Handout 19: Confidence Intervals

Over the past few classes we have discussed properties that make for good estimators $\widehat{\theta}$ of some unknown parameter θ. These are often called *point estimators* as they specify the single best guess (i.e., a point) for θ. Today we will discuss providing a range of guesses for a parameter value; this is called a *confidence interval*. Confidence intervals make probabilistic guarantees about how likely it is that θ is contained in the predicted interval. That is:

$$\mathbb{P}\left[\widehat{\theta}_{lower} \leq \theta \leq \widehat{\theta}_{upper}\right] \geq 1 - \alpha$$

For some reasonable α near 0.

Suppose that $X_1, X_2, \ldots, X_n \overset{iid}{\sim} F$ with a finite mean μ and finite variance σ^2. Consider the following statistic:

$$t = \frac{\bar{X} - \mu}{s/\sqrt{n}}$$

Where, as before,

$$s = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - (\bar{X}))^2.$$

In this case we know by the central limit theorem that t is asymptotically distributed as $N(0,1)$.[1] Define z_α as the α cutoff for the standard normal distribution,

[1] Because the distribution does not depend on the parameter of interest, μ, we say that it t a pivot statistic.

$$\alpha = \int_{z_\alpha}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}\, dx,$$

We then have:

$$\mathbb{P}\left(-z_{\alpha/2} \leq \frac{\bar{X} - \mu}{s/\sqrt{n}} \leq z_{\alpha/2}\right) \approx 1 - \alpha.$$

Which can be re-written as:

$$\mathbb{P}\left(\bar{X} - (s/\sqrt{n})z_{\alpha/2} \leq \mu \leq \bar{X} + (s/\sqrt{n})z_{\alpha/2}\right) \approx 1 - \alpha$$

Or, even as:

$$\bar{X} \pm \frac{s}{\sqrt{n}} \cdot z_{\alpha/2}$$

This particular case is important because we can use it without knowing the exact distribution family that X is drawn from. Therefore, we can use this formula to approximately capture the mean from any sample taken independently from a distribution with finite variance.