DS 102 Data, Inference, and Decisions **Spring 2020**

Lecture 14: Causal Inference II

Lecturer: Moritz Hardt

14.1 Recap

In the previous lecture, introduced the idea of causality. We saw that causal effects cannot be expressed in the language of probability, and built up a framework for making formal statements about causality.

Structural causal models, which can be thought of as programs for generating data, specify instructions for building a data-generating distribution step-by-step from independent noise variables. More formally, they consist of a list of assignments to generate a distribution on $(X_1, \ldots, X_d)$ from independent noise variables $(N_1, \ldots, N_d)$

$$
\begin{aligned}
&1.\ X_1 := f_1(V_1, N_1) \\
&2.\ X_2 := f_2(V_2, N_2) \\
&\qquad \cdots \\
&3.\ X_d := f_d(V_d, N_d)
\end{aligned}
$$

where $V_i \subseteq \{X_1, \ldots, X_d\}$ is the set of parents of X_i.

We defined do-interventions (or do-assignments) as a formal way to talk about taking actions in a given causal model. Specifically, $\mathbf{do}(X_i := x)$ means we replace the i^{th} assignment in the structural causal model with $X_i := x$.

Once we have defined do-intervention, we can formalize the causal effect of X on Y as $\mathbb{P}(Y = y \mid \mathbf{do}(X := x))$. When X is a binary treatment variable (for example, whether or not an individual takes a medication), we can define a related quantity called the **treatment effect**,

$$\mathbb{E}[Y \mid \mathbf{do}(X := 1)] - \mathbb{E}[Y \mid \mathbf{do}(X := 0)],$$

which measures the average effect of the treatment X on the outcome Y relative to the baseline.

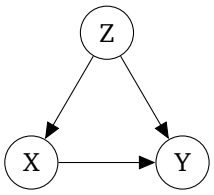

Figure 14.1: Causal graph where Z is a confounding variable.

We also introduced the idea of a causal graph, which represents the dependence structure of a given causal model. In the causal graph, there is an incoming edge into X_i from each parent in V_i. We looked at some specific examples of causal graphs such as Figure 14.1, wherein the variable Z is a confounder. The confounding influence of Z means that $\mathbb{P}(Y = y \mid \mathbf{do}(X := x)) \neq \mathbb{P}(Y = y \mid X = x)$. We also examined structures like the one in Figure 14.2 where Z is a mediator. Mediators are fundamentally different from confounders; in particular, in this case $\mathbb{P}(Y = y \mid \mathbf{do}(X := x)) = \mathbb{P}(Y = y \mid X = x)$ since Z is a part of the causal effect. In general, when we make a do-intervention on a source node (that is, a node that has no parents), it is the same thing as conditioning.

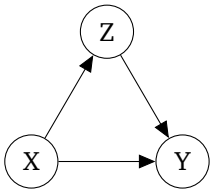

Figure 14.2: Causal graph where Z is a mediating variable.

In addition to the structures associated with confounding and mediating random variables, we can consider structures like that in Figure 14.3. In this case, Z is called a **collider** since it has an incoming effect from both X and Y. Unlike the case where conditioning on a confounder helps us understand the true causal effect, conditioning on a collider can create anti-correlation between X and Y even when (without conditioning) they are truly uncorrelated in the population. This phenomenon is known as **Berkson's law** or **collider bias**.

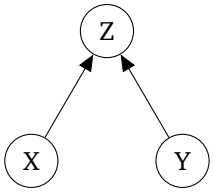

Figure 14.3: Causal graph where Z is a collider.

Example 14.1. Suppose we are studying individuals in a hospital. We use $Z \in \{0, 1\}$ to denote whether or not an individual is admitted to the hospital. Getting admitted has multiple sufficient causes. Let $X \in \{0, 1\}$ represent whether or not an individual has a broken leg. Similarly, let $Y \in \{0, 1\}$ indicate whether an individual has pneumonia. In the general population, broken legs and pneumonia are uncorrelated. However, if we are in the hospital and observe that an individual has a broken leg, what does that tell us about pneumonia? In this case, since have a broken leg is by itself a sufficient condition for being admitted to the hospital, it explains away the other possible causes. If we condition on hospital admission (Z), we might incorrectly conclude that X and Y are anti-correlated.

With this basic framework in place, today we will discuss approaches for actually estimating causal effects.

14.2 Adjustment Formula

The fundamental question we would like to address is the following:
when/how can we estimate causal effects from data?

It turns out that we can rewrite this question in a more operational way. Since we can always estimate conditional probabilities from data, we can ask the following equivalent question:
when/how can we express a do-intervention with formula that involves only conditional probabilities?

Recall that, in general, $P(Y = y \mid \mathbf{do}(X := x)) \neq P(Y = y \mid X = x)$ due to confounding. Consider the causal structure in Figure 14.1, where Z is a discrete random variable that takes one of k possible values. In this case, the following **adjustment formula** precisely specifies the causal effect of X on Y in terms of conditional probabilities.

Claim 14.2 (Adjustment Formula, discrete case)**.**

$$\mathbb{P}(Y = y \mid \boldsymbol{do}(X := x) = \sum_z \mathbb{P}(Y = y \mid X = x, Z = z)\mathbb{P}(Z = z).$$

Let us interpret the righthand side of this adjustment formula. Note that each conditional probability $\mathbb{P}(Y = y \mid X = x, Z = z)$ corresponds to performing a "separate analysis" for each value of Z. Then, to get the overall causal effect, we combine the results of all of these separate analyses weighted by how often $Z = z$ occurs. Also note that, despite superficial similarities, this adjustment formula is *not* the same as the law of total probability.

Proof of Adjustment Formula. First, note that

$$\mathbb{P}(Y = y \mid \mathbf{do}(X := x),\ Z = z) = \mathbb{P}(Y = y \mid X = x,\ Z = z)$$

since fixing the value of Z blocks the confounding influence of Z in the causal graph (Figure 14.1). Then, by applying the law of total probability to the model where we make the do-intervention $\mathbf{do}(X := x)$,

$$\begin{aligned}\mathbb{P}(Y = y \mid \mathbf{do}(X := x)) &= \sum_z P(Y = y \mid \mathbf{do}(X := x),\ Z = z)\mathbb{P}(Z = z) \\ &= \sum_z P(Y = y \mid X = x,\ Z = z)P(Z = z).\end{aligned}$$

□

We have shown that the adjustment formula holds for the simple causal graph Figure 14.1. What can we say when we have an arbitrary causal graph? If X has a causal effect on Y, there are directed paths from $X \to Y$ in the causal graph which all add up to a real causal effect. All the pathways that have incoming edges to X are the "bad pathways" which do not have only forward edges from X to Y. These bad paths are also called **backdoor paths** because they start with an incoming edge into X. Since all confounding pathways between X and Y are backdoor paths, if we condition on all of the parents of X, we block all the confounding pathways. Thus, the adjustment

formula generalizes directly to arbitrary graphs if we replace Z with all the parents of X. This highlights one of the core ideas in causal inference: we find a set of variables in our causal model that blocks the confounding pathway(s) and apply the adjustment formula to estimate the causal effect. Every causal inference estimator uses this idea under the hood.

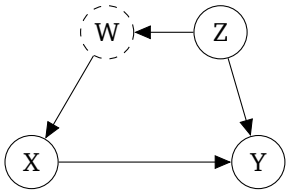

Figure 14.4: Causal graph where W is an unobserved confounding variable and Z is a fork.

So far, we have assumed that we observed all of the nodes in our causal graph. This may not always be the case; there might be some variable that cannot be measured in the data that is still a confounder. For example, when we are trying to decide whether smoking causes cancer, attitude towards personal safety (i.e. whether an individual is risk neutral or risk averse) is a confounder which is typically not observed. This case corresponds to a causal graph like Figure 14.4. We can block the backdoor path $X \leftarrow W \leftarrow Z \rightarrow Y$ by adjusting for Z alone. Note that if Z were a collider we would have $X \leftarrow W \rightarrow Z \leftarrow Y$ and conditioning on Z would no longer work to eliminate the confounding path since Berkson's law will come into play. There is something called the **backdoor criterion** which defines a general method for determining which variables we can adjust for using the adjustment formula (i.e. the set of variables which blocks all the backdoor pathways) when we do not observe all the parents of X.

14.3 Propensity Scores

In the previous section, we saw that the adjustment formula gives us a way to estimate causal effects in an arbitrary causal graph by summing over all possible values of the variables which block our confounding pathways. Here, we discuss an alternative approach which works when Z is continuous or its support is too large to sum over.

Let $X \in \{0, 1\}$ be a binary treatment variable. The quantity $e(z) = \mathbb{E}[X \mid Z = z]$ is known as the **propensity score** and gives the likelihood of treatment given $Z = z$. We claim that the following relationship holds:

Claim 14.3. *Suppose that the adjustment formula holds for Z, and $e(z) \neq 0$ for all z. Then,*

$$\mathbb{E}[Y \mid \textbf{\textit{do}}(X := 1)] = \mathbb{E}\left[\frac{YX}{e(Z)}\right] \tag{$\star$}$$

Note that this equation is useful because it exposes a straightforward way to estimate the expected outcome Y when we make the intervention $\textbf{do}(X := x)$. Specifically, we can first learn a model

$\hat{e}(z) \approx e(z)$ from data, for example by using logistic regression. We then estimate $(\star)$ from samples (x_i, y_i, z_i) as:

$$\frac{1}{n}\sum_{i=1}^{n}\frac{x_i y_i}{\hat{e}(z_i)}.$$

This approach is called **inverse propensity score weighting**.

Proof of $(\star)$. To begin,

$$\begin{aligned}
\mathbb{E}[Y \mid \mathbf{do}(X := 1)] &= \sum_y y\mathbb{P}(Y = y \mid \mathbf{do}(X := 1)) \\
&= \sum_y y\left(\sum_z \mathbb{P}(Y = y \mid X = 1, Z = z)\mathbb{P}(Z = z)\right) \\
&= \sum_y y\left(\sum_z \frac{\mathbb{P}(Y = y \mid X = 1, Z = z)\mathbb{P}(Z = z)\mathbb{P}(X = 1 \mid Z = z)}{\mathbb{P}(X = 1 \mid Z = z)}\right) \\
&= \sum_y y\left(\sum_z \frac{\mathbb{P}(Y = y,\ X = 1,\ Z = z)}{\mathbb{P}(X = 1 \mid Z = z)}\right) \\
&= \sum_y y\left(\sum_{z,x\in\{0,1\}} \frac{\mathbb{1}\{X = 1\}\mathbb{P}(Y = y,\ X = x,\ Z = z)}{\mathbb{P}(X = 1 \mid Z = z)}\right) \\
&= \sum_{y,z,x\in\{0,1\}} \frac{y\mathbb{1}\{X = 1\}\mathbb{P}(Y = y,\ X = x,\ Z = z)}{\mathbb{P}(X = 1 \mid Z = z)} \\
&= \mathbb{E}\left[\frac{YX}{e(Z)}\right]
\end{aligned}$$

where we used the adjustment formula in the second line, then multiplied and divided by $e(z) = \mathbb{P}(X = 1 \mid Z = z) \neq 0$ to obtain the third line. □

Note that the same argument can we used to show that $\mathbb{E}[Y \mid \mathbf{do}(X := 0)] = \mathbb{E}\left[\frac{Y(1-X)}{1-e(Z)}\right]$