Lecture 3:
Causal Models and Counterfactuals
Defining a target causal parameter

A roadmap for causal inference

1. Specify **Causal Model** representing real background knowledge
2. Specify **Causal Question**
3. Specify **Observed Data** and link to causal model
4. **Identify** : Knowledge + data sufficient?
5. Commit to an **estimand** as close to question as possible, and a **statistical model** representing real knowledge.
6. **Estimate**
7. **Interpret** Results

Outline

1. Counterfactuals
   - Link between Structural Causal Models (SCM) and counterfactuals
2. Examples of counterfactual target parameters
   - Average Treatment Effect
   - (Working) marginal structural models

References

- Petersen and van der Laan, Epidemiology 2014
- Targeted Learning, Rose and van der Laan. 2011, Springer, Berlin Heidelberg New York, 2011. Chapter 2
  - Hence forth "TLB"
- Pearl. "An Introduction to Causal Inference" *Int J Biostat*, 6(2): Article 7, 2010.
- Robins and Hernan. "Estimation of the causal effects of time-varying exposures" In: Longitudinal Data Analysis, Chapter 23. Chapman & Hall/CRC Press, Boca Raton, FL, 2009
- Robins, Hernan, Brumback. "Marginal Structural Models and Causal Inference in Epidemiology". *Epidemiology,* 11(5):550-560, 2000.
  - (For Longitudinal Data)
- Neugebauer and van der Laan. "Nonparametric causal effects based on marginal structural models" *Journal of Statistical Planning and Inference*, 137(2): 419-434, 2007.

Defining a target causal parameter

- Recall our motivation:

  experimental conditions under which we observe a system ≠ experimental conditions we are most interested in
- The process of translating our background knowledge into a SCM required us to be specific about our knowledge of existing experimental conditions
- The process of translating our scientific question into a target causal parameter requires us to be specific about our ideal experimental conditions

Defining a target causal parameter

- Step 1. Decide which variable or variables we want to intervene on
  - "Exposure" or "Treatment"
  - We are interested in a system that modifies the way these variables are generated
  - For now focus on one variable at a single time point
    - Lots of times you are interested in intervening on more than one variable
    - We will get to that
  - We refer to this variable as the intervention variable, and typically use "A" to represent it

Defining a target causal parameter

- Step 2. Decide what kind of intervention you are interested in
  - For now, we will focus on "static" interventions
  - Interventions that deterministically set A equal to some fixed value(s) of interest
  - There are other options (e.g. set A in response to other variables)
- Step 3. Specify an outcome (or outcomes)
  - Again, we'll focus on a single outcome at a single time point for now

Example- Vitamins and Breast Cancer

- Interested in the effect of vitamins on breast cancer

1. What is the intervention variable?
2. What is the intervention?
3. What is the outcome?

Example- Vitamins and Breast Cancer

- In the "experiment" that generated our data, only a self-selected group took vitamins
- What modifications to these experimental conditions might we be interested in…?
  - What would the risk of developing breast cancer have been if we had instituted a policy in which the whole population took vitamins?
  - Need a way to formally express this target parameter
    - $P(Y=1|A=1)=P(Y=1)$? Not good…
    - Need to make clear we are not talking about the observed data distribution

Counterfactuals

- Y_a for an individual is understood as the value that variable Y would have taken for that individual if that individual had received treatment A=a
  - "Counterfactual" because the individual may not have actually received treatment A=a
  - Also referred to as "Potential Outcomes"
    - Neyman *Statistical Science*, 5:472 – 480, 1923
    - Rubin A*nnals of Statistics*, 6:34 – 58, 1978
    - Robins *Mathematical Modelling*, 7:1393 – 1512, 1986

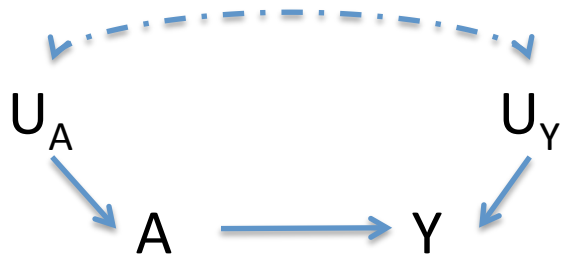
In SCM counterfactuals are derived from the causal model…

- Structural equations are autonomous
  - Changing one function does not change the other functions
  - This means we can intervene on part of the system and see how changes are transmitted through the rest of the system
- If we want to make inferences about data generated by the same system under different conditions, we have to know which parts of the system will change and which parts will stay the same

Example: Autonomy of Structural Equations

- *X=(A=Vitamins, Y=Breast cancer); U=(U_A,U_Y)~P_U*

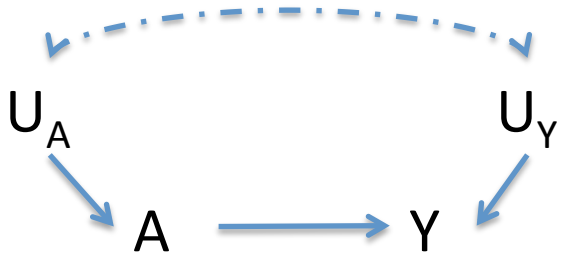
$$A = f_A(U_A)$$
$$Y = f_Y(A, U_Y)$$



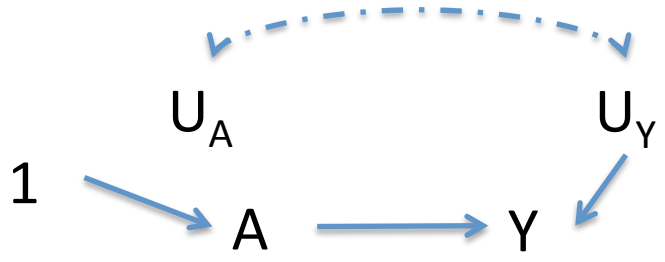
  - Changing how the decision to use vitamins is made does not change the effect of vitamin use on breast cancer
  - Ex. Assigning everyone to vitamins versus letting women make their own decision about using vitamins
- Autonomy reasonable here?

Interventions on the SCM

- The autonomy of structural equations means that we can make a targeted modification to the set of equations in order to represent our intervention of interest
- Ex. Intervene on the system to set A=1
  - Replace f_A with constant function *A=1*



$$A = f_A(U_A)$$
$$Y = f_Y(A, U_Y)$$



$$A = 1$$
$$Y = f_Y(A, U_Y)$$

Counterfactuals defined using SCM

- $Y_a(u)$ is defined as the solution to the equation f_Y under an intervention on the system of equations to set $A=a$ (with input $U=u$)
  - Again we can think of u as the background factors of each subject
  - $Y_a(u)$ is a realization
    - It is implied by F and u
- P_U and F induce a probability distribution on Y_a just as they do on Y
  - $Y_a(U)$ is the post-intervention (or counterfactual) random variable
    - All the randomness comes from U

Example: Counterfactuals derived from SCM

- Endogenous variables: X={W, A, Y}
  - W=SES
  - A=vitamin use
  - Y= breast cancer
- Errors: $U=(U_W,U_A,U_Y)\sim P_U$
- Structural equations:

$$
\begin{aligned}
W &= f_W(U_W) \\
A &= f_A(W, U_A) \\
Y &= f_Y(W, A, U_Y)
\end{aligned}
$$

- <u>Distribution of (U,X) generated by:</u>

1. Draw *U* from some distribution
2. Generate *W* as a deterministic function of U_W
3. Generate *A* as a deterministic function of *W* and U_A
4. Generate *Y* as a deterministic function of *W*, *A*, U_Y

Example: Counterfactuals derived from SCM

- Endogenous variables:
  X={W, A, Y}
  - W=SES
  - A=vitamin use
  - Y= breast cancer
- Errors: $U=(U_W,U_A,U_Y)\sim P_U$
- Structural equations:

$$W = f_W(U_W)$$

$$A = 1$$

$$Y = f_Y(W, A, U_Y)$$

- <u>Distribution of (U,X) generated by:</u>

1. Draw *U* from some distribution
2. Generate *W* as a deterministic function of U_W
3. Set A deterministically=1
4. Generate $Y_1(U)$ as a deterministic function of *W*, *A*, U_Y

Alternative Framework: Casual Inference as a missing data problem

- Define the "Full Data"
  - The data you would have had in your ideal (impossible) experiment

$$\text{Full Data: } X^F = (L_a : a \in \mathcal{A}) \sim \mathcal{F}_{\mathcal{X}},$$
$$\text{where } \mathcal{A} \text{ is the set of possible treatment levels.}$$

- Example

$$\begin{aligned} W &= f_W(U_W) \\ A &= f_A(W, U_A) \\ Y &= f_Y(W, A, U_Y) \end{aligned}$$

$$\text{Full Data: } X^F = (W, (Y_a : a \in \{0, 1\})) \sim F_X$$

  - Alternative notation: won't use in this class

Counterfactuals in the missing data framework

- Counterfactuals are assumed to exist for each individual and for each treatment value of interest
  - They are typically assumed to be deterministic
  - They have a distribution in the population: $Y_a \sim P_{Ya}$
- Example: A=Vitamin Use
  - For a given individual: Y_1 is the breast cancer status that that individual would have had, had she taken vitamins
  - $P(Y_1=1)$ is the risk of breast cancer in the population if everyone had taken vitamins

Where do these counterfactuals (and their distribution) come from?

- Could treat them as primary quantities
  - We assume their existence
  - We assume a link between these counterfactuals and the observed data
    - The "Consistency assumption": $Y_A=Y$
  - This lets us treat causal inference as a missing data problem
- Additional references to this framework
  - Morgan and Winship (2007). Counterfactuals and Causal Inference. Methods and Principles for the Social Sciences
  - Hernan and Robins (textbook in process). Causal Inference.

The missing data paradigm

- Example – Vitamins and Breast cancer
  - Ideally all women would have taken vitamins and we could have observed their breast cancer outcomes
  - The outcomes we are interested in (Y_1) are observable for the women who did take vitamins
  - The outcomes we are interested in (Y_1) are missing for the women who did not take vitamins

SCM and the missing data framework

- A single SCM is a model on lots and lots of possible Full Data sets
  - Corresponding to all the possible interventions we might make on components of X…
  - The SCM defines the allowed distributions for (is a model on) each of these sets of Full Data
- Full Data and missing data frameworks can be a useful concept when developing estimation approaches…

Where are we?

- We have learned how to represent a modification to the experimental conditions as an intervention on the SCM/causal graph
- We have used this concept to define counterfactuals as derived quantities
- We have seen how the original SCM is also a model on the distribution of these counterfactuals

Defining a target causal parameter

1. Decide which variable we want to "intervene on"
2. Decide what intervention we are interested in

Steps 1 and 2 define our counterfactuals of interest (and our SCM defines a model for the distribution of these counterfactuals)

3. Specify what parameter of the distribution of these counterfactuals we are interested in…

Ex: Defining target causal parameters

- Endogenous variables:

  X={W, A, Y}
  - W=SES
  - A=vitamin use
  - Y= breast cancer
- Structural equations

$$\begin{aligned} W &= f_W(U_W) \\ A &= f_A(W, U_A) \\ Y &= f_Y(W, A, U_Y) \end{aligned}$$

- Errors: $U=(U_W,U_A,U_Y)\sim P_U$
- Interventions of interest: Set A=1 and A=0
- Counterfactuals of Interest:

$$Y_a = f_Y(W, a, U_Y),\ a \in \mathcal{A} = \{0, 1\}$$

where $\mathcal{A}$ refers to treatment levels of interest

Example: Average Treatment Effect:

- How would expected outcome have differed if everyone in the population had been treated vs. if no one in the population had been treated?
  - This is a common target of inference.
  - This is what most RCTs are trying to estimate….

$$E_{U,X}Y_1 - E_{U,X}Y_0$$

Distribution of Y_a is given by P_U and F, or alternatively, by $P_{U,X}$

Other counterfactual parameters

- Causal Relative Risk $E_{U,X}Y_1 / E_{U,X}Y_0$
- Causal Odds Ratio $$\frac{E_{U,X}Y_1/(1-E_{U,X}Y_1)}{E_{U,X}Y_0/(1-E_{U,X}Y_0)}$$
- May be interested in a causal effect within certain strata of the population…

$$\begin{aligned} W &= f_W(U_W) \\ A &= f_A(W, U_A) \\ Y &= f_Y(W, A, U_Y) \end{aligned}$$

$$E_{U,X}(Y_1 - Y_0|V), V \subset W$$

Marginal Structural Models

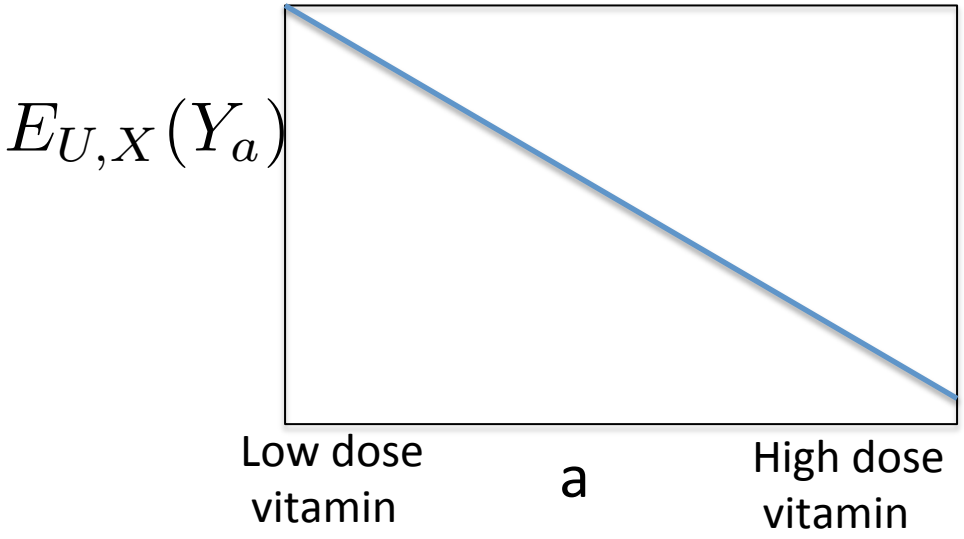
- Just another way to define your target parameter…
- Provides a summary measure of how the counterfactual outcome changes as a function of treatment (and possibly pre-treatment covariates)
- Useful when A (or (A,V)) has many possible levels…

Example: Marginal Structural Models

- Question: Effect of vitamin dose on cholesterol levels?
- Intervention variable is now continuous (or ordinal with a lot of levels)
- The ideal experiment might intervene to set vitamin dose to each possible value and generate corresponding cholesterol outcomes

Example: Marginal Structural Models

- If we knew the shape of this dose response curve we could summarize it using a model…
  - Robins: MSM: A model on the marginal distribution of the counterfactual outcomes

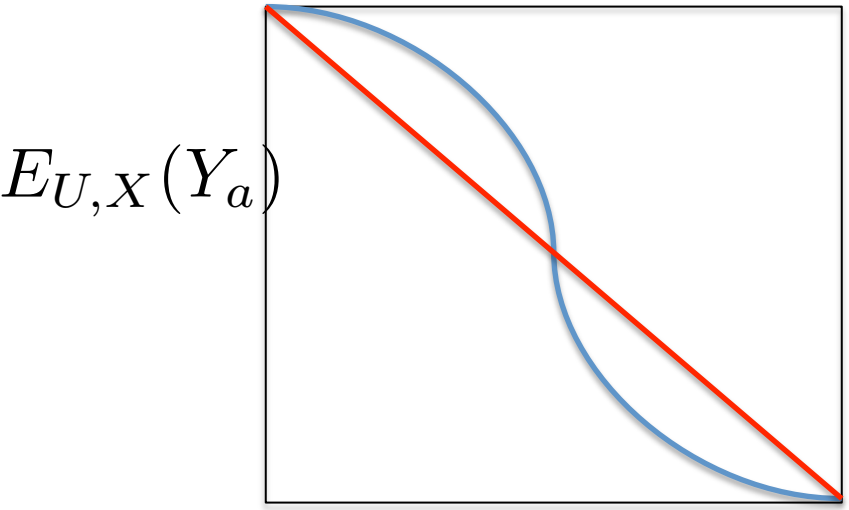


$$E_{U,X}(Y_a) = m(a|\beta)$$

$$m(a|\beta) = \beta_0 + \beta_1 a$$

Example: Marginal Structural Working Models

- We usually don't know enough to confidently specify a parametric model for this dose response curve
- Nonetheless, we may be willing to settle for some summary measure of the true dose-response curve….



Define target parameter as a projection of the true causal curve onto a working model

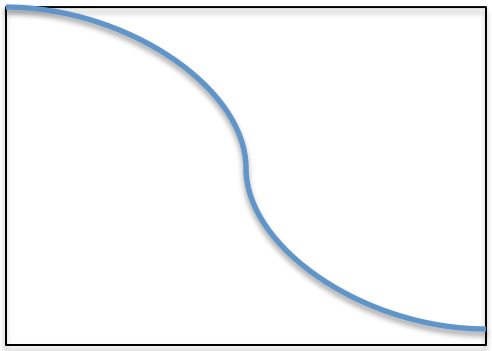
$$m(a|\beta) = \beta_0 + \beta_1 a$$

$$\beta(P_{U,X}|m) \equiv \arg\min_{\beta} E_{U,X}\left[\sum_{a\in\mathcal{A}}(Y_a - m(a|\beta))^2\right]$$

Marginal Structural Working Models

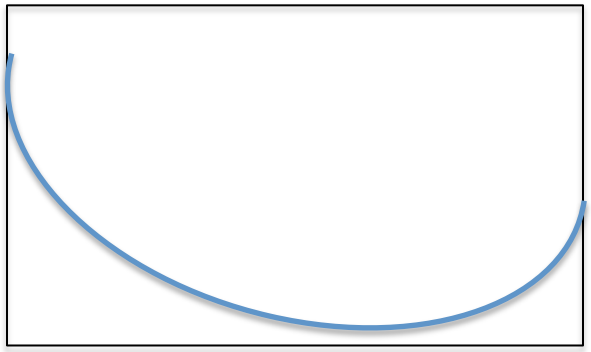
- How to think of working MSMs:
  - Take your full data (ideal experiment) and then use it to fit a regression model as a means to summarize the causal relationship of interest
  - How interesting a given summary (working model) is depends on the underlying causal curve and your question

$E_{U,X}(Y_a)$



a

$E_{U,X}(Y_a)$



a

Marginal Structural Working Models

- Defining our target parameter by specifying a working marginal structural model (rather than a parametric MSM) is more in-line with general philosophy in this class
  - Don't introduce new assumptions about the data generating system at this stage
  - See Neugebauer Non-parametric MSMs reading

Key Points

- Counterfactuals provide a language for translating a scientific question into a formal causal query
- Counterfactuals can be defined based on interventions on a structural causal model
- Target causal parameters defined by contrasting the distribution of the counterfactual outcome under different interventions
- Examples
  - Average Treatment Effect
  - Using a (working) marginal structural model