# Lecture 31: Estimation - Part IV

Satyajit Thakor
IIT Mandi

# Recall - Interval estimation

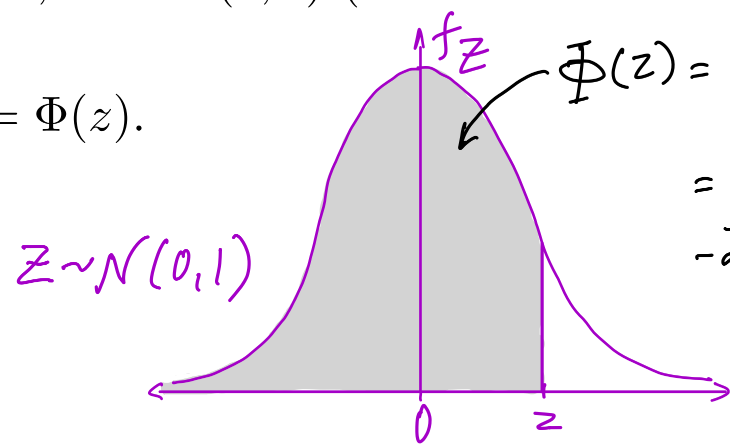
- An interval estimate of a population parameter $\theta$ is an interval of the form $\hat{\theta}_L < \theta < \hat{\theta}_U$, where $\hat{\theta}_L$ and $\hat{\theta}_U$ depend on the value (e.g., $\hat{\theta}$) of the statistic $\hat{\Theta}$ for a particular sample and also on the distribution of the parameter $\hat{\Theta}$.
- If, for instance, we find $\hat{\theta}_L$ and $\hat{\theta}_U$ such that

$$P(\underbrace{\hat{\theta}_L}_{\text{lower}} < \theta < \underbrace{\hat{\theta}_U}_{\text{upper}}) = 1 - \alpha,$$

  for $0 < \alpha < 1$, then we have a probability of $1 - \alpha$ of selecting a random sample that will produce an interval containing $\theta$.
- In this case, the interval $\hat{\theta}_L < \theta < \hat{\theta}_U$ is called $100(1 - \alpha)$ percent confidence interval estimate of $\theta$.
- Next we will try to find $100(1 - \alpha)$ percent confidence interval estimate of a sample mean $\bar{X}$.

# Recall - Interval estimation

- For an example of $100(1-\alpha)$ percent confidence interval estimate of a sample mean $\bar{X}$, assume that $\bar{X}$ is normally distributed.
- Note that this is a very logical assumption for large $n$ since the CLT suggests that the distribution of $\bar{X}$ can be well approximated by the normal distribution $\mathcal{N}(\mu, \sigma^2/n)$ (recall: Slides 4-5 of Lecture 26).
- Also, recall that if $\bar{X}$ has the distribution $\mathcal{N}(\mu, \sigma^2/n)$ then $\bar{X} = \mu + (\sigma/\sqrt{n})Z$, where, $Z \sim \mathcal{N}(0, 1)$ (recall: Slide 7 of Lecture 25).
- Now, denote $P(Z \leq z) = \Phi(z)$.

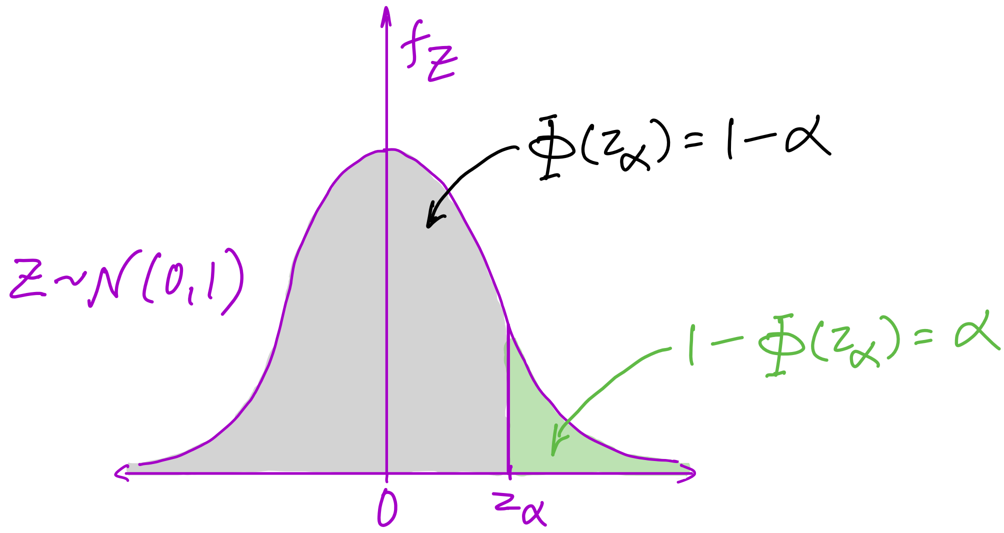



$\Phi(z) = P(Z \leq z)$

$= \int_{-\infty}^{z} f_Z(z)\, dz$

# Recall - Interval estimation

- Let $z_\alpha$ be the value such that $\Phi(z_\alpha) = 1 - \alpha$.
- For any give $0 < \alpha < 1$, we can find out $z_\alpha$ numerically (using the CDF table for standard normal).



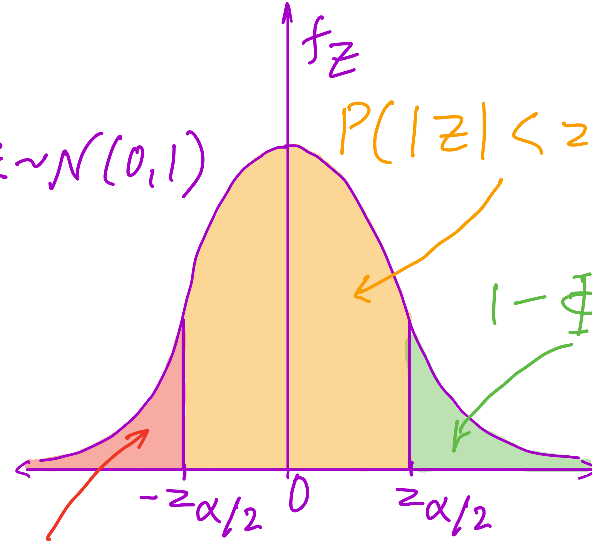
# Interval estimation

- $P(|Z| < z_{\alpha/2}) = 1 - \alpha$
- Proof:

$$P(|Z| < z_{\alpha/2})$$
$$= P(-z_{\alpha/2} < Z < z_{\alpha/2})$$
$$= \Phi(z_{\alpha/2}) - \Phi(-z_{\alpha/2})$$
$$= 1 - \frac{\alpha}{2} - \frac{\alpha}{2}$$
$$= 1 - \alpha.$$

$P(|Z| < z_{\alpha/2}) = 1 - \alpha$

$1 - \Phi(z_{\alpha/2}) = \alpha/2$




$$\underbrace{\Phi(-z_{\alpha/2}) = \alpha/2}$$

$\because f_Z$ is symmetric around $0$.

$$\Rightarrow \Phi(-z_{\alpha/2}) = 1 - \Phi(z_{\alpha/2}) = \alpha/2.$$

# Interval estimation

- For our sample mean example, we have $\bar{X} = \mu + (\sigma/\sqrt{n})Z$ where $Z$ is a standard normal.

Then, $P(|Z| < z_{\alpha/2}) = 1-\alpha$

$$\Rightarrow P(|Z| < z_{\alpha/2})$$

$$= P(-z_{\alpha/2} < Z < z_{\alpha/2})$$

$$= P\left(-z_{\alpha/2} < \frac{\bar{X}-\mu}{\sigma/\sqrt{n}} < z_{\alpha/2}\right)$$

$$= P\left(-z_{\alpha/2}\frac{\sigma}{\sqrt{n}} < \bar{X}-\mu < z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right)$$

$$= P\left(\bar{X}-z_{\alpha/2}\frac{\sigma}{\sqrt{n}} < \mu < \bar{X}+z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right) = 1-\alpha$$

# Interval estimation

- Based on the derivation we did in the last few slides, we define a $100(1-\alpha)$ percent confidence interval for $\mu$ as follows:

- If $\bar{x}$ is the mean of a random sample of size $n$ from a population with known variance $\sigma^2$, a $100(1-\alpha)$ percent confidence interval for $\mu$ is given by

$$\underbrace{\bar{x} - z_{\alpha/2}\frac{\sigma}{\sqrt{n}}}_{\hat{\theta}_L} < \mu < \underbrace{\bar{x} + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}}_{\hat{\theta}_U}$$

where $z_{\alpha/2}$ is the value such that $\Phi(z_{\alpha/2}) = 1 - \alpha/2$.

Two-sided interval estimation.

- There are also one-sided estimates.

e.g., a parameter is "at least" or "at most" certain value with certain confidence. $[\theta_L, \infty)$ $(-\infty, \theta_U]$

# Interval estimation

- Example: The average zinc concentration recovered from a sample of measurements taken in 36 different locations in a river is found to be 2.6 grams per milliliter. Find the 95% confidence intervals for the mean zinc concentration in the river. Assume that the population standard deviation is 0.3 gram per milliliter.

- Let $X$ be the r.v. : "zinc concentration."

- Given: sample mean $\bar{x} = 2.6$ grams/ml & $\sigma_X = 0.3$ grams/ml.

- $\bar{x}$ is an estimate of $\mu_X$.

- For 95% confidence interval $\underbrace{\alpha = 0.05.}_{\because\, 95 = 100(1-\alpha)}$

# Interval estimation

- Then, recall that, $P(Z < z_{\alpha/2}) = 1 - \alpha/2$

$$\Rightarrow \Phi(z_{0.025}) = 1 - 0.025$$
$$= 0.975$$

- By the CDF table, $z_{0.025} = 1.96$.

- The $100(1-\alpha)\%$ confidence interval of $\mu$ is

$$\bar{x} - z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}} < \mu < \bar{x} + z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}$$

- Hence, 95% confidence interval of $\mu$ is

$$2.6 - (1.96) \cdot \frac{0.3}{\sqrt{36}} < \mu_X < 2.6 + (1.96) \cdot \frac{0.3}{\sqrt{36}}$$

$$\Rightarrow 2.50 < \mu_X < 2.69$$

# Hypothesis testing

- That is, it is quite likely (type II error prob. is 0.2517) that we shall reject the new vaccine when, in fact, it is superior (50% effective) to what is now in use (compared to 25% effective).
- Now, let the particular alternative hypothesis be $p = .7 > 1/4$. Then,

- That is, t is extremely unlikely that the new vaccine would be rejected when it was 70% effective after a period of 2 years.

# Hypothesis testing

- How to decrease the probability of type I and II errors?
- The probability of committing both types of error can be reduced by increasing the sample size.
- a