These notes come from Chapter 10 of the textbook and provide an introduction to resampling methods for conducting inference, particularly the bootstrap.

Resampling Methods

H: 10.1

Our approach to inference so far has been to establish the limiting distribution of some parameter of interest; for example, $\sqrt{n}(\hat{\theta} - \theta) \xrightarrow{d} N(0, \mathbf{V}_\theta)$, and then to construct an estimate of $\mathbf{V}_\theta$. Given this estimate, we could construct a test statistic, for example a t-statistic for some $\mathbb{H}_0$, or construct a confidence interval, etc.

The idea of the resampling methods that we'll study in this section are, essentially, to substitute computational power for the (potentially complex) mathematical calculations that we have been using before. Resampling methods are popular in many applications. For example, the bootstrap is popular in quantile regression applications (which we'll talk about if we have time this semester) where (i) it is relatively complicated to complicate the asymptotic distribution and (ii) even after you calculate the asymptotic distribution, it is relatively hard to estimate it.

The book talks briefly about two resampling methods that I'll just briefly mention here. The **jackknife** is the distribution from n leave-one-out estimators (e.g., repeatedly estimating θ using all observations except particular ones). **Sub-sampling** is like the boostrap that we'll talk about below except that you draw subsamples of the original data (with less than n observations) without replacement.

The Bootstrap Algorithm

H: 10.6

There are several variations of the bootstrap, but let's start with the most common one, which is typically called either the **nonparametric bootstrap** or the **empirical bootstrap**.

Step 1: Construct a **bootstrap sample** by making n iid draws, with replacement, from the original sample. We'll denote particular draws by (Y_i^*, X_i^*), and the entire bootstrap sample by $\{Y_i^*, X_i^*\}_{i=1}^n$.

Step 2: Construct the bootstrap estimate $\hat{\theta}^*$ by applying whatever approach you originally used to estimate $\hat{\theta}$ to the bootstrap sample. For example, if you are interested in the linear projection model, you would estimate $\hat{\beta}^*$ by the linear regression of Y_i^* on X_i^*.

Steps 1 and 2 give us an estimate from the distribution of estimates obtained by iid sampling from the original data. However, the real usefulness of the bootstrap, is that (unlike our original sample from the population), we can repeat this process a large number of times. In particular, let B denote the number of bootstrap samples that we draw, then, for $b = 1, \ldots, B$, we can draw new bootstrap samples and calculate $\hat{\theta}_b^*$, where the subscript indicates that it is the bootstrap estimate from the b^{th} bootstrap sample.

Other Types of Bootstrap Procedures

The nonparametric bootstrap procedure above is the most common one, but there are other variations that are worth mentioning.

The **weighted bootstrap** involves perturbing (i.e., causing it to vary) the objective function for some particular estimation procedure. For example, if you were trying to estimate $\mathrm{E}[Y]$, the bootstrap estimate would be given by

$$\hat{\mu}^* = \arg\min_b \frac{1}{n}\sum_{i=1}^n w_i(Y_i - m)^2$$

where w_i are iid weights (in particular, they are weights that are independent of each other and independent of the original data) that satisfy $\mathrm{E}[w] = 1$ and $\mathrm{var}(w) = 1$. A leading choice is to make iid draws from an exponential distribution with mean 1 (in R, you can run `rexp(n)`). After solving this, you would get $\hat{\mu}^* = \frac{1}{n}\sum_{i=1}^n w_i Y_i$.

Similarly, if you to compute bootstrap estimates of β from a regression, it would amount to computing

$$\hat{\beta}^* = \arg\min_b \frac{1}{n}\sum_{i=1}^n w_i(Y_i - X_i'b)^2$$

If you solve this, you will get

$$\hat{\beta}^* = \left(\frac{1}{n}\sum_{i=1}^n w_i X_i X_i'\right)^{-1} \frac{1}{n}\sum_{i=1}^n w_i X_i Y_i$$

It is worth pointing out that this approach is actually quite similar to the nonparametric bootstrap. In that case, the weights are from a multinomial distribution. They have mean 1 and variance 1. However, they are not independent of each other.

Another common approach is the **multiplier bootstrap** (sometimes this is called the **score bootstrap**). In this case, bootstrap draws are constructed by perturbing the "score"/"influence function" (i.e., the part of the). For example, if we go back to the regression setup, we would compute bootstrap estimates by

$$\hat{\beta}^* = \hat{\beta} + \left(\frac{1}{n}\sum_{i=1}^n X_i X_i'\right)^{-1} \frac{1}{n}\sum_{i=1}^n w_i X_i \hat{e}_i$$

where w_i are iid weights with $\mathrm{E}[w] = 0$ (note that this is different from the weighted bootstrap) and $\mathrm{var}(w) = 1$. Common choices are (i) $W \sim N(0,1)$ or (ii) $W = 1$ with probability $1/2$ and $W = -1$ with probability $1/2$.

There are other variations of the bootstrap that we'll not cover; if you are interested, H: 10.29 covers the wild bootstrap, which is another popular version of the bootstrap, and (I think) is commonly used for nonparametric regression.

Bootstrap Variance and Standard Errors

H: 10.7

Once we have a large number of bootstrap estimates, we can estimate features of the bootstrap distribution of $\hat{\theta}_b^*$. The **bootstrap estimator of variance** of $\hat{\theta}$ is given by

$$\hat{\mathbf{V}}_{\hat{\theta}}^{boot} = \frac{1}{B-1}\sum_{b=1}^{B}\left(\hat{\theta}_b^* - \bar{\theta}^*\right)\left(\hat{\theta}_b^* - \bar{\theta}^*\right)'$$

where

$$\bar{\theta}^* = \frac{1}{B}\sum_{b=1}^{B}\hat{\theta}_b^*$$

Side-Comment: As a side-comment, notice there are a couple of small differences here relative to what we have done before. First, dividing by $B-1$ (rather than B) in the first term is a degree-of-freedom adjustment. Second, $\hat{\mathbf{V}}_{\hat{\theta}}^{boot}$ is an estimate of the variance of $\hat{\theta}$; it will typically converge to 0 as $n \to \infty$, and you could potentially multiply by n. This is similar to the distinction between $\mathbf{V}_{\hat{\beta}}$ (the variance of $\hat{\beta}$) and $\mathbf{V}_{\beta}$ (the asymptotic variance of $\sqrt{n}(\hat{\beta}-\beta)$). I typically do multiply this term by n to get an estimate of the asymptotic variance, but I am just going to follow the book here. I don't think it matters much except that you need to keep it straight what you are actually estimating.

When $\hat{\theta}$ is a scalar, the **bootstrap standard error** is given by

$$\widehat{\text{s.e.}}_{\hat{\theta}}^{boot} = \sqrt{\hat{\mathbf{V}}_{\hat{\theta}}^{boot}}$$

As in the previous set of notes, it would be very common in applications to report $\hat{\theta}$ and $\widehat{\text{s.e.}}_{\hat{\theta}}^{boot}$. Moreover, bootstrap standard errors can be used to construct confidence intervals; e.g.,

$$C^{nb} = \left[\hat{\theta} \pm 1.96\ \widehat{\text{s.e.}}_{\hat{\theta}}^{boot}\right]$$

where (I think) "nb" stands for "normal approximation bootstrap". We'll talk about other approaches to construct bootstrap confidence intervals below.

As an additional comment, although one would typically choose B to be a large number, it is still finite. This means that all bootstrap statistics, e.g., $\hat{\mathbf{V}}_{\theta}^{boot}$ are estimates and therefore are random. In particular, this means that its value will change if you were to compute them more than once. This is to be expected, though typically they should be "close" if you were to compute them more than once.

Percentile Interval

H: 10.8

This is another approach to constructing a confidence interval using the bootstrap. Recall that we have constructed $\{\hat{\theta}_1^*, \ldots, \hat{\theta}_B^*\}$. Let q_α^* denote the α quantile of these bootstrap estimates (for example, if $\alpha = .05$, this would be the 5th percentile of the bootstrap estimates). A $(1 - \alpha)$ percentile bootstrap confidence interval is given by

$$C^{pc} = [q^*_{\alpha/2}, q^* 1 - \alpha/2]$$

For example, if $B = 1000$ and $\alpha = 0.05$, then $C^{pc} = [\hat{\theta}^*_{25}, \hat{\theta}^*_{975}]$ which denote the 25th smallest and 975th largest bootstrap estimates of $\hat{\theta}$. Notice that this does not require calculating any standard error (which can sometimes be useful) and, as discussed in the text, this approach may perform better than the normal approximation confidence interval in some cases as well.