Handout 04: Asymptotic T Statistic

Last time, we have derived a confidence interval using the T-distribution for estimation of the mean of a population distribution that we know to be normal (but do not know the mean or variance). Today, we will apply this technique to some actual data in R. But, in general, real data is never perfectly normal. So, why is the approach still valid? Let's see!

First, let's define k_X to be a constant given by a scaled version of the fourth-central moment of the population distribution:

$$k_X = \text{Var}\left[\frac{X_i - \mu_X}{\sigma_X}\right]^2.$$

Will will assume that k_X is finite. Then, in the limit of large n, the following chain of relationships:[1]

$$\frac{S_X^2}{\sigma_X^2} = \frac{1}{n-1}\sum_i \left[\frac{X_i - \bar{X}}{\sigma_X}\right]^2 \underset{\mathcal{P}}{\rightarrow} \frac{1}{n}\sum_i \left[\frac{X_i - \mu_X}{\sigma_X}\right]^2 \underset{\mathcal{D}}{\rightarrow} N(1, \frac{k_X}{n}) \underset{\mathcal{P}}{\rightarrow} 1.$$

[1] We write $\underset{\mathcal{P}}{\rightarrow}$ to indicate convergence in probability, and $\underset{\mathcal{D}}{\rightarrow}$ to indicate an asymptotic distribution.

We will not go into the spectic proof details here, partially because part of the proof requires some functional analysis. At the same time, each of the steps should seem fairly intuitive. And the second limit is just a straightforward application of the central limit theorem.

Now, plugging this into the formula for the T statistic, we have:

$$T = \frac{\frac{\mu_X - \bar{X}}{\sqrt{\sigma_X^2/n}}}{\sqrt{\frac{(n-1)S_X^2}{\sigma_X^2 \cdot (n-1)}}} = \frac{\frac{\mu_X - \bar{X}}{\sqrt{\sigma_X^2/n}}}{\sqrt{\frac{S_X^2}{\sigma_X^2}}} \underset{\mathcal{D}}{\rightarrow} \frac{N(0,1)}{1} = N(0,1).$$

The numerator is just a straightforward application of the central limit theorem to $\bar{X}$; the denominator comes from the chain of relationships above. We can combine them in the "natural" way that you might assume due to a result called **Slutsky's Theorem**.