Lecture 10:

Causal Inference

Causal Inference

Roughly, the statement “X causes Y” means that changing the value of X will change the distribution of Y.

When X causes Y, X and Y will be associated (not independent), but the reverse is generally not true – association does not imply causation.

There are at least two frameworks for discussing causation – the one using **counterfactual** random variables, and the one using **directed acyclic graphs** (aka graphical models).

We will mostly discuss the counterfactual model today.

The Counterfactual Model

The Counterfactual Model

Suppose X is a binary treatment variable: $X = 1$ means “treated” and $X = 0$ means “not treated”. (Treatment in a very broad sense, alternatively “exposed / not exposed”).

Let Y be some outcome variable such as presence or absence of disease. To distinguish association from causation, we’ll decompose the response, Y, in the following way:

Introduce two new r.v.-s: (C_0, C_1) called **potential outcomes**:

$$Y = \begin{cases} C_0 & \text{if} \quad X = 0 \\ C_1 & \text{if} \quad X = 1 \end{cases}, \quad \text{so, shorter, } Y = C_X$$

This is called the **consistency relationship**

The Counterfactual Model

Here's a toy dataset to make it clear:
The asterisks (*) are for unobserved values.

X	Y	C_0	C_1
0	4	4	*
0	7	7	*
0	2	2	*
0	8	8	*
1	3	*	3
1	5	*	5
1	8	*	8
1	9	*	9

When $X = 0$, we don't observe C_1, so we say that C_1 is a **counterfactual** (counter the fact). Similarly, if $X = 1$, C_0 is counterfactual.

There are four types of subjects:

Type	C_0	C_1
Survivors	1	1
Responders	0	1
Anti-responders	1	0
Doomed	0	0

Think of the potential outcomes (C_0, C_1) as of hidden variables that contain all the relevant information about the subject.

The Counterfactual Model

Define the **average causal effect** or **average treatment effect** as:

$$\theta = \mathbb{E}(C_1) - \mathbb{E}(C_0)$$

The parameter θ has the following interpretation: θ is the mean if everyone were treated ($X = 1$) minus the mean if everyone were not ($X = 0$). But there are other ways of measuring the causal effect, for example, if C_0 and C_1 are binary, define **the causal odds ratio** and the **causal relative risk:**

$$\frac{\mathbb{P}(C_1 = 1)}{\mathbb{P}(C_1 = 0)} \div \frac{\mathbb{P}(C_0 = 1)}{\mathbb{P}(C_0 = 0)} \qquad\qquad \frac{\mathbb{P}(C_1 = 1)}{\mathbb{P}(C_0 = 1)}$$

The Counterfactual Model

For simplicity, we'll work with the average causal effect θ. (But the main ideas are the same whatever causal effect we use).

Define the **association** as $\alpha = \mathbb{E}(Y|X=1) - \mathbb{E}(Y|X=0)$

Theorem: (Association is not causation). In general, $\theta \neq \alpha$.

X	Y	C_0	C_1
0	0	0	0*
0	0	0	0*
0	0	0	0*
0	0	0	0*
1	1	1*	1
1	1	1*	1
1	1	1*	1
1	1	1*	1

<- **Example**. Notice that $C_0 = C_1$, so this treatment has no effect. Indeed, av. causal effect:

$$\theta = \mathbb{E}(C_1) - \mathbb{E}(C_0) = \frac{4 \cdot 0 + 4 \cdot 1}{8} - \frac{4 \cdot 0 + 4 \cdot 1}{8} = 0$$

the association: $\alpha = \dfrac{4 \cdot 1}{4} - \dfrac{0 \cdot 1}{4} = 1$

The Counterfactual Model

Suppose the outcome var. is 1 for “healthy” and 0 for “sick”. $X = 0$ means subject doesn’t take vitamin C, $X = 1$ means he/she does. In our example healthy people $(C_0, C_1) = (1,1)$ tend to take vitamin C, and unhealthy $(C_0, C_1) = (0,0)$ don’t. Association between (C_0, C_1) and X creates an association between X and Y.

X	Y	C_0	C_1
0	0	0	0*
1	0	0	0*
1	0	0	0*
1	0	0	0*
1	1	1*	1
1	1	1*	1
1	1	1*	1
1	1	1*	1

Suppose we wrongly interpret this causally, and conclude that vitamin C prevents illness. We might <- encourage everyone to take vitamin C.

Now $\alpha = 4/7 - 0/1 = 4/7$ went down from 1. The causal effect never changed, but not distinguishing association and causation confuses – it seems we made things worse instead of better.

The Counterfactual Model

In that example, $\theta = 0$ and $\alpha = 1$. It is not hard to create examples in which $\alpha > 0$ and yet $\theta < 0$ – association and causal effects can have different signs.

So we cannot use the association to estimate the causal effect θ. $\theta \neq \alpha$, since (C_0, C_1) is not independent of X – treatment assignment was not independent of person type!

Can we estimate the causal effect? Sometimes yes! In particular, random assignment to treatment makes it possible. *(See next page)*

The Counterfactual Model

Theorem: Suppose we randomly assign subjects to treatment and that $\mathbb{P}(X=0)>0$ and $\mathbb{P}(X=1)>0$. Then $\alpha=\theta$. Hence, any consistent estimator of α is a consistent estimator of θ. In particular,

$$\widehat{\theta} = \hat{\mathbb{E}}(Y|X=1) - \hat{\mathbb{E}}(Y|X=0) = \overline{Y}_1 - \overline{Y}_0$$

is a consistent estimator of θ, where $\overline{Y}_1 = \dfrac{1}{n_1}\displaystyle\sum_{i=1}^{n} Y_i X_i$,

$$\overline{Y}_0 = \frac{1}{n_0}\sum_{i=1}^{n} Y_i(1-X_i), \quad n_1 = \sum_{i=1}^{n} X_i \text{ and } n_0 = \sum_{i=1}^{n}(1-X_i)$$

The Counterfactual Model

If Z is a covariate, define the **conditional causal effect** by

$$\theta_z = \mathbb{E}(C_1 | Z = z) - \mathbb{E}(C_0 | Z = z)$$

For example, if Z denotes gender $Z = 0$ (women) and $Z = 1$ (men), then θ_0 is the causal effect among men.

In a randomized experiment,

$$\theta_z = \mathbb{E}(Y | X = 1, Z = z) - \mathbb{E}(Y | X = 0, Z = z)$$

and we can estimate the conditional effect using appropriate sample averages.

The Counterfactual Model

Summary of the Counterfactual Model:

Random variables: (C_0, C_1, X, Y)

Consistency relationship: $Y = C_X$

Causal Effect: $\theta = \mathbb{E}(C_1) - \mathbb{E}(C_0)$

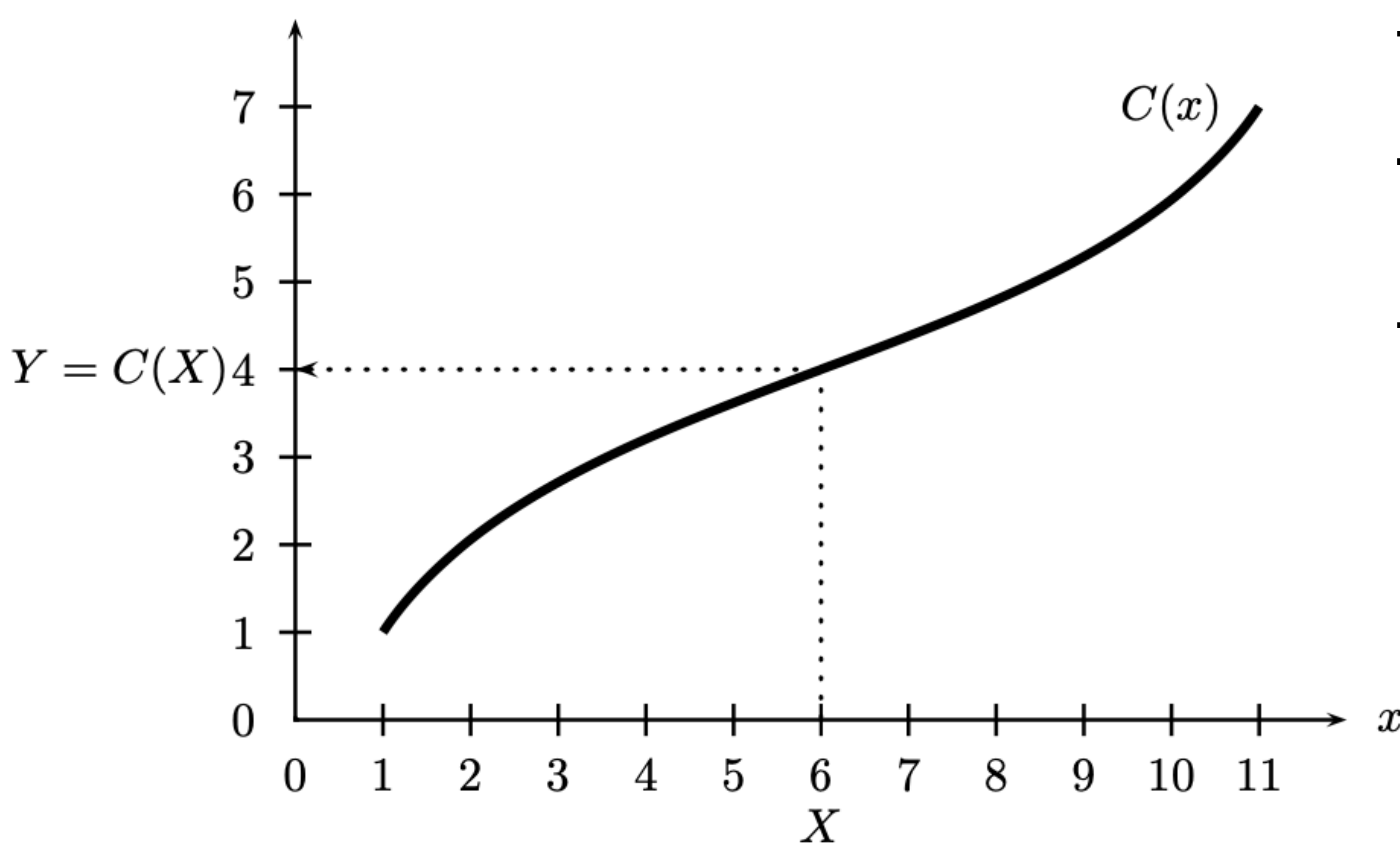
Association: $\alpha = \mathbb{E}(Y | X = 1) - \mathbb{E}(Y | X = 0)$

Random assignment -> $(C_0, C_1) \perp X$ -> $\theta = \alpha$

Beyond Binary Treatments

Beyond Binary Treatments

Suppose $X \in \mathscr{X}$. For example, X could be the dose of a drug, then $X \in \mathbb{R}$. The counterfactual vector (C_0, C_1) becomes a **counterfactual function** $C(x)$ – the outcome a subject would have if received dose x. Observed response is given by $Y \equiv C(x)$ – consistency relation



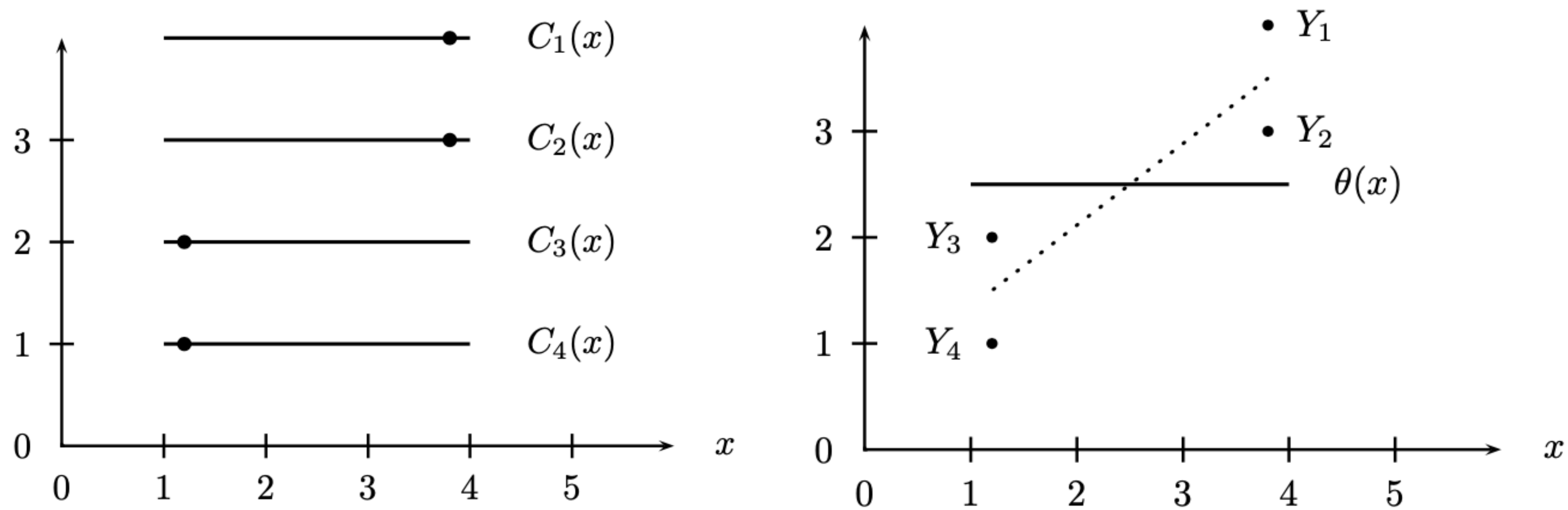
The **causal regression function** is $\theta(x) = \mathbb{E}(C(x))$.

The regression function, measuring association, is $r(x) = \mathbb{E}(Y | X = x)$

Beyond Binary Treatments

Theorem: In general, $\theta(x) \neq r(x)$. However, when X is assigned randomly, $\theta(x) = r(x)$.

Example: Here, $\theta(x)$ is constant but $r(x)$ is not:



$C_i(x)$ are constant over x for all i, so there is no causal effect: the causal regression function $\theta(x) = (C_1(x) + \ldots + C_4(x))/4$.

$r(x) = \mathbb{E}(Y | X = x)$ is not constant – there's association.

Observational Studies and Confounding

Observational Studies and Confounding

A study in which treatment is not randomly assigned is called an **observational study** – subjects have individual exposure values X.

The discrepancy occurs here because the potential outcome C is not independent of treatment X. But, suppose, we can split subjects into groups so that within a group X and $\{C(x) : x \in \mathcal{X}\}$ are indep. That happens if subjects are very similar within groups. For example, suppose we find people of similar age, gender, ethnicity. Among them we can assume the choice of X is random. These other variables are called **confounding variables** – denote them all as Z. Then the above idea (**no unmeasured confoundung**) is that:

$$\{C(x) : x \in \mathcal{X}\} \perp X | Z$$

Observational Studies and Confounding

Theorem: Suppose $\{C(x) : x \in \mathscr{X}\} \perp X \mid Z$. Then,

$$\theta(x) = \int \mathbb{E}(Y \mid X = x, Z = z)\, dF_Z(z)\, dz$$

if $\widehat{r}(x, z)$ is a consistent estimate of the regression function $\mathbb{E}(Y \mid X = x, Z = z)$, then a consistent estimate of $\theta(x)$ is

$$\widehat{\theta}(x) = \frac{1}{n} \sum_{i=1}^{n} \widehat{r}(x, Z_i)$$

in particular, if $r(x, z) = \beta_0 + \beta_1 x + \beta_2 z$ – linear, then

$\widehat{\theta}(x) = \widehat{\beta}_0 + \widehat{\beta}_1 x + \widehat{\beta}_2 \overline{Z}_n$ where $(\widehat{\beta}_0, \widehat{\beta}_1, \widehat{\beta}_2)$ are the least squares estimators

Observational Studies and Confounding

Remark: Compare $\theta(x) = \displaystyle\int \mathbb{E}(Y \mid X = x, Z = z)\, dF_Z(z)\, dz$ to

$$\mathbb{E}(Y \mid X = x) = \int \mathbb{E}(Y \mid X = x, Z = z)\, dF_{Z|X}(z \mid x)$$

Epidemiologists call 1-st one the **adjusted treatment effect**. The process of computing it is called **adjusting (or controlling) for confounding**. Selection of Z (confounders) requires scientific insight. Even after adjusting for them, we can't be sure there are no others. So take such studies with healthy skepticism. These become believable when 1) results are replicated in many studies, 2) each of studies controlled for plausible confounding variables, 3) there is a plausible scientific explanation for a causal relationship.