Bootstrap and Subsampling

C.Conlon

February 22, 2021

Applied Econometrics II

Bootstrap

- Bootstrap takes a different approach.
  - Instead of estimating $\hat{\theta}$ and then using a first-order Taylor Approximation...
  - What if we directly tried to construct the sampling distribution of $\hat{\theta}$?
- Our data $(X_1, \ldots, X_n) \sim P$ are drawn from some measure P
  - We can form a nonparametric estimate $\hat{P}$ by just assuming that each X_i has weight $\frac{1}{n}$.
  - We can then simulate a new sample $X^* = (X_1^*, \ldots X_n^*) \sim \hat{P}$.
    - Easy: we take our data and construct n observations by sampling with replacement
  - Compute whatever statistic of X^*, $S(X^*)$ we would like.
    - Could be the OLS coefficients $\beta_1^*, \ldots, \beta_k^*$.
    - Or some function β_1^*/β_2^*.
    - Or something really complicated: estimate parameters of a game $\hat{\theta}^*$ and now find Nash Equilibrium of the game $S(X^*, \hat{\theta^*})$ changes.
  - Do this B times and calculate at $Var(S_b)$ or $CI(S_1, \ldots, S_b)$.

Bootstrap: Bias Correction

The main idea is that $\hat{\theta}^{1*}, \ldots, \hat{\theta}^{B*}$ approximates the sampling distribution of $\hat{\theta}$. There are lots of things we can do now:

- We already saw how to calculate $Var(\hat{\theta}^{1*}, \ldots, \hat{\theta}^{B*})$.

$$\frac{1}{B-1}\sum_{b=1}^{B}(\hat{\theta}^{*}_{(b)} - \overline{\theta^{*}})^2$$

- Calculate $E(\hat{\theta}^{*}_{(1)}, \ldots, \hat{\theta}^{*}_{(B)}) = \overline{\theta^{*}} = \frac{1}{B}\sum_{b=1}^{B}\hat{\theta}^{*}_{(b)}$.

Bootstrap: Bias Correction

- We can use the estimated bias to bias correct our estimates

$$\begin{aligned} Bias(\hat{\theta}) &= E[\hat{\theta}] - \theta \\ Bias_{bs}(\hat{\theta}) &= \overline{\theta^*} - \hat{\theta} \end{aligned}$$

Recall $\theta = E[\hat{\theta}] - Bias[\hat{\theta}]$:

$$\hat{\theta} - Bias_{bs}(\hat{\theta}) = \hat{\theta} - (\overline{\theta^*} - \hat{\theta}) = 2\hat{\theta} - \overline{\theta^*}$$

- Correcting bias isn't for free - variance tradeoff!
- Linear models are (hopefully) unbiased, but most nonlinear models are consistent but biased.

Bootstrap: Confidence Intervals

There are actually three ways to construct bootstrap CI's:

1. Obvious way: sort $\hat{\theta}^*$ then take $CI : [\hat{\theta}^*_{\alpha/2}, \hat{\theta}^*_{1-\alpha/2}]$.
2. Asymptotic Normal: $CI : \hat{\theta} \pm 1.96\sqrt{V(\hat{\theta}^*)}$. (CLT).
3. Better Way: let $W = \hat{\theta} - \theta$. If we knew the distribution of W then: $Pr(w_{1-\alpha/2} \leq W \leq w_{\alpha/2})$:

   $$CI : [\hat{\theta} - w_{1-\alpha/2}, \hat{\theta} - w_{\alpha/2}]$$

   We can estimate with $W^* = \hat{\theta}^* - \hat{\theta}$.

   $$CI : [\hat{\theta} - w^*_{1-\alpha/2}, \hat{\theta} - w^*_{\alpha/2}] = [2\hat{\theta} - \theta^*_{1-\alpha/2}, 2\hat{\theta} - \theta^*_{\alpha/2}]$$

   Why is this preferred? Bias Correction!

Bootstrap: Why do people like it?

- Econometricians like the bootstrap because under certain conditions it is higher order efficient for the confidence interval construction (but not the standard errors).
  - Intuition: because it is non-parametric it is able to deal with more than just the first term in the Taylor Expansion (actually an Edgeworth Expansion).
  - Higher-order asymptotic theory is best left for real econometricians!
- Practitioner's like the bootstrap because it is easy.
  - If you can estimate your model once in a reasonable amount of time, then you can construct confidence intervals for most parameters and model predictions.

Bootstrap: When Does It Fail?

- Bootstrap isn't magic. If you are constructing standard errors for something that isn't asymptotically normal, don't expect it to work!
- The Bootstrap exploits the notion that your sample is IID (by sampling with replacement). If IID does not hold, the bootstrap may fail (but we can sometimes fix it!).
- Bootstrap depends on asymptotic theory. In small samples weird things can happen. We need $\hat{P}$ to be a good approximation to the true P (nothing missing).

Bootstrap: Variants

The bootstrap I have presented is sometimes known as the nonparametric bootstrap and is the most common one.

Parametric Bootstrap ex: if $Y_i = \beta_0 + \beta_1 X_i + \epsilon_i$ then we can estimate $(\hat{\beta}_0, \hat{\beta}_1)$ via OLS.

Now we can generate a bootstrap sample by drawing an x_i at random with replacement $\hat{\beta}_0 + \hat{\beta}_1$ and then drawing independently from the distribution of estimated residuals $\hat{\epsilon}_i$.

Wild Bootstrap Similar to parametric bootstrap but we rescale ϵ_i to allow for heteroskedasticity

Block Bootstrap For correlated data (e.g.: time series). Blocks can be overlapping or not.

Bootstrap vs Delta Method

- Delta Method works best when working out Jacobian $D(\theta)$ is easy and statistic is well approximated with a linear function (not too curvy).
- I would almost always advise Bootstrap unless:
    - Delta method is trivial e.g.: β_1/β_2 in linear regression.
    - Computing model takes many days so that 10,000 repetitions would be impossible.
- Worst case scenario: rent time on Amazon EC2!
    - I "bought" over $1,000 of standard errors recently.
- But neither is magic and both can fail!

Subsampling

The bootstrap has a close cousin subsampling.

- In practice it looks similar, but the underlying theory is quite different.
- It relies on weaker assumptions and works even in some cases where the bootstrap fails.
- Again "fails" means that the 95% confidence interval has coverage that isn't very close to 95%.

Subsampling: How does it work?

1. Draw a smaller sample $X^* = (X_1^*, \ldots, X_{a_n}^*)$ without replacement of size a_n where as $n \to \infty$ we have $a_n \to \infty$ and $\frac{a_n}{n} \to 0$.
    - e.g. $a_n = \log n$ or $a_n = \sqrt{n}$. Note that $a_n/10$ doesn't work.
2. Compute the relevant statistic $\theta(X^*)$ or $g(\theta(X^*))$.
3. Repeat this B times and construct the CDF:

$$L_n(t) = \frac{1}{B} \sum_{b=1}^{B} \mathrm{I}\left(\sqrt{a_n}(\widehat{\theta}_b - \widehat{\theta}_n) \leq t\right)$$

4. Calculate the quantiles of the CDF and CI:

$$\hat{t}_{\alpha/2} = L_n^{-1}(\alpha/2), \quad \hat{t}_{1-\alpha/2} = L_n^{-1}(1-\alpha/2)$$

$$C_n = \left[\hat{\theta}_n - \frac{\hat{t}_{1-\alpha/2}}{\sqrt{n}}, \hat{\theta}_n - \frac{\hat{t}_{\alpha/2}}{\sqrt{n}}\right]$$

Subsampling: Caveats

- The proof for why subsampling works is complicated. See `https://web.stanford.edu/~doubleh/lecturenotes/lecture13.pdf`.
- Downsides:
    - Subsampling really leans on $n \to \infty$ more than bootstrap. People often use bootstrap to understand finite sample performance (is this a good idea though?).
    - Choice of a_n is difficult. Calculating the optimal value can be quite complicated and there aren't great rules of thumb.
- But if you're in a weird case where bootstrap fails (parameter on the boundary, etc.) try subsampling and see!

Thanks!