R Lab 6: Inference

topics:

- Inference on ATE
- Final project guidance
- Parallel processing!!!

IMPT REMINDERS

- Josh: remember to record this

- Solutions to this lab will be posted soon too; R code will be very helpful for final project

ATE inference

- Reminder: ATE is not typically equivalent to the coefficient in a model. We need different ways of calculating standard errors, CIs, P-values.
- Big idea: Can use influence curve in certain situations, and non-parametric bootstrap in certain (most) situations.
- IC variance estimates can be conservative … but …
- Bootstrap variance estimates are slow!

ATE inference

- g-comp / simple substitution: IF outcome regression $\mathbb{E}(Y \mid A, W)$ is correctly specified, can use Delta Method or non-parametric bootstrap. BUT you probably aren't correctly specified. Recommendation: Bootstrap.

- IPTW: Same. Recommendation: Bootstrap.

- TMLE: Can use bootstrap, but faster to use the influence curve!

The influence curve for TMLE for observation i at the true data generating distribution $\mathbb{P}_0$ is given by

$$IC(O_i) = \left(\frac{\mathbb{I}(A_i = 1)}{\mathbb{P}_0(A = 1|W_i)} - \frac{\mathbb{I}(A_i = 0)}{\mathbb{P}_0(A = 0|W_i)} \right) \left[Y_i - \mathbb{E}_0(Y|A_i, W_i) \right] + \mathbb{E}_0(Y|A = 1, W_i) - \mathbb{E}_0(Y|A = 0, W_i) - \Psi(\mathbb{P}_0)$$

Practical issues with bootstrap

- Especially if using Super Learner, can be SLOW.
- Requires certain conditions that are rarely checked. Check that the distro is reasonably normal-looking!
- Common mistake: Get your point estimate first, then use the variance of the bootstrap samples to get a +/- 1.96*SE(bootstrap samples) to get the CI… or, alternatively, just use the 2.5% and 97.5% quantiles of the bootstrap distribution to get the CI. This is actually a complex topic, we will just gloss over here.
- Demo coming up!

TMLE IC subtleties

- Two estimated functions: outcome regression $\bar{Q}$ and propensity score g…
    - IF g is consistently estimated, IC inference will work, though may be conservative (not asymptotically efficient)
    - IF g is known exactly, as in an RCT, IC inference works and will be asymptotically efficient
    - IF $\bar{Q}$ is also consistently estimated, then IC inference works and will be asymptotically efficient.
- Our goal in SL is to have flexible enough algorithms to make consistent estimation of both more likely.
- Caveat: Positivity violations or very rare outcomes can make IC-based inference anti-conservative… be careful.

Let’s try it!