

INTRODUCTION AND RANDOMIZED CONTROL TRIALS

JIALIANG CHEN

ABSTRACT. This serves as complementary material to Prof. Jeffery Smith's lecture notes.

CONTENTS

1. Introduction	2
1.1. Fundamental Evaluation Problem	2
2. Randomized Control Trial	2

1. INTRODUCTION

This part of the course is entirely design-based and will focus only on treatment effects. The interest will always be only one single parameter. We will go through *Randomized Control Trials*, *Heterogeneous Treatment Effects*, *Non-parametric Regression*, *Matching and Weighting estimators* and *Model Selection and Justification*.

1.1. Fundamental Evaluation Problem.

Consider the ATET parameter

$$\mathbb{E}[Y_1 - Y_0 \mid D = 1] = \mathbb{E}[Y_1 \mid D = 1] - \mathbb{E}[Y_0 \mid D = 1].$$

The first term is not a big deal. The second term, however, is the unobserved counterfactual mean, and is hard to get. Different methods depending on contexts, data and assumptions try to provide a good estimate of this term. For example, experiments solve this by forcing the units that would otherwise be treated to experience the untreated state. All of the fancy methods including Discontinuity Design, Difference-in-Difference, matching, inverse probability weighting are trying to get an estimate of this.

2. RANDOMIZED CONTROL TRIAL

In this section, we assume the structural equation as

$$Y_i = \beta_0 + \beta_T T_i + \epsilon_i,$$

where $i = 1, 2, 3, \dots$, $Y_i = Y_{1i}T_i + Y_{0i}(1 - D_i)$ is the observed outcome, and T_i indicates the actual treatment status (i.e., unit i indeed receives treatment or not).

Human are not lab rats. So in a randomized control trial, those people who got assigned to the treatment group might not actually get the treatment. On the other hand, those people who got assigned to the control group might get to find other (similar) treatment somewhere else. Therefore, we need a more involved model. The selection process we will be considering here is the following one.

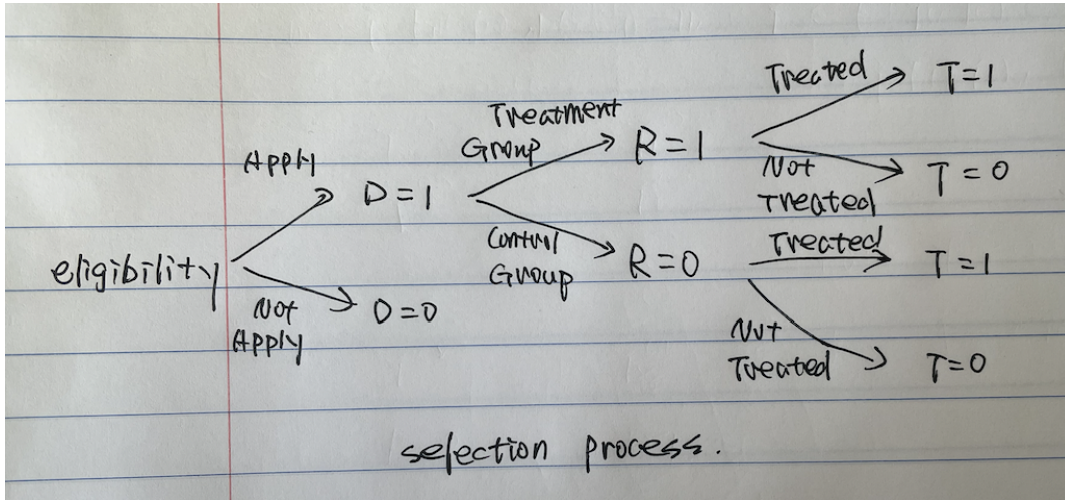


FIGURE 1. The Selection Process

First, the eligible units choose to apply or not. Then those applied units are randomly assigned to treatment and control groups. This is the world of *non-perfect compliance*. Finally, we will have *the treatment group no-shows*, that is, $D_i = 1, R_i = 1$ but $T_i = 0$, as well as *the control group substitutions*, i.e., $D_i = 1, R_i = 0$ but $T_i = 1$.

There is another complication, it is very likely that we will have *the endogeneity problem*. That is, $\mathbb{E}(T, \epsilon) \neq 0$ or a stronger (but unnecessary) version $\mathbb{E}[\epsilon \mid T] \neq 0$. In this case, we consider the random variable R , which stands for random assignment, as an instrument random variable for D . We need to have $\text{Cov}(R, D) \neq 0$ and $\mathbb{E}[R, \epsilon] = 0$ being satisfied for a legitimate instrument. The first condition is reasonable, since people will react to incentives. The second condition is guaranteed if we have random assignment.