Potential Outcomes and SUTVA

Joel Ferguson

9/12/2022

Rubin Causal Model

Many social science disciplines are primarily concerned with answering questions that reflect a causal link between quantities using observational (i.e. non-experimental) data. Two issues immediately stand out once we start thinking about how to estimate causal quantities:

- We never see a unit with multiple treatment statuses at once. This is "the fundamental problem of causal inference.''
- Units that are observed under one treatment status may have systematically different characteristics (observable or unobservable) from units observed under another status, and those characteristics may help determine the outcome of interest.

A simple example of how these issues play out in causal inference is to consider the effects of a de-worming drug (this is related to a paper we will be exploring a bit later when we discuss SUTVA). We'll simulate data in which only individuals with ex-ante heavy worm burdens opt to take de-worming medicine, leading to a perverse conclusion using ex-post observational data. In particular, all individuals with a worm burden above a threshold T will be assigned to treatment.

```
N <- 1000 # 1000 rows
worm_load_pre <- rchisq(N,1) # Worm load drawn from a chi squared dist
treat_cutoff <- quantile(worm_load_pre,probs=0.9) # Find 90th percentile to choose who gets treated
max_worm <- max(worm_load_pre) # Find the max worm load, which will be used to make POs
df <- data.frame("worm_load_pre" = worm_load_pre) %>% # Make a df
  mutate(treat = worm_load_pre>=treat_cutoff) # Those with high worm loads are treated
df$Y_1 <- (1-sqrt(df$worm_load_pre/max_worm))*df$worm_load_pre # People with higher worm load see bigge
df$Y <- df$treat*df$Y_1 + (1-df$treat)*df$worm_load_pre # Calculate the observed outcome
df$TE_i <- df$Y_1 - df$worm_load_pre # Calculate individual treatment effects
df$i <- c(1:nrow(df))

ggplot(data=df)+
  geom_density(aes(x=Y,fill=treat),alpha=0.5)+
  scale_fill_brewer(palette = "Set2")+
  labs(title="Observed Outcome Distributions",
       fill="Treatment Status",
       x="Y",y="Density")
```

Observed Outcome Distributions
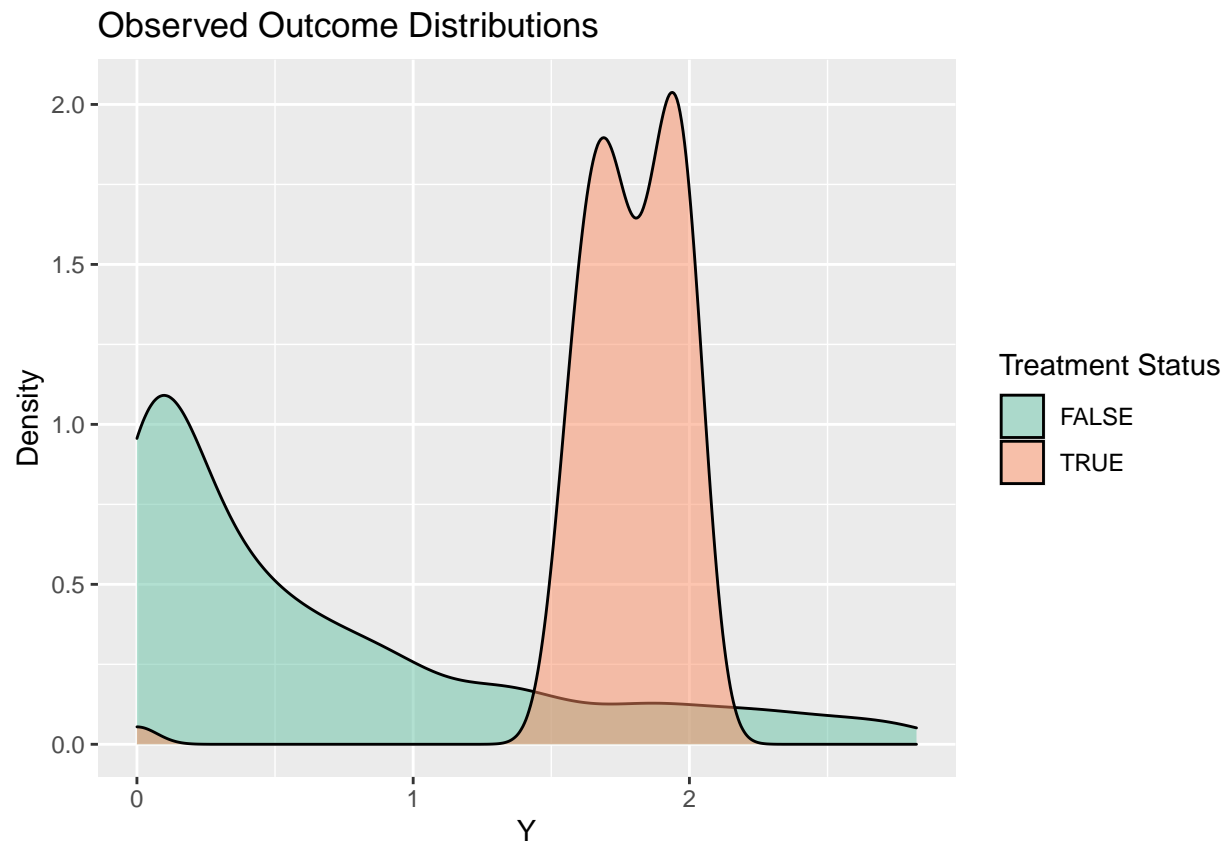

```r
ggplot(data=df %>%
         select(i,Y_0=worm_load_pre,Y_1) %>%
         pivot_longer(cols=c("Y_0","Y_1"),
                      names_to = "Potential Outcome",values_to = "PO"))+
  geom_density(aes(x=PO,fill=`Potential Outcome`),alpha=0.5)+
  scale_fill_brewer(palette = "Set2")+
  labs(title="Potential Outcome Distributions",
       x="Y(D)",y="Density")
```

Potential Outcome Distributions
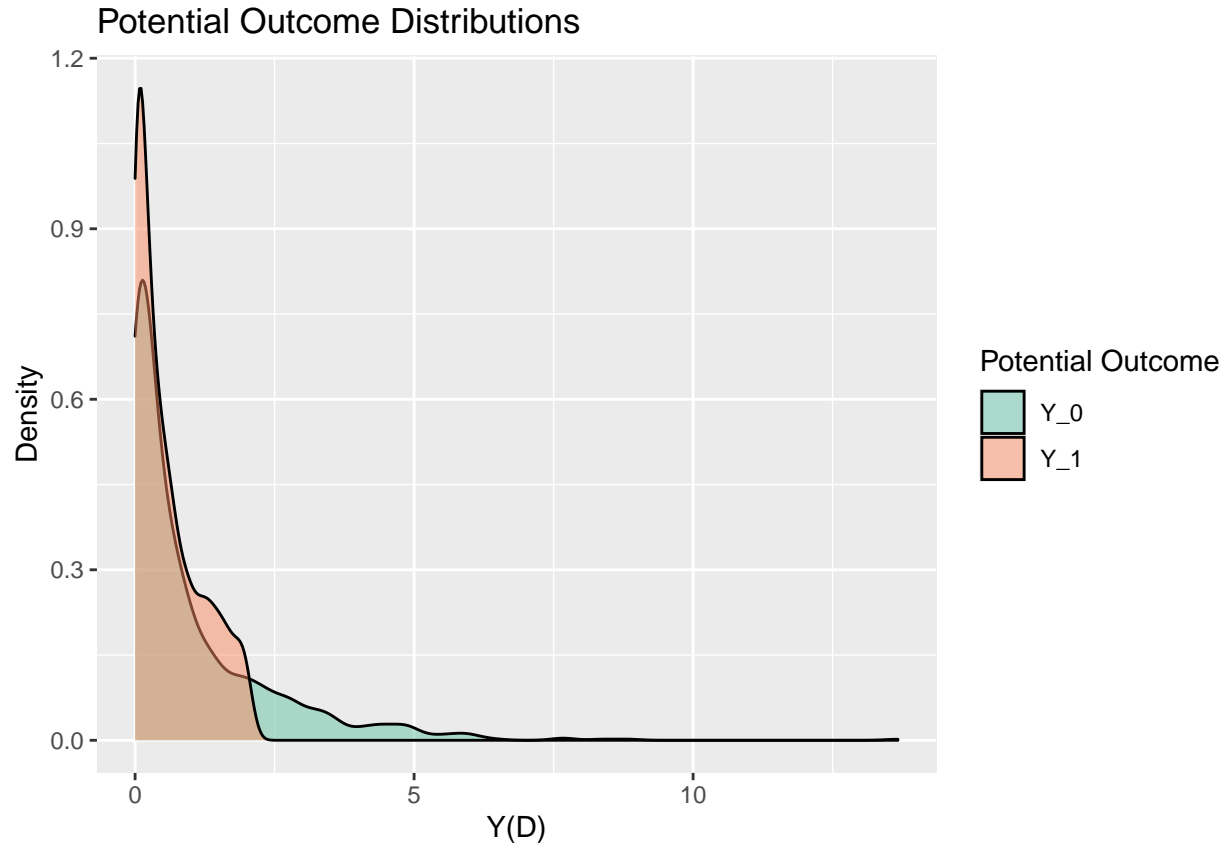

```r
ate_obs <- mean(df$Y_1[df$treat==1])-mean(df$worm_load_pre[df$treat==0])
print(paste0("ATE from observation:",ate_obs))
```

```
## [1] "ATE from observation:1.1355243716439"
```

```r
ate_true <- mean(df$TE_i)
print(paste0("True ATE:",ate_true))
```

```
## [1] "True ATE:-0.435219257831549"
```

This example shows both of the issues mentioned above in action (it also shows how simulation can help us understand causal estimation). Because in practice we don't observe the effect of medicine on worm loads for individuals who don't take the medicine and individuals who do take the medicine have high worm loads, the naive observational study in this simulation concludes that de-worming medicine *increases* worm load. However, using our full knowledge of the data generating process, we know that this is not the case.

More formally, the problem is that treatment status D_i, is a function of the untreated potential outcome $Y_i(0)$. As such, we can write the observational difference in means as

$$\tau^{obs} = E[Y_i|D_i=1]-E[Y_i|D_i=0] = E[Y_i(1)|D_i=1]-E[Y_i(0)|D_i=0] = E[Y_i(1)|Y_i(0) \geq T]-E[Y_i(0)|Y_i(0) < T]$$

This is very clearly different from the (or at least an) object of interest, the average treatment effect $\tau = E[Y_i(1)-Y_i(0)]$. In general, we can only hope to recover a causal effect if treatment assignment is *independent* of potential outcomes (conditional on some covariates). When this is satisfied, conditioning on treatment status provides no additional information to the expected value of potential outcomes (again, conditional on some covariates).

SUTVA

While the independence of treatment status and potential outcomes is necessary for causal identification, it is not sufficient. In general, we also require a restriction on how potential outcomes vary with the treatment assignment *vector*. In the example above, potential outcomes only depended on each individual's own treatment status. Thus, we implicitly enforced what is known as the **Stable Unit Treatment Value Assumption**, which essentially states that an individual unit's potential outcomes do not depend on other units' treatment statuses. Mathematically, this can be expressed as

$$E[Y_i(D_i)] = E[Y_i(D_i)|D]$$

where D represents the treatment vector.

Why do we need SUTVA (or some other restriction on the relationship between potential outcomes and treatment assignment)? Consider the case where every unit's potential outcomes can differ with every treatment assignment vector. In this case, an experiment (i.e. randomization of treatment statuses) only recovers a treatment effect that is local in the sense that it is sensitive to the treatment assignment vector D. In fact, in the most general version of this scenario this estimate doesn't necessarily tell us *anything* about how a unit's potential outcomes would be different under an alternative realization of treatment assignment.

Let's use another simulation to see this issue in action. We're going to generate a sample in which SUTVA does not hold in a very drastic way: individuals' potential outcomes are random functions of the entire treatment vector. Under this data generating process, a single experiment cannot recover the average treatment effect; we would require one experiment for *every* possible treatment assignment vector[1]!

```
N <- 10 # Need a small N this time because many possible treatment vectors
Y_0 <- matrix(rnorm(10*2^10), # Draw untreated POs for every treatment vector
             nrow=N)
Y_1 <- matrix(sapply(c(1:2^10), function(x) rnorm(10,mean=runif(1,0,10))), # Same for treated POs but w
             nrow = N)

# Function to map a base 10 int to a length-10 binary vector
makeD <- function(x){
  D <- as.vector(as.binary(x))
  D <- c(rep.int(0,times=10-length(D)),D)
  return(D)
}

Ds <- as.matrix(sapply(c(1:2^10-1),makeD)) # Make treatment assignment vectors
Y <- Ds*Y_1 - (1-Ds)*Y_0 # Make observed outcomes
D <- sample(c(1:2^10),1) # Choose a treatment assignment

# Function to calculate ATE for a particular experiment
calc_ate <- function(D){
  tau_exp <- mean(Y[as.logical(Ds[,D]),D])-mean(Y[as.logical(1-Ds[,D]),D])
  return(tau_exp)
}

tau_D <- calc_ate(D) # Get ATE for the randomly chosen treatment vector
tau <- mean(sapply(c(1:2^10), calc_ate),na.rm = T) # Calculate ATE for every experiment (can't calculat

print(paste("Experiment ATE:",tau_D))
```

---

[1]When we make relaxations like this, we need to be very careful about what ATE even means. We usually suppress the distribution we're taking expectations over, which can make things confusing when there are multiple sources of randomness (e.g. randomized treatment and other determinents of outcomes). Here I take ATE to be defined as $\tau = \mathbb{E}_D[Y_i(1) - Y_i(0)]$

```
## [1] "Experiment ATE: 7.44360321808974"
```

```
print(paste("True ATE:",tau))
```

```
## [1] "True ATE: 4.9120904217483"
```

As this example shows, when SUTVA doesn't hold our experiment estimates something different from what we (generally) seek to recover. However, even when SUTVA doesn't hold, we can still place restrictions on how potential outcomes vary with treatment vectors to uncover the causal quantity of interest from a single experiment. This is done by essentially defining equivalence classes of experiments at the unit level. SUTVA defines very large equivalence classes: every experiment in which the unit of interest receives the same treatment assignment is equivalent. Our example above has very small equivalence classes: they are all singletons (i.e. each experiment is its own equivalence class). Naturally, there are intermediate cases.

A famous example of an experiment that uses a relaxed assumption of potential outcome invariance to treatment assignment in place of SUTVA is *Miguel and Kremer (2004)*, in which the effectiveness of de-worming medication in increasing school attendance in Kenya was evaluated. As worm infections are communicable, de-worming provides a positive spillover externality: an untreated student near many treated student may experience a reduction in worm load due to lower prevalence in her environment. Previous studies which had randomly assigned treatment at the individual level had found small effects on attendance, likely due to these spillover effects. Miguel and Kremer sought to limit spillovers by randomizing treatment status at the *school* level.

However, even this clever modification likely cannot fully limit spillovers to the unit of randomization. The proposed solution: parametrically define spillover effects.

We first estimate program impacts in treatment schools, as well as cross-school treatment externalities:[24]

$$(1) \qquad Y_{ijt} = a + \beta_1 \cdot T_{1it} + \beta_2 \cdot T_{2it} + X'_{ijt}\delta + \sum_d (\gamma_d \cdot N^T_{dit}) + \sum_d (\phi_d \cdot N_{dit})$$

$$+ u_i + e_{ijt}.$$

Y_{ijt} is the individual health or education outcome, where i refers to the school, j to the student, and $t \in \{1, 2\}$ to the year of the program; T_{1it} and T_{2it} are indicator variables for school assignment to the first and second year of deworming treatment, respectively; and X_{ijt} are school and pupil characteristics. N_{dit} is the total number of pupils in primary schools at distance d from school i in year t, and N^T_{dit} is the number of these pupils in schools randomly assigned to de-worming treatment. For example, in Sections 5 and 6, $d = 03$ denotes schools that are located within three kilometers of school i, and $d = 36$ denotes schools that are located between three to six kilometers away.[25] Individual disturbance terms are assumed to be independent across schools, but are allowed to be correlated for observations within the same school, where the school effect is captured in the u_i term.

The assumption here is that potential outcomes only depend on the number of treated students within 3 kilometers and 3-6 kilometers away. This appears to be an important factor to consider empirically, as (at least in some specifications) these spillovers appear to attenuate treatment effects.

TABLE IX

SCHOOL PARTICIPATION, DIRECT EFFECTS AND EXTERNALITIES[a]
DEPENDENT VARIABLE: AVERAGE INDIVIDUAL SCHOOL PARTICIPATION, BY YEAR

	OLS (1)	OLS (2)	OLS (3)	OLS (4) May 98– March 99	OLS (5) May 98– March 99	OLS (6) May 98– March 99	IV-2SLS (7) May 98– March 99
Moderate-heavy infection, early 1999						−0.028*** (0.010)	−0.203* (0.094)
Treatment school (T)	0.051*** (0.022)						
First year as treatment school (T1)		0.062*** (0.015)	0.060*** (0.015)	0.062* (0.022)	0.056*** (0.020)		
Second year as treatment school (T2)		0.040* (0.021)	0.034* (0.021)				
Treatment school pupils within 3 km (per 1000 pupils)			0.044** (0.022)		0.023 (0.036)		
Treatment school pupils within 3–6 km (per 1000 pupils)			−0.014 (0.015)		−0.041 (0.027)		
Total pupils within 3 km (per 1000 pupils)			−0.033** (0.013)		−0.035* (0.019)	0.018 (0.021)	0.021 (0.019)
Total pupils within 3–6 km (per 1000 pupils)			−0.010 (0.012)		0.022 (0.027)	−0.010 (0.012)	−0.021 (0.015)
Indicator received first year of deworming treatment, when offered (1998 for Group 1, 1999 for Group 2)					0.100*** (0.014)		
(First year as treatment school Indicator) ∗ (Received treatment, when offered)					−0.012 (0.020)		
1996 district exam score, school average	0.063*** (0.021)	0.071*** (0.020)	0.063*** (0.020)	0.058 (0.032)	0.091** (0.038)	0.021 (0.026)	0.003 (0.023)

Ted and his coauthors have recently revisited this style of weakening SUTVA in *Egger et al (2019)*, which analyzes the effects of a large-scale cash transfer in Kenya.

$$y_{iv} = \alpha + \beta Amt_v + \sum_{r=0}^{R} \beta_r Amt_{\neg v,r} + \delta_1 y_{iv,t=0} + \delta_2 M_{iv} + \varepsilon_{iv}. \tag{2}$$

The novel terms here are the amount Amt_v of cash transferred to one's own village v over the entire study, and the amount $Amt_{\neg v,r}$ of cash transferred to villages other than v in a series of bands with inner radius r km and outer radius $r+2$ km. We normalize both to be measured

Developing alternative assumptions to SUTVA that allow treatment effects to be recovered or bounded from a single experiment is a very active area of econometric research (see e.g. Bryan Graham's work on estimation with network effects).