Lecture 29

Today we begin our discussion of interval estimation. Rather than just reporting a single number it seems more sensible, due to sampling variability, that reporting an interval of plausible values makes more sense. We typically call these interval estimates:

Confidence Interval

- Report as interval ie (lower limit, upper limit)
- Has something we call a "confidence level" which is a measure of the reliability of the interval. Typically we like this to be high. Commonly used values are 90%, 95% and 99%.

Confidence interval for μ with σ known and normal data

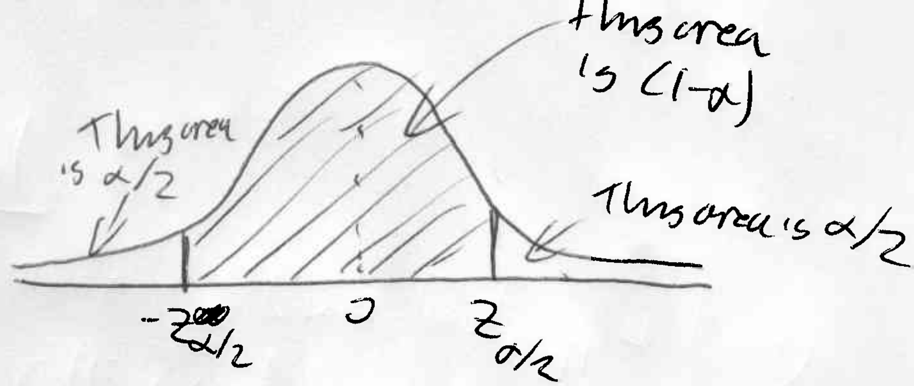
Assume that $X_1, \ldots, X_n$ represent a random sample from a normal distribution $N(\mu, \sigma)$. Recall from previous work that this implies that $\bar{X} \sim N(\mu, \frac{\sigma}{\sqrt{n}})$ and therefore

$$Z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}}$$ has $N(0,1)$ distribution

Let $z_{\alpha/2}$ be a constant such that

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

ie



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

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

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

In otherwords with probability $(1-\alpha)$ the population parameter μ lies between

$$\left(\bar{X} - z_{\alpha/2}\frac{\sigma}{\sqrt{n}} \,,\; \bar{X} + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right)$$

equivalently. We can say that a $100(1-\alpha)\%$ confidence interval for μ is given by

$$\boxed{\bar{x} \pm z_{\alpha/2}\frac{\sigma}{\sqrt{n}}}$$

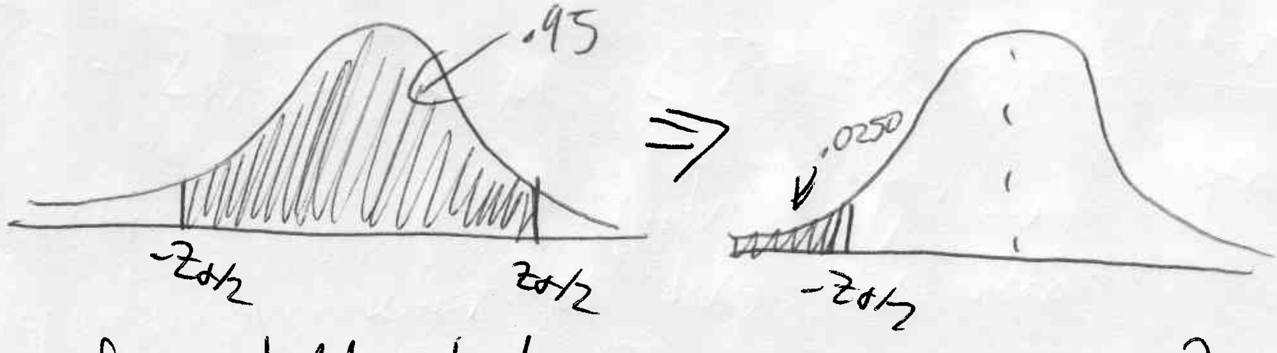
note sample mean evaluated at data

what is $z_{\alpha/2}$?

Confidence level	α	$z_{\alpha/2}$
90%	.10	1.645
95%	.05	1.96
99%	.01	2.576

These are from normal dist table.

eg for 95%

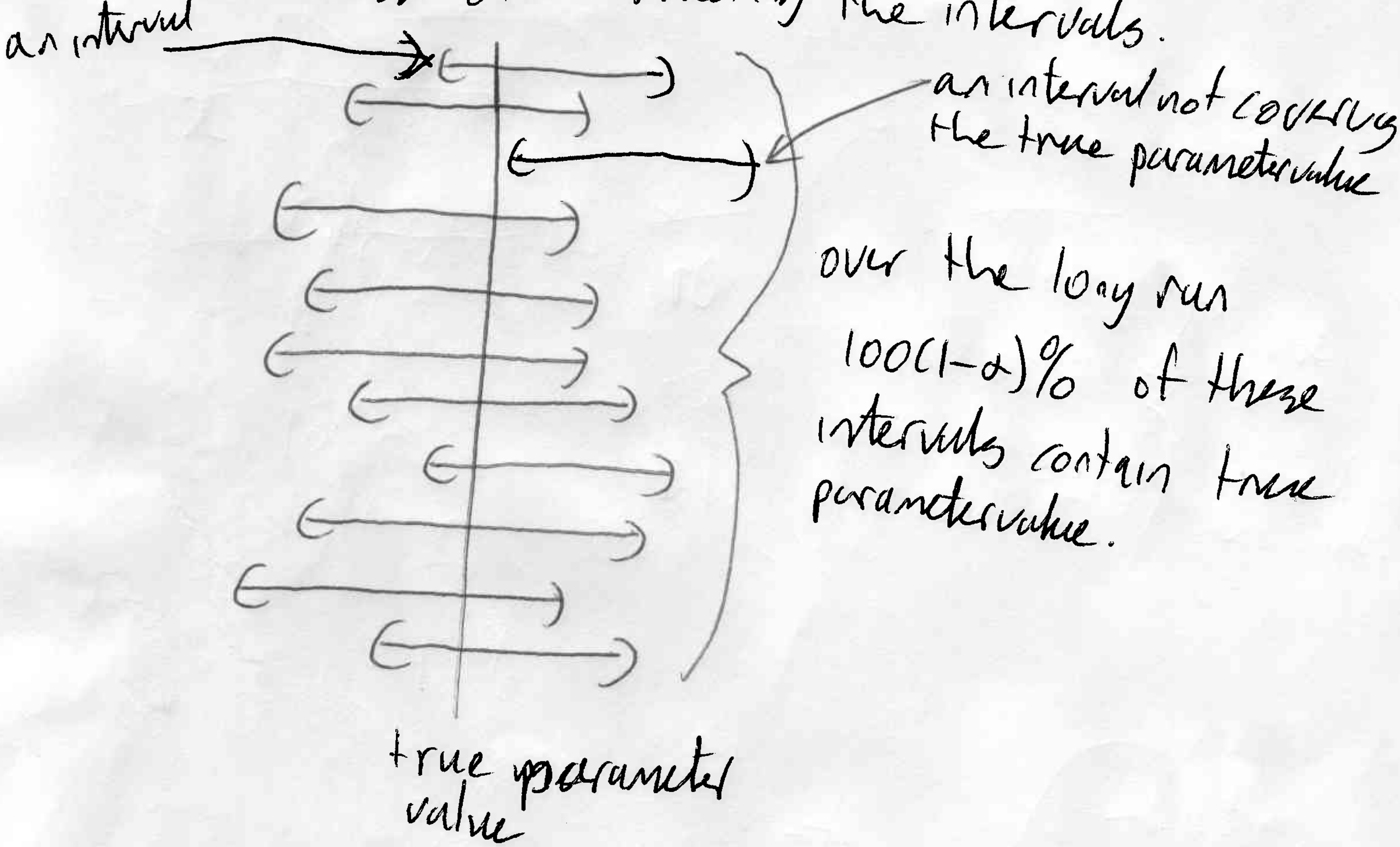


from table what z gives $\Phi(z) = .0250$? -1.96!!

Interpreting a Confidence interval

1. Note we can not say that μ really lies in our particular interval even with some probability.

2. Instead the probability statement refers to the method of constructing the intervals.



3. A higher confidence level will give you a wider interval.

4. A larger sample size will give you a narrower interval.

5

5. we don't take 100% CI because they are non informative ie $(-\infty, \infty)$

6. Confidence interval width can be viewed as precision. Confidence level as reliability. high precision $\Leftrightarrow$ low reliability, low precision $\Leftrightarrow$ high reliability } trade off.

Choosing sample size

we can relate the sample size to the confidence interval width

$$w = 2 z_{\alpha/2} \frac{\sigma}{\sqrt{n}}$$

$$\Rightarrow n = \left(2 z_{\alpha/2} \frac{\sigma}{w}\right)^2$$

ie if we wish to have a desired confidence interval width we need on ensure that we sample a certain size.