Model-based matching for causal inference in observational studies

Kellie Ottoboni

Department of Statistics, UC Berkeley
Berkeley Institute for Data Science

February 17, 2016

Observational Studies vs Experiments

- In randomized control trials, treatment is assigned to individuals at random.
- In observational studies, the way individuals select into treatment groups is unknown.
- **Confounding** occurs when a variable affects both treatment assignment and the outcome. It biases our estimates of the treatment effect.

Matching

- **Ideal:** group individuals by important confounders to estimate subgroup treatment effects and then average over subgroups
- **Reality:** many covariates, perhaps continuous, make it difficult to stratify
- **Solution:** use a one-dimensional score to match or group individuals
- Stratify according to the "best" prediction of the response, based on all covariates except for the treatment

Model-based Matching

- Under standard assumptions (unconfoundedness), the **average treatment effect** is nonparametrically identified
- Estimate it using the difference in average residuals between treated and controls
- Simulation RMSE plot

Model-based Matching

- Use stratified permutation test to test the null hypothesis of no treatment effect whatsoever
- Stratifying on $\hat{Y}$ allows us to detect non-constant and non-linear treatment effects
- Simulation power plot

Future Directions

- Do different test statistics give greater power?
- What is the optimal way to stratify?
- How to quantify uncertainty – standard errors and confidence intervals?