that these hypotheses are exclusive. In the bottom right corner, there is a small video inset showing a man in a blue shirt speaking. At the bottom of the slide, there are logos for 'swayam' and 'INDIA RISES WITH EDUCATION'.

Confidence Intervals and Hypothesis Testing

- The quantity how close a certain approximation is to a real parameter is defined by **confidence intervals**.
- When we say that we have a 0.99 confidence interval for μ , we mean that we have an interval that contains μ with a probability of 0.99.
- **Hypothesis testing** is a statistical tool to define the confidence interval.
- In such tests, we test whether a certain **hypothesis** that we make about the parameter(s) is true or not. For instance, we want to test whether $\mu = \mu_0$ or not.
- We define two hypothesis:
 - **Null Hypothesis**, $H_0: \mu = \mu_0$, and **Alternative Hypothesis**, $\mu \neq \mu_0$.
 - Note the null and alternative hypotheses are not necessary complementary, but they are exclusive.

So, now another important concept which is required to understand the you know like the following concepts is essentially what is called as the Confidence Interval and Hypothesis Testing ok. So, the quantity basically says how close a certain approximation is to a real parameter is often defined by confidence intervals. See this is a very important concept because when we are doing a difference of mean attack, we need to predict the mean right. So, therefore, this is the parameter which you are trying to estimate, but then the question would be how many observations you need to estimate it with the high confidence and that is where this concept of confidence interval comes in.

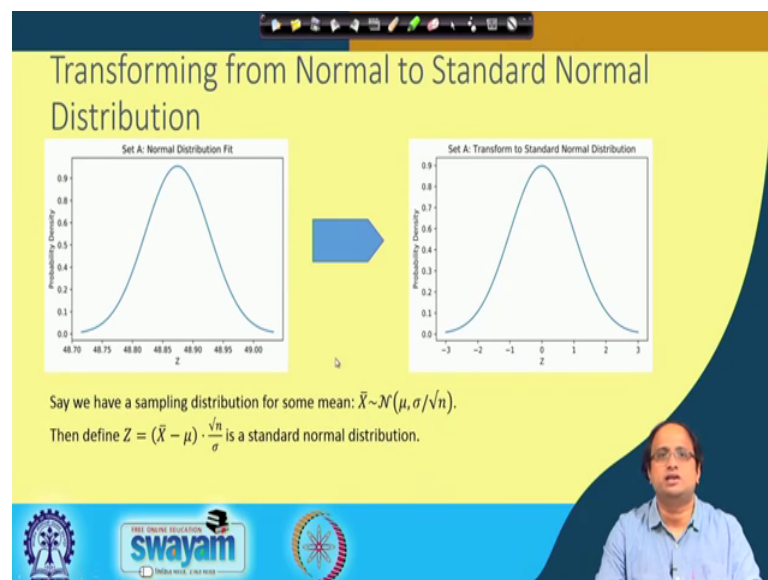
Because it tells us that it basically is a quantity or defines a quantity which tells how close a certain approximation is to a real parameter ok. So, when we say for example, that we have got a 0.99 confidence interval for μ , we mean to say that we have an interval that contains μ with the probability of 0.99 ok. That means, we have been we are saying that there is an interval which kind of engulfs μ with a high probability and hypothesis testing is a statistical tool to define this confidence interval ok. So, in such tests what we do is as follows. We basically test whether we basically make a hypothesis ok.

So, hypothesis would be essentially a claim about the parameter or parameters and for example, we want to you know like test whether μ is equal to μ_0 or whether μ is not equal to μ_0 . So, what we do is that we basically make 2 hypotheses; one is a null

hypothesis denoted as H_0 ; writing as described as $\mu = \mu_0$ and there is an Alternative Hypothesis which is $\mu \neq \mu_0$.

An important point which can be mentioned here is that the null and the alternative hypothesis although in this example are complementary, it need not be complementary. They are not necessarily complementary; that means they are not necessarily exhaustive. But they are exclusive; that means they are mutually independent. They are different; I mean they are rather than saying independent; it is better to say that they are exclusive.

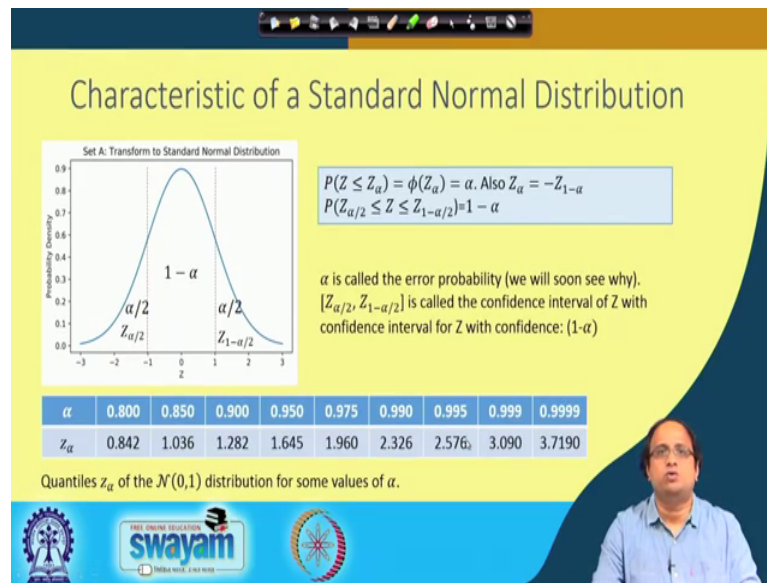
(Refer Slide Time: 14:57)



So, now so let us take our example. So, this is our power trace for example. So, we basically had we have basically feed that if you did that into a normal distribution and we would like to convert this into a standard normal distribution and I say that in a standard normal distribution, you have got the mean as 0 and the standard deviation right or sigma essentially as 1. Therefore, the normal distribution is 0 comma 1.

So, now the usual way of transforming that is that suppose you have got the sampling distribution for mean denoted as μ comma σ by root n as we have seen in the previous slide, we will make a transformation from \bar{X} . So, we basically substitute \bar{X} as $\bar{X} - \mu$ into square root of n by σ . If you do this transformation, then you get Z and this Z essentially will follow a standard normal distribution which means it will follow n with mean as 0 and standard deviation or σ as 1.

(Refer Slide Time: 15:59)



So, now we would like to develop or find out the characteristic of a standard normal distribution which we will be using in our hypothesis testing. So, this is the standard normal distribution curve. Now this particular curve has got some important properties ok. These properties are typically right essentially very important and central to the understanding of hypothesis testing and they are essentially described by something which is called as Z alpha or what are called as Quantiles ok.

So, let us see what they mean the idea is that if you take the area under the curve, the area under the curve would be essentially 1 or unity and here, you will see that there is there are certain notations which are called as Z alpha and as I said the cumulative value which is essentially phi of Z alpha indicates the probability that the Z statistics takes a value which is less than or equal to Z alpha.

So, $P(Z \leq Z_{\alpha})$ is nothing but phi of Z alpha and that is equal to alpha in this in this particular curve and also note that Z alpha is equal to minus of Z 1 minus alpha because it is kind of symmetric over the 0 well ok. So, this is the 0 line because the mean is in this case 0. So, we have got Z alpha equal to minus of Z 1 minus alpha ok.

So, this you can easily check from the normal distribution curves ok. And so now, that that implies that if I take a region from Z alpha by 2 to Z 1 minus alpha by 2; then the probability that Z will lie from Z alpha by 2 to Z 1 minus alpha by 2 is equal to 1 minus alpha, that is because if you take Z 1 minus alpha by 2; then a probability that it is less

than $Z_{1 - \alpha/2}$ is; so for example, right you would take $\alpha/2$, the probability that it is less than $Z_{\alpha/2}$ is $\alpha/2$ ok.

And as we know that from here $Z_{\alpha/2}$ is equal to minus of $Z_{1 - \alpha/2}$ and by the symmetry we know that the probability that it is greater than $Z_{1 - \alpha/2}$ is also $\alpha/2$ and as I said that since the area under the curve is 1; that means, the probability that Z will lie in this region that is from $Z_{\alpha/2}$ to $Z_{1 - \alpha/2}$ is nothing but $1 - \alpha/2 + \alpha/2$ that is $1 - \alpha$.

So, therefore right what we will see why there is a reason why we call α as the error probability, I will explain very soon why. But you can, but rather let me state here that this region or this closed region from $Z_{\alpha/2}$ to $Z_{1 - \alpha/2}$ is called the confidence interval of Z with a confidence of $1 - \alpha$ ok. So, now, this is this you know like the statistics or the properties of this standard normal distribution is often described by a table ok. So, here is a small summary of the table, but there are more data that you can easily get from a bigger table.

For example, like if I take α and if the α value is 0.05 for example, then the corresponding $Z_{\alpha/2}$ is 1.960. So, I hope now we understand that what it means because this follows from this fact that is $P(Z \leq Z_{\alpha/2}) = \alpha/2$; that means, if I take a value of α like 0.05; that is 0.05 somewhere here, then the you know like that.

So, this if this is your. So, if your $Z_{\alpha/2}$ is for example, 1.960, then I would basically draw a line here through this point that is 1.960 and that would imply that the probability that Z is lesser than this; that is it lies on the left hand side is given by this value which is denoted by $\alpha/2$ ok. So, for example, the α value is 0.05 ok. So, therefore, the quantiles essentially defines this probability distribution and which we will be using in our hypothesis testing shortly.

(Refer Slide Time: 20:15)

Derive the Confidence Interval for μ

- Using, $Z = (\bar{X} - \mu) \cdot \frac{\sqrt{n}}{\sigma}$ we have:
- $P(Z_{\alpha/2} \leq (\bar{X} - \mu) \cdot \frac{\sqrt{n}}{\sigma} \leq Z_{1-\alpha/2})$
 $= P\left(\frac{\sigma}{\sqrt{n}} Z_{\alpha/2} \leq (\bar{X} - \mu) \leq \frac{\sigma}{\sqrt{n}} Z_{1-\alpha/2}\right)$
 $= P\left(\bar{X} - \frac{\sigma}{\sqrt{n}} Z_{1-\alpha/2} \leq \mu \leq \bar{X} - \frac{\sigma}{\sqrt{n}} Z_{\alpha/2}\right).$
- Using, $Z_{\alpha/2} = -Z_{1-\alpha/2}$, thus we continue:
 $= P\left(\bar{X} - \frac{\sigma}{\sqrt{n}} Z_{1-\alpha/2} \leq \mu \leq \bar{X} + \frac{\sigma}{\sqrt{n}} Z_{1-\alpha/2}\right).$

$\left[\bar{X} - \frac{\sigma}{\sqrt{n}} Z_{1-\alpha/2}, \bar{X} + \frac{\sigma}{\sqrt{n}} Z_{1-\alpha/2}\right]$ is a $(1 - \alpha)$ confidence interval for μ .

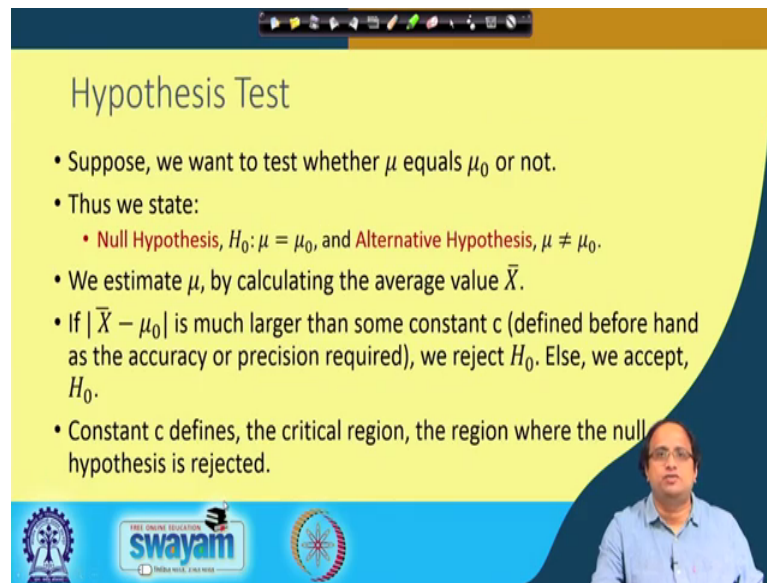
swayam

So, let us try to define now the confidence interval for mu which we are trying to estimate. So, as I said that Z essentially follows a standard normal distribution curve, where Z is equal to x bar minus mu into root n by sigma. So, that would imply that we are basically trying to limit Z between Z alpha by 2 and Z 1 minus alpha by 2 right that is the confidence interval that we just now saw here ok. So, now, using this right we basically write this equation and that implies that I can now you know like do a few simplification I can multiply both these sides by sigma by root n.

Therefore, I have got sigma by root n Z alpha by 2 which is less than equal to x bar minus mu is less than equal to sigma by root n into Z 1 minus alpha by 2 and again right. Since I want the confidence interval for mu, I bring in I do a little bit of maneuvering and therefore, I see that mu is bounded between X bar minus sigma by root n Z alpha by 2 and it is also lower bounded by X bar minus sigma by root n Z 1 minus alpha by 2 ok. Now, note that Z alpha by 2 is nothing but minus Z 1 minus alpha by 2.

So, therefore, I can replace this error for by 2 by minus of Z 1 minus alpha by 2 and therefore, I get that mu is bounded between X bar minus X bar plus minus sigma by square root of n Z 1 minus alpha by 2. So, therefore, my confidence interval for mu with the confidence of 1 minus alpha would be X bar minus sigma by square root of n Z 1 minus alpha by 2 to X bar plus sigma by square root of n Z 1 minus alpha by 2.

(Refer Slide Time: 21:59)



Hypothesis Test

- Suppose, we want to test whether μ equals μ_0 or not.
- Thus we state:
 - Null Hypothesis, $H_0: \mu = \mu_0$, and Alternative Hypothesis, $\mu \neq \mu_0$.
- We estimate μ , by calculating the average value \bar{X} .
- If $|\bar{X} - \mu_0|$ is much larger than some constant c (defined before hand as the accuracy or precision required), we reject H_0 . Else, we accept, H_0 .
- Constant c defines, the critical region, the region where the null hypothesis is rejected.

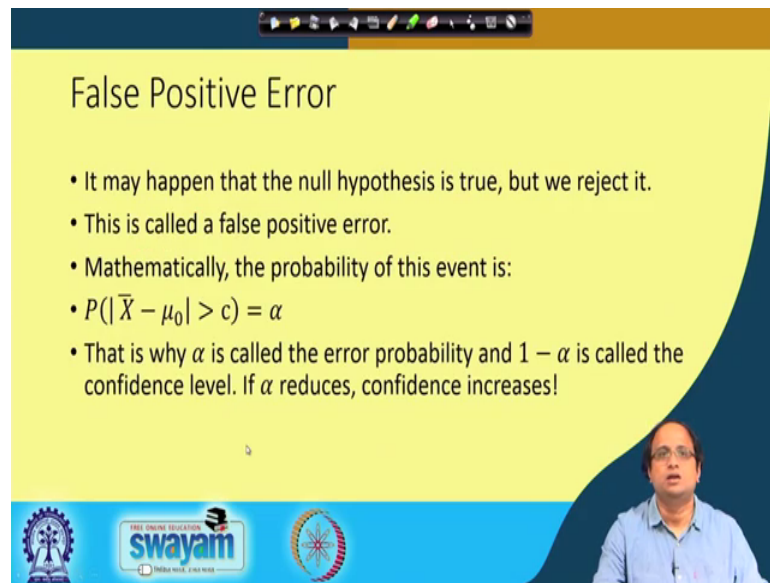
Logos at the bottom: Swamyam (Free Online Education), Anna University, and other educational institutions.

So, now once we have derived this confidence interval, we are all set to understand what is hypothesis testing. So, as I said that if you want to test that whether μ equals to μ_0 or not or whether μ is equal to μ_0 or μ is not equal to μ_0 , you make 2 hypotheses. So, you state two hypotheses the null hypothesis is that μ is equal to μ_0 and the alternative hypothesis or H_1 is μ is not equal to μ_0 . So, now, we estimate μ by calculating the average value \bar{X} because we do not know what is the actual of μ .

So, now, the idea is that if $\bar{X} - \mu_0$ is much larger than some predefined constant; μ_0 is the value that you are trying to see whether μ is equal to that and if you find out that this empirical mean is not equal to that; that means, the absolute value of this error is much larger than some predefined constant. So, this predefined constant is often given beforehand depending upon the accuracy or precision that is required, then we will reject H_0 . Because that means, that it is too big; then allowed ok, then possibly and that possibly means that \bar{X} is not μ_0 ok.

Or I mean rather μ is not equal to μ_0 . Else that means, if it lies inside this region, then we accept H_0 . We accept the null hypothesis. So, this constant is vital again and it defines what is called as a critical region, the region where the null hypothesis is rejected because if it lies in this region it is much larger than this constant c then and it lies in that region, then you are basically rejecting the null hypothesis ok.

(Refer Slide Time: 23:31)



False Positive Error

- It may happen that the null hypothesis is true, but we reject it.
- This is called a false positive error.
- Mathematically, the probability of this event is:
- $P(|\bar{X} - \mu_0| > c) = \alpha$
- That is why α is called the error probability and $1 - \alpha$ is called the confidence level. If α reduces, confidence increases!

swamyam
FREE ONLINE EDUCATION
INDIA WISE, LEAD WISE

So, therefore, right this brings us to the fact why alpha is called as an Error. You remember why it called alpha as an error. So, it is because of this fact that it basically is nothing but the false positive error. So, it may happen that the null hypothesis is true, but we reject it and this is called as a False Positive Error. Mathematically the probability of this event is nothing but $P(|\bar{X} - \mu_0| > c) = \alpha$ and that is equal to alpha.

So, that is why alpha is called as an error probability and what we want often is that we will try to reduce this error probability and if we do that, then that implies that $1 - \alpha$ will get increased and therefore, the confidence will get increased and that is intuitive; that is why you are reducing error and you are increasing your confidence of estimate and that is understood. So, therefore, if alpha reduces then confidence increases.

(Refer Slide Time: 24:23)

Estimating the Number of Traces to Estimate μ

- Suppose, we need to estimate the mean with precision c (say 0.01).
- $P(|\bar{X} - \mu_0| > c) = \alpha$,
- Or, $P(|\bar{X} - \mu_0| > c) = P\left(|\bar{X} - \mu_0| \frac{\sqrt{n}}{\sigma} > c \frac{\sqrt{n}}{\sigma}\right) = P\left(|Z| > c \frac{\sqrt{n}}{\sigma}\right) = 2P\left(Z > c \frac{\sqrt{n}}{\sigma}\right) = \alpha$
- Or, $P\left(Z > c \frac{\sqrt{n}}{\sigma}\right) = \frac{\alpha}{2} \Rightarrow c \frac{\sqrt{n}}{\sigma} = Z_{1-\alpha/2}$
- This implies, $n = \frac{\sigma^2}{c^2} Z_{1-\alpha/2}^2$.

The slide also features a normal distribution curve with the area under the curve to the right of a point labeled $z_{1-\alpha/2}$ shaded in red. The total shaded area is labeled $\alpha/2$. The presenter is a man with glasses wearing a light blue shirt.

So, now we are all set to understand or estimate the number of traces to estimate μ ok. Suppose, we need to estimate the mean with a precision say 0.01 ok. So, say P of; so, I say that P of $|\bar{X} - \mu_0| > c$, this is equal to α ; this is our error probability. Ordinarily that means, that probability that $\bar{X} - \mu_0$ is greater than c you can also write that by multiplying both sides by \sqrt{n} by \sqrt{n} by σ . Note that σ is positive.

So, you can multiply them without changing the sign and this is nothing but your Z right or the absolute value of Z which is your Z statistic. Now, the absolute value of Z is basically greater than $c \sqrt{n}$ by σ and this means that this probability right if we just remove this absolute sign, then that means, that this is nothing but 2 into probability of Z greater than equal to $c \sqrt{n}$ by σ and that is equal to α . So, this is an important thing to understand. So, so I hope that it is clear. So, let me try here to explain this point.

So, basically what I am trying to say is that this is your normal curve right and you are basically trying to say that when you are saying that $|Z| > c \sqrt{n}$ by σ , then that means, you are bothered about the absolute value of Z being greater than this ok. So that means, you are basically bothered about both these regions; this region as well as this region. Because the absolute value is important because when, but when you

are writing this fact that is Z is greater than this we are only writing this region. So, in order to account for that, you basically bring in this factor of 2 ok.

So, you bring in this factor of two and therefore, write $P(Z > c \sqrt{n} / \sigma)$ is equal to $\alpha / 2$ and that implies that this is nothing but $Z_{1 - \alpha / 2}$ ok. Remember right that this essentially as I say that if this is $Z_{1 - \alpha / 2}$, then the probability that it is greater than this was shown to be $\alpha / 2$ ok. So, therefore, right we can essentially write that this parameter that is $c \sqrt{n} / \sigma$ is nothing and equated nothing but $Z_{1 - \alpha / 2}$ and this immediately tells me that n is equal to $\sigma^2 / c^2 \times Z_{1 - \alpha / 2}^2$ whole square ok. So, gives me a closed form expression of estimating the number of traces which are required to estimate the value of μ .

So, what we will see next is we will see few more you know like inferences on this and try to see how we can apply this on you know like on getting the difference of mean attack analyzed in a much more predictable fashion, but we will take that in the next class.