Interval Estimates of Causal effects

April 9, 2014

Standard Errors of ATE Estimate

- ATE = E(Y(1)-Y(0)).
- Point estimate of ATE in randomized experiments:

$$\pi = \sum_i(Y_i Z_i)/\sum_i(Z_i) - \sum_i(Y_i(1-Z_i))/\sum_i(1-Z_i)$$

$$= \sum_{i:Zi=1} Y_i/n_1 - \sum_{i:Zi=0} Y_i / n_0$$

- Variance of π:

$$\text{var}(\pi) = \sum_{i:Zi=1} \text{var}(Y_i)/(n_1)^2 + \sum_{i:Zi=0} \text{var}(Y_i)/(n_0)^2$$

- Usually assume variance of Y_i are the same, σ^2 within each group.

Standard Errors: Delta Method

- Then: var(π)= $\sigma_1^2 / n_1 + \sigma_0^2 / n_0$
- Now we only need estimate σ^2
- One easy way is to estimate σ^2 by the variance of the observed Y's in each group.
- The s.e. of ATE estimate π is the square root of var(π).
- 95% CI of π is: $\pi \pm 1.96 * se(\pi)$

Limitations of Delta Method

- But this method (delta method) can be difficult when the causal estimand (and thus estimator) is more complicated.
- For example, the CACE estimator is a ratio of two estimators.
- Also, using the empirical variance to estimate the true variance is not always reliable.
- Alternative: re-sampling – bootstrap.

Bootstrap

- Step 1: Sample *n* units with replacement from the *n* observed units. Calculate the ATE or whatever estimand (like CACE) from this new sample.
- Step 2: Repeat Step 1 many times (say, 1000), record the point estimate from each sample.
- The 95% CI is just the empirical 95% CI of the 1000 point estimates.

Bootstrap for propensity score estimates

- Remember in observational studies, we often use propensity score methods to balance the covariate distributions.
- Two-stage methods: (1) Estimate prop. score; (2) Estimate ATE based on the estimated prop score from stage 1, using matching, or weighting, or subclassification.
- How can we get the s.e. of the ATE estimates from the propensity score approach?

Bootstrap for propensity score estimates

- Exact same idea: obtain a large number (say, 1000) samples from resampling with replacement from the observed sample.
- For each resample, repeat the propensity score approach (that is, fit the same logistic regression model to estimate prop score, and then use matching (or weighting or subclassification) to obtain a point estimate of ATE given the estimated prop score.
- Repeat the same analysis for each of the 1000 resamples, and record the point estimate of ATE each time.
- The 95%CI is the 95% empirical interval (i.e., the 2.5 percentile and the 97.5 percentile) of the 1000 estimated ATEs.

Bootstrap for CACE

- In dealing with randomized experiment with noncompliance, the same idea of bootstrap applies.
- Sample with replacement from the observed sample (Z, W, Y) many times.
- Calculate CACE from each of these sample, and get the 95%CI.

Bootstrap for multilevel data

- Recall the propensity score weighting methods for multilevel data
- We can again use bootstrap to estimate the standard errors.
- Key question: when there is multilevel structure, how to re-sample? Re-sample individuals or re-sample clusters?
- Resample clusters. Indeed, total number of individuals may change from sample to sample, but it is valid.