

Model-based matching for causal inference in observational studies

Kellie Ottoboni
with Philip B. Stark

Department of Statistics, UC Berkeley
Berkeley Institute for Data Science

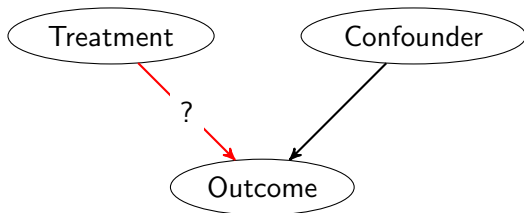
March 15, 2016

Outline

- 1 Introduction
- 2 Model-based Matching
- 3 Examples
 - Toads and Packstock in Yosemite
 - Salt and Mortality
- 4 Conclusions
- 5 References

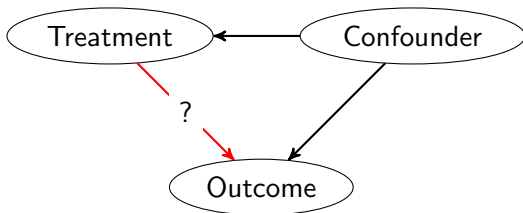
Observational Studies vs Experiments

- **Problem:** Estimate the causal effect of a treatment on outcome of interest
- In randomized experiments, treatment is assigned to individuals at random.
- In observational studies, the way individuals select into treatment groups is unknown.



Observational Studies vs Experiments

- **Problem:** Estimate the causal effect of a treatment on outcome of interest
- In randomized experiments, treatment is assigned to individuals at random.
- In observational studies, the way individuals select into treatment groups is unknown.



Neyman-Rubin Causal Model

- Population of $i = 1, \dots, N$ individuals. Each individual has two **potential outcomes**.
- $Y_i(1)$ is individual i 's outcome if he receives treatment
- $Y_i(0)$ is individual i 's outcome if he is in the control group
- The treatment effect for individual i is $\tau_i = Y_i(1) - Y_i(0)$

$Y_1(1)$	$Y_1(0)$
----------	----------

$Y_3(1)$	$Y_3(0)$
----------	----------

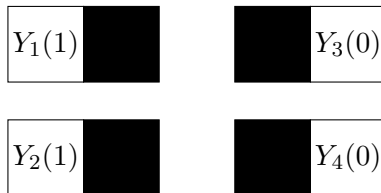
$Y_2(1)$	$Y_2(0)$
----------	----------

$Y_4(1)$	$Y_4(0)$
----------	----------

Fundamental Problem of Causal Inference [Holland, 1986]

- We may never observe both $Y_i(1)$ and $Y_i(0)$
- T_i is a treatment indicator: 1 if i is treated, 0 if i is control
- The observed outcome for individual i is

$$Y_i = T_i Y_i(1) + (1 - T_i) Y_i(0)$$



Estimands

- Average treatment effect

$$\mathbb{E}(Y_i(1) - Y_i(0))$$

- Average treatment effect on the treated

$$\mathbb{E}(Y_i(1) - Y_i(0) \mid T_i = 1)$$

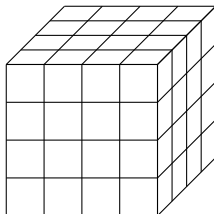
- Conditional average treatment effect

$$\mathbb{E}(Y_i(1) - Y_i(0) \mid X_i)$$

- If treatment effect varies by covariates X , then averages might not be informative

Matching

- **Ideal:** group individuals by X_i to estimate subgroup treatment effects and then average over subgroups
- **Reality:** many covariates, perhaps continuous, make it difficult to stratify

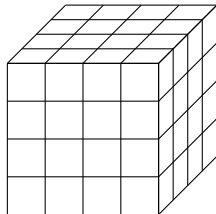


Matching

- If d covariates are split into k bins, we have d^k groups.
- To guarantee that we have at least one treated and one control in each group with 95% probability, we need

$$n \geq \frac{2 \log(1 - (0.95)^{1/k^{d+1}})}{\log(\frac{k^d - 1}{k^d})}$$

- If $d = 5$ and $k = 2$, $n \geq 225$.
- If $d = 10$ and $k = 2$, $n \geq 10,844$.



Matching

- **Solution:** use a one-dimensional score to match or group individuals

Propensity score matching

- The **propensity score** is an individual's probability of being assigned treatment, conditional on their covariates

$$p(x) = \mathbb{P}(T = 1 \mid X = x)$$

- The propensity score is a balancing score: $X \perp\!\!\!\perp T \mid p(X)$
- For individuals with the same propensity score, treatment assignment is as if random

Propensity score matching

Theorem (Rosenbaum and Rubin [1983])

If treatment assignment is independent of potential outcomes given X ,

$$(Y(1), Y(0)) \perp\!\!\!\perp T \mid X$$

and if every unit has a chance of receiving treatment,

$$0 < p(X) < 1 \text{ for all } X$$

then $(Y(1), Y(0)) \perp\!\!\!\perp T \mid p(X)$.

In particular, potential outcomes are balanced between groups

$$\mathbb{E}(Y(t) \mid T = 1, p(X)) = \mathbb{E}(Y(t) \mid T = 0, p(X)) \text{ for } t = 0, 1$$

Propensity score matching

This result identifies the average treatment effect in terms of quantities we can estimate:

$$\begin{aligned}\mathbb{E}(Y(1) - Y(0)) &= \mathbb{E}_{p(x)} [\mathbb{E}(Y(1) - Y(0) \mid p(x))] \\ &= \mathbb{E}_{p(x)} [\mathbb{E}(Y(1) \mid p(x)) - \mathbb{E}(Y(0) \mid p(x))] \\ &= \mathbb{E}_{p(x)} [\mathbb{E}(Y \mid p(x), T = 1) - \mathbb{E}(Y \mid p(x), T = 0)]\end{aligned}$$

Propensity score matching

How to use the result to estimate the average treatment effect?

- Often, the true propensity score is unknown. Estimate $\hat{p}(x)$ using logistic or probit regression
- Match or stratify units on $\hat{p}(x)$
- Within groups, take difference in mean outcomes between treated and controls
- Average estimated treatment effects across groups

Propensity score matching

In practice, estimating $\hat{p}(x)$ causes problems.

- Logistic and probit regression assume a functional form for relationship between covariates and treatment
- Assumes that probability of treatment takes same functional form for all individuals
- May actually worsen balance if estimated incorrectly
- Standard errors are difficult to compute for estimates from matching

Outline

- 1 Introduction
- 2 Model-based Matching
- 3 Examples
 - Toads and Packstock in Yosemite
 - Salt and Mortality
- 4 Conclusions
- 5 References

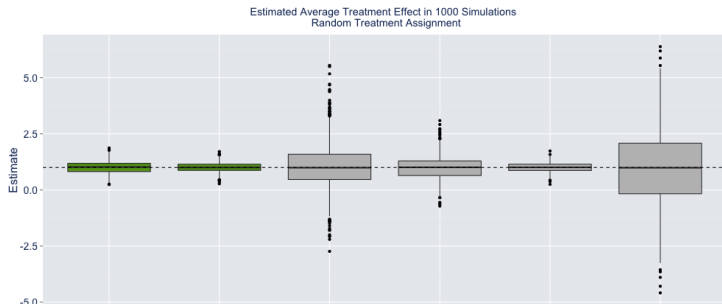
Model-based Matching

Idea: Instead of modeling the propensity score, model the outcome

Stratify on \hat{Y} , the “best” prediction of the response based on all covariates except for the treatment

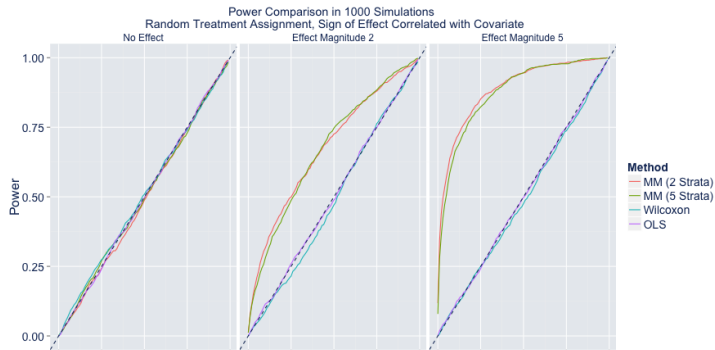
Model-based Matching

- Under standard assumptions (conditional independence of treatment and potential outcomes given X), the **average treatment effect** is nonparametrically identified
- Estimate it using the difference in average residuals, $Y - \hat{Y}$, between treated and controls



Model-based Matching

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



Outline

1 Introduction

2 Model-based Matching

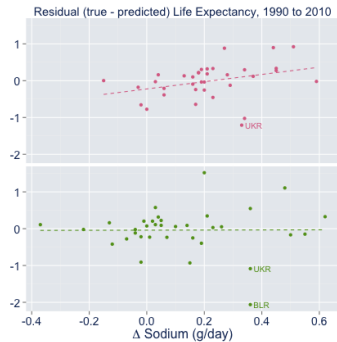
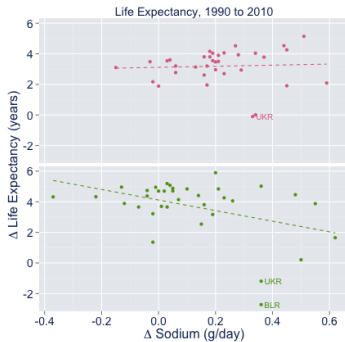
3 Examples

- Toads and Packstock in Yosemite
- Salt and Mortality

4 Conclusions

5 References

Salt



Female Male

Outline

- 1 Introduction
- 2 Model-based Matching
- 3 Examples
 - Toads and Packstock in Yosemite
 - Salt and Mortality
- 4 Conclusions
- 5 References

Future Directions

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

Outline

- 1 Introduction
- 2 Model-based Matching
- 3 Examples
 - Toads and Packstock in Yosemite
 - Salt and Mortality
- 4 Conclusions
- 5 References

References

- Paul W. Holland. Statistics and Causal Inference. *Journal of the American Statistical Association*, 81(396):945–960, 1986. ISSN 0162-1459. doi: 10.2307/2289064. URL <http://www.jstor.org/stable/2289064>.
- Paul R. Rosenbaum and Donald B. Rubin. The central role of the propensity score in observational studies for causal effects. *Biometrika*, 70(1):41–55, April 1983. ISSN 0006-3444, 1464-3510. doi: 10.1093/biomet/70.1.41. URL <http://biomet.oxfordjournals.org/content/70/1/41>.