

Estimation, Part 2: The Bootstrap

Confidence Intervals

We saw in the last set of slides that we could construct confidence intervals by knowing something about how the sampling distribution of the parameter that we care about is distributed.

In certain cases (we saw specifically for the case of a mean), that we can know analytically the sampling distribution. (Eg. $\frac{\bar{x} - \mu}{s/\sqrt{n}}$ follows a t -distribution).

Problem: These relied on certain assumptions (population approximately normal, large enough sample size), or relied on us only wanting to estimate particular parameters. Otherwise, our conclusions about the sampling distributions might not hold.

The Bootstrap

Big Idea: Rather than analytically trying to determine what the sampling distribution is, we can just approximate it.

We take a sample, and this gives us an approximation of the probability distribution function (pdf) for the population. We can then draw samples from this (by resampling with replacement from our sample), and look at how the relevant sample statistics are distributed.

Once we have a good idea about the sampling distribution, we can use this to construct our confidence interval.

The Bootstrap

(After widget demonstration)

General recipe for a 95% confidence interval:

1. Take a sample of size n from the population.
2. Find the statistic of interest for this sample.
3. Draw a large number (say, 10,000) of samples, with replacement, of size n from the original sample.
4. For each sample, calculate the statistic of interest.
5. Find the 2.5th and 97.5th percentile, a and b of the calculated statistics.
6. The 95% bootstrap confidence interval is

$$[m - (b - m), m + (m - b)]$$