Handout 05: Resampling Techniques

Let $\hat{\theta}$ be an unbiased point estimator for an unknown parameter θ. We have seen that if we can construct a pivot with a known distribution based on this estimator and parameter, we can manipulate these to derive confidence intervals and run hypothesis tests. Today, we will see an important computational technique to apply this technique when we do not know the full distribution of $\hat{\theta}$. A common example, for instance, would be in the case where we have a sample mean but are not confident in assuming that the central limit theorem can be accurately applied to our data.

In a world where we could run an experiment a large number of times, we could reliably estimate the distribution of $\hat{\theta}$ by looking at the distribution of outcomes from each experiment. Usually, this is not feasible (and when it is, there is no need to do it anyway). An approximate way to simulation this, though, is possible. We can create a **bootstrap sample** of n observations taken from sampling with replacement from the original data. If we do this many times, we will get a full approximation of the distribution of the point estimator. We will not get into the theoretical proofs of the bootstrap, but know that they have good convergence properties under weak regularity conditions.

There are a few different ways to produce confidence intervals and hypothesis tests from bootstrap samples. We will find the **percentile bootstrap** to be the best for our current applications. To create a percentile bootstrap let $\hat{\theta}_{\alpha/2}$ and $\hat{\theta}_{1-\alpha/2}$ be the corresponding estimated percentiles of the point estimator. Then, the confidence interval is constructed by using these as the lower and upper estimates. A simple approach to hypothesis testing can be done by testing whether the null hypothesis value falls within the confidence interval.

The bootstrap is just one example of a broad set of approaches called resampling methods. Another example are **permutation tests**, which are specifically designed to facilitate hypothesis testing. While there is one generic approach to the bootstrap that can be applied to any estimator, permutation tests need a bit more customization. Let's consider a permutation test alternative to the two-sample T-test, which can be explained fairly clear in prose. First, compute the standard T-statistic. Then, randomly reassign each of the observations to each of the two groups and recalculate the test statistic. Repeat a large number of times. A p-value is computed by measuring how often the statistic from the randomly assigned labels is larger in absolute value than the observed set.