**Bootstrapping**

**Hypothesis Testing**

To do hypothesis tests, we generally need to know how to relate the presumed statistic of our population (assuming the Null Hypothesis), η0, to the probability distribution of our measurement, Pη0(m). And then we’d say, for instance, depending on the test,

  

if we’re looking for evidence that our sample measurement invalidates the Null Hypothesis. There is another approach that we can use. We could effectively reverse our reasoning, and take our measurement result as the Null Hypothesis, and determine the p-value of η0 being as far away as it is. So something like,

  

But this begs the question of how we’d get the sample measurement probability distribution from just a single sample measurement. This is what bootstrapping is all about. Apparently, what we do is create so-called bootstrap samples, by sampling our sample, *with replacement*, a large number of times. These bootstrap samples should be the same size as the original. So for instance if our sample were {x1, x2, x3, x4}. Then our bootstrap samples might be: {x1, x2, x2, x4}, {x3, x2, x1, x4}, {x3, x1, x4, x4}, etc. And then we construct measurements, m1, m2, m3, etc., from these bootstrap samples. And then we construct a histogram from these measurements. This histogram should closely match the true probability distribution Pη=m\*(m). And from the histogram, we can evaluate our p-value.

**Confidence Intervals**

To do confidence intervals, we generally need to know how to relate the presumed statistic of our population, η, to the probability distribution of our measurement, Pη(m). Then we can say, for the min and max η, that:

 

This simplified for the case of a symmetric, translationally invariant distribution Pη(m). We could say the bounds were given by:



and, even more simply, by:



These last two formulas invoke only Pη=m\*(m), rather than Pη=η\_max(m), or Pη=η\_min(m). So we don’t need to have any theoretical knowledge about how the population parameter η affects the measurement probability distribution, just knowledge about the measurement probability distribution for the case when η = m\*. This is fortuitous because the sample points we collected are those for η = m\*. So all we need is a way to construct the measurement probability distribution from our sample points. This is prima facie impossible though, because we have only *one* set of sample points, say {x1, x2, x3, …, xn}, from which we constructed *one* measurement, m = m\*. How can we get more measurements from our single sample? This is what bootstrapping is all about. Apparently, what we do is create so-called bootstrap samples, by sampling our sample, *with replacement*, a large number of times. These bootstrap samples should be the same size as the original, i.e., n items. So for instance if our sample were {x1, x2, x3, x4}. Then our bootstrap samples might be: {x1, x2, x2, x4}, {x3, x2, x1, x4}, {x3, x1, x4, x4}, etc. And then we construct measurements, m1, m2, m3, etc., from these bootstrap samples. And then we construct a histogram from these measurements. This histogram should fairly closely match the true probability distribution Pη=m\*(m). And from the histogram, we can get zα/2 and σm.