1

COUNTERFACTUALS

DEFINITIONS :

1 IDENTIFIABILITY

Definition 2.1 (Identifiability) *A causal quantity (e.g. $\mathbb{E}[Y(t)]$) is identifiable if we can compute it from a purely statistical quantity (e.g. $\mathbb{E}[Y \mid t]$).*

(Theorem 2.1):

$$\mathbb{E}[Y(1) - Y(0)] = \mathbb{E}_X\left[\mathbb{E}[Y \mid T = 1, X] - \mathbb{E}[Y \mid T = 0, X]\right] \quad (2.14)$$

$\mathbb{E}[Y(1) - Y(0)]$ is the causal estimand that we are interested in. In order to actually estimate this causal estimand, we must translate it into a statistical estimand: $\mathbb{E}_X\left[\mathbb{E}[Y \mid T = 1, X] - \mathbb{E}[Y \mid T = 0, X]\right]$.[15]

When we say "identification" in this book, we are referring to the process of moving from a *causal* estimand to an equivalent *statistical* estimand. When we say "estimation," we are referring to the process of moving from a statistical estimand to an estimate. We illustrate this in the flowchart in Figure 2.5.

$$\boxed{\text{Causal Estimand}} \xrightarrow{\text{Identification}} \boxed{\text{Statistical Estimand}} \xrightarrow{\text{Estimation}} \boxed{\text{Estimate}}$$

Figure 2.5: The Identification-Estimation Flowchart – a flowchart that illustrates the process of moving from a target causal estimand to a corresponding estimate, through identification and estimation.

What do we do when we go to actually estimate quantities such as $\mathbb{E}_X\left[\mathbb{E}[Y \mid T = 1, X] - \mathbb{E}[Y \mid T = 0, X]\right]$? We will often use a model (e.g. linear regression or some more fancy predictor from machine learning) in place of the conditional expectations $\mathbb{E}[Y \mid T = t, X = x]$. We will refer to estimators that use models like this as *model-assisted estimators*. Now that we've gotten some of this terminology out of the way, we can proceed to an example of estimating the ATE.

[15] Active reading exercise: Why can't we directly estimate a causal estimand without first translating it to a statistical estimand?

2. PARAMETRIC SCM = SCM

A PARAMETRIC SCM HAS 2 COMPONENTS:

a) A CAUSAL GRAPH

b) THE PARAMETRIC MODEL THAT GOVERN THE SCM. → THIS IS PART OF THE DEFINITION OF SCM

UNIT - LEVEL COUNTERFACTUAL :

$\boxed{\text{IDENTIFICATION}}$ OF UNIT - LEVEL (E.G. INDIVIDUAL) COUNTERFACTUAL REQUIRE A PARAMETRIC MODEL. A CAUSAL GRAPH IS NOT SUFFICIENT.

WHY ?.

BECAUSE THE SAME CAUSAL GRAPH MIGHT ENTAIL THE SAME OBSERVATIONAL DISTRIBUTION, BUT DIFFERENT STATISTICAL ESTIMANDS FOR THE COUNTERFACTUALS DEPENDING ON THE PARAMETRIC MODEL.

EXAMPLES OF THIS:

REVIEW THIS EXAMPLE

PEARL "CAUSALITY" BOOK, PAGE 33-38.

BUT THIS EXAMPLE IS NOT GREAT BECAUSE MODEL 1 AND MODEL 2 CORRESPOND TO DIFFERENT SCMS

ECI BOOK, PAGE 98

PEARL "CAUSALITY PRIMER" BOOK. CH 4, P104

POPULATION-LEVEL COUNTERFACTUALS

- ATT = AVERAGE TREATMENT EFFECT ON THE TREATED

$$E[Y(1) - Y(0) \mid T=1]$$

- CATT = CONDITIONAL AVERAGE TREATMENT EFFECT ON THE TREATED $E[Y(1)-Y(0) \mid T=1, X]$

(4)

SO ATT AND CATT ARE EXAMPLES OF CAUSAL ESTIMANDS THAT INVOLVE POPULATION-LEVEL COUNTERFACTUALS

. IDENTIFYING POPULATION-LEVEL COUNTERFACTUALS DO NOT REQUIRE ASSUMING A PARAMETRIC MODEL. HOWEVER WE REQUIRE ASSUMING WE CAN CONDITION ON A SET {Z} OF VARIABLES SATISFYING THE BACK DOOR CRITERION (CONDITIONAL IGNORABILITY)

THIS IS EQUIVALENT TO SAY THAT WE NEED TO ASSUME A CAUSAL GRAPH, BUT WE DON'T NEED TO ASSUME A PARAMETRIC FORM FOR THE EQUATIONS THAT GOVERN THE SCM FOR THE SAKE OF IDENTIFICATION.

OF COURSE, FOR ESTIMATION, WE NEED TO ASSUME A PARAMETRIC MODEL, BUT

AGAIN, WE CAN GET A STATISTICAL ESTIMAND WITHOUT A PARAMETRIC MODEL.

- THERE IS ONE SITUATION WHERE I DON'T EVEN NEED THE CAUSAL GRAPH FOR IDENTIFYING POP-LEVEL COUNTERFACTUALS SUCH AS ATT / CATT (i.e. WITHOUT ASSUMING CONDITIONAL EXOGENEITY)

THIS IS THE CASE IF T IS BINARY AND I HAVE ACCESS TO BOTH OBSERVATIONAL AND EXPERIMENTAL DATA.

$$\text{RDC} \rightarrow \underset{a' \in \ 0,1}{\operatorname{argmax}} \ E[\pi_{a'} | I = a, M]$$

$$\begin{aligned}
P(\pi_{a'}, M) &= P(\pi_{a'}, M, a') + P(\pi_{a'}, M, a) \\
&= P(\pi_{a'}|M, a')P(M, a') + P(\pi_{a'}|M, a)P(M, a)
\end{aligned}$$

$$\begin{aligned}
P(\pi_{a'}|M) &= P(\pi_{a'}|M, a')P(a'|M) + P(\pi_{a'}|M, a)P(a|M) \\
&= P(\pi|M, a')P(a'|M) + P(\pi_{a'}|M, a)P(a|M) \text{ (Consistency)}
\end{aligned}$$

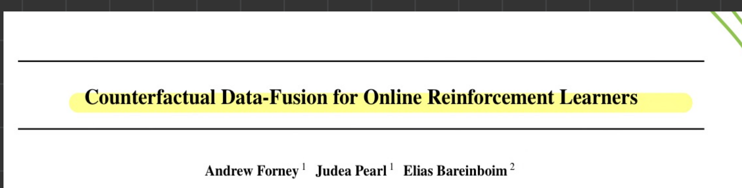
$$\begin{aligned}
P(\pi_{a'}|M, a) &= \frac{1}{P(a|M)}\left[P(\pi_{a'}|M) - P(\pi|M, a')P(a'|M)\right] \\
&= \boxed{\underbrace{\frac{1}{P(a|M)}}_{\text{observational}} \overbrace{\left[P\Big(\pi|M, do(a')\Big)\right.}^{\text{experimental}} - \underbrace{\left.P(\pi|M, a')P(a'|M)\right]}_{\text{observational}}}
\end{aligned}$$

Notice, that in this expression, the conditioning set $\{Z=z\}$ does not satisfy conditional ignorability as I have an unobserved confounder. Still, the CATT is identifiable from experimental + observational data!

- If $|T| > 2$, and I don't want to assume conditional ignorability

I could also conduct RDC-type randomization

ie. avoid requiring the causal graph for ATT/CATT identifiability



Counterfactual Data-Fusion for Online Reinforcement Learners

Andrew Forney [1] Judea Pearl [1] Elias Bareinboim [2]

NOTE ON PRE-BID RANDOMIZATION PAPER (7)

1), and an unbiased estimator of $E(W|Z=1)$ is the fraction of Test userIDs that are exposed. □

Lemma 6.3. In the Pre-Bid Randomization setting, the ATT is the ratio of two observable averages:

$$ATT = ACE/P(U=1) = ACE/w = (R_T - R_C)/w, \tag{10}$$

where w is the win rate, or equivalently,

$$E[Y(1) - Y(0)|U=1] = E[Y(1) - Y(0)]/P(U=1), \tag{11}$$

Proof. We can decompose the ACE (Def. 6.2) as a sum of two terms:

$$\begin{aligned} ACE &= E[Y(1) - Y(0)] = R_T - R_C && (12)\\ &= P(U=1)\cdot(R_{TW} - R'_{CW}) + && (13)\\ &\quad P(U=0)\cdot(R_{TL} - R'_{CL}) && (14)\\ &= w\cdot ATT + (1-w)\cdot\underbrace{(R_{TL} - R'_{CL})}_{\to 0}, && (15) \end{aligned}$$

NOTICE, IN THIS CASE WE ARE IDENTIFYING THE ATT WITHOUT ASSUMING THE CAUSAL GRAPH, BUT NOTE WE NEED ACCESS TO BOTH EXPERIMENTAL (ATE = ACE) AS WELL AS OBSERVATIONAL DATA (W → PROBABILITY OF WIN) FOR IDENTIFYING THE ATT.

NOTE

- IN THE CASE OF THAT INDEX, I CAN ESTIMATE ATT = ATE/W, AND NOT USING PEARL'S CAUSALITY BOOK FORMULA IN PAGE 397 BECAUSE OF THE PRE-BID TYPE RANDOMIZATION.

I COULD ALSO ESTIMATE THE ATT WITH PEARL'S PAGE 397 (WHICH IS EQUIVALENT TO THE ONE DERIVED IN PAGE 5 IN THIS NOTES), BUT I WOULD NEED POST-BID RANDOMIZATION $P(\pi \mid y, do(a'))$, SINCE 'a' HERE REPRESENTS WIN BID / LOSE BID, WHICH IS EQUIVALENT TO BEING EXPOSED TO THE AD.

- IF THE GOAL OF THE STUDY WERE TO ESTIMATE THE CATT (CONDITIONAL AVERAGE TREATMENT EFFECT ON THE TREATED) AND I CAN DO PRE-BID RANDOMIZATION I CAN ESTIMATE IT AS FOLLOWS:

1), and an unbiased estimator of $E(W|Z=1)$ is the fraction of Test userIDs that are exposed.

Lemma 6.3. In the Pre-Bid Randomization setting, the ATT is the ratio of two observable averages:

$$ATT = ACE/P(U=1) = ACE/w = (R_T - R_C)/w, \tag{10}$$

where w is the win rate, or equivalently,

$$E[Y(1) - Y(0)|U=1] = E[Y(1) - Y(0)]/P(U=1), \tag{11}$$

Proof. We can decompose the ACE (Def. 6.2) as a sum of two terms:

$$\begin{aligned} ACE &= E[Y(1) - Y(0)] \;=\; R_T - R_C && (12)\\ &= P(U=1)\cdot(R_{TW} - R'_{CW}) \;+ && (13)\\ &\quad\; P(U=0)\cdot(R_{TL} - R'_{CL}) && (14)\\ &= w \cdot ATT \;+\; (1-w)\cdot(R_{TL} - R'_{CL}), && (15) \end{aligned}$$

$$\begin{aligned} CATE &= E[Y(1) - Y(0) \mid X] \\ &\doteq E[Y \mid do(T), X] \\ &= R_T - R_C \mid X \\ &= P(U=1 \mid X) \cdot (R_{TW} - R'_{CW} \mid X) \;+ \\ &\quad\; P(U=0 \mid X) \cdot (\underbrace{R_{TL} - R'_{CL}}_{0} \mid X) \end{aligned}$$

$$CATE = P(U=1 \mid X) \cdot CATT$$

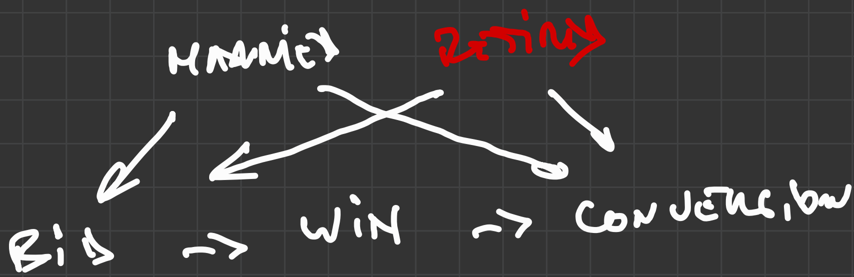
$$CATT = CATE / P(U=1 \mid X)$$

THE ABOVE WOULD WORK FINE IF I DON'T HAVE UCs. IN THE PRESENCE OF UCs, THE CATE IS NOT IDENTIFIED FROM EXPERIMENTAL DATA ALONE. NOTE THAT CATE REQUIRES CONDITIONAL IGNORABILITY

10

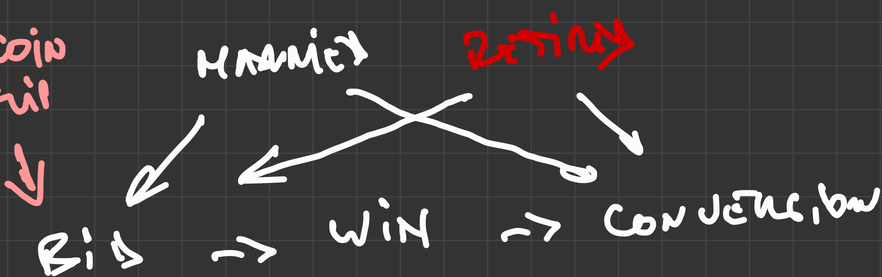
Unconfoundedness Throughout this chapter, whenever we are estimating an ATE, we will assume that W is a sufficient adjustment set, and whenever we are estimating a CATE, we will assume that $W \cup X$ is a sufficient adjustment set. In other words, for ATE estimation, we assume that W satisfies the backdoor criterion (Definition 4.1); equivalently for ATE estimation, we assume that we have conditional exchangeability given W (Assumption 2.2). And similarly for CATE estimation, assuming $W \cup X$ is a sufficient adjustment set means that we are assuming that $W \cup X$ satisfies the backdoor criterion / gives us unconfoundedness. This unconfoundedness assumption gives us parametric identification[2] and allows us to focus on estimation in this chapter.

[2] By "parametric identification," we mean identification under the parametric assumptions of our statistical models. For example, these assumptions are for extrapolation if we don't have positivity.



ORIGINAL DATA

COIN FLIP



POST-BID RANDOMIZED DATA

$$CATT = \frac{E[C \mid do(B=1), M] - E[C \mid do(B=0), M]}{P(W=1 \mid B=1, M)}$$

↓

FOLLOWS FROM PAGE 9

NOTE THAT {M} IS NOT A SUFFICIENT ADJUSTMENT SET (I WOULD ALSO NEED 'RETIRED', BUT IT'S UNOBSERVED)

SO IN THIS CASE (ie. WITH UCs) CAN I IDENTIFY THE CATT?

YES, BUT NOT FROM PRE-BID RANDOMIZATION. I WOULD NEED POST-BID RANDOMIZATION

I COULD USE RAC (PAGE 5) BUT THE TYPE OF EXPERIMENTAL DATA I NEED INVOLVE $P(A \mid do(a'), M)$

WHERE a, a' IS BEING EXPOSED TO THE AD GIVEN WIN 12

THAT IS, MY EXPERIMENT SHOULD INVOLVE RANDOMIZATION EXPOSURE TO AD (BUT AS INDICATED IN THE PRE-BID STAGE. RATHER, THIS WOULD BE IDEAL GIVEN THAT WINNING THE AD BUT NOT EXPOSING A CLIENT TO THE AD IS A WASTE OF MONEY).

KEY CONCLUSION.

SO, EITHER FOR CATE OR CAST IDENTIFIABILITY, EXPERIMENTAL DATA ALONE IS NOT SUFFICIENT WHEN I HAVE UCs.

BOTH AIM TO CAPTURE ITS HETEROGENEOUS TREATMENT EFFECT