Causal Inference in Statistics: A Primer

Derek Li

1 Graphical Models and Their Applications

1.1 d-Separation

Definition 1.1 (d-Separation)**.** A path p is blocked by a set of nodes Z iff

1. p contains a chain of nodes $A \to B \to C$ or a fork $A \leftarrow B \to C$ s.t. the middle node B is in Z (i.e., B is conditioned on), or
2. p contains a collider $A \to B \leftarrow C$ s.t. the collision node B is not in Z, and no descendant of B is in Z.

If Z blocks every path between two nodes X and Y, then X and Y are d-separated, conditional on Z, and thus are independent conditional on Z.

Note. It is similar to Bayes Ball Rules.

1.2 Model Testing and Causal Search

If we have a graph G that we believe might have generated a data set S, d-separation will tell us which variables in G must be independent conditional on which other variables. We can test conditional independence using a data set.

Example 1.1. Suppose we find W and Z_1 are independent given X since X d-separates W from Z_1. We regress W on X and Z_1, i.e., we find the line $w = r_X x + r_1 z_1$ that best fits our data. If $r_1 \neq 0$, then W depends on Z_1 given X and the model is wrong.

2 The Effects of Interventions

2.1 Interventions

- $P(Y = y|X = x)$ is the probability that $Y = y$ conditional on finding $X = x$. $P(Y = y|do(X = x))$ is the probability that $Y = y$ when we intervene to make $X = x$.
- $P(Y = y|X = x)$ is the population distribution of Y among individuals whose X value is x. $P(Y = y|do(X = x))$ is the population distribution of Y if *everyone in the population* had their X value fixed at x.

2.2 The Adjustment Formula

Rule 2.1 (The Causal Effect Rule)**.** Given a graph G in which a set of variables PA are designated as the parents of X, the causal effect of X on Y is given by

$$\begin{aligned} P(Y = y|do(X = x)) &= \sum_z P(Y = y|X = x, PA = z)P(PA = z) \\ &= \sum_z \frac{P(X = x, Y = y, PA = z)}{P(X = x|PA = z)} \end{aligned}$$

where z ranges over all the combinations of values that the variables in PA can take.

The truncated product formula is

$$P(x_1, \cdots, x_n|do(x)) = \prod_i P(x_i|pa_i), \forall i \text{ with } X_i \notin X$$

2.3 The Backdoor Criterion

Definition 2.1 (The Backdoor Criterion)**.** Given an ordered pair of variables (X, Y) in a directed acyclic graph G, a set of variables Z satisfies the backdoor criterion relative to (X, Y) if no node in Z is a descendant of X, and Z blocks every path between X and Y that contains an arrow into X.

If a set of variables Z satisfies the backdoor criterion for X and Y, then the causal effect of X on Y is

$$P(Y = y|do(X = x)) = \sum_z P(Y = y|X = x, Z = z)P(Z = z)$$

2.4 The Front-Door Criterion

Definition 2.2 (Front-Door)**.** A set of variables Z is said to satisfy the front-door criterion relative to an ordered pair of variables (X, Y) if

1. Z intercepts all directed paths from X to Y.
2. There is no unblocked path from X to Z.
3. All backdoor paths from Z to Y are blocked by X.

Theorem 2.1 (Front-Door Adjustment)**.** If Z satisfies the front-door criterion relative to (X, Y) and if $P(x, z) > 0$, then the causal effect of X on Y is identifiable and

$$P(y|do(x)) = \sum_z P(z|x) \sum_{x'} P(y|x', z)P(x')$$

2.5 Conditional Interventions and Covariate-Specific Effects

Rule 2.2. The z-specific effect $P(Y = y|do(X = x), Z = z)$ is identified whenever we can measure a set S of variables s.t. $S \cup Z$ satisfies the backdoor criterion. The z-specific effect is

$$P(Y = y|do(X = x), Z = z) = \sum_s P(Y = y|X = x, S = s, Z = z)P(S = s)$$

Note. The adjustment set is $S \cup Z$, not just S, and the summation goes only over S, not including Z. If Z is a subset of S, we have $S \cup Z = S$, and S alone need satisfy the backdoor criterion.

2.6 Mediation

For any three variables X, Y, and Z, where Z is a mediator between X and Y, the controlled direct effect (CDE) on Y of changing the value of X from x to x' is

$$\text{CDE} = P(Y = y|do(X = x), do(Z = z)) - P(Y = y|do(X = x'), do(Z = z))$$

In general, the CDE of X on Y, mediated by Z, is identifiable if the following two properties hold:

1. There exists a set S_1 of variables that blocks all backdoor paths from Z to Y.
2. There exists a set S_2 of variables that blocks all backdoor paths from X to Y, after deleting all arrows entering Z.

3 Counterfactuals

3.1 Defining and Computing Counterfactuals

3.1.1 The fundamental Law of Counterfactuals

Let M_x be the modified version of M, with the equation of X replaced by $X = x$. The formal definition of the counterfactual $Y_x(u)$ is

$$Y_x(u) = Y_{M_x}(u)$$

In general, counterfactuals obey the consistency rule

$$X = x \Rightarrow Y_x = Y$$

If X is binary, then

$$Y = XY_1 + (1 - X)Y_0$$

3.1.2 The Three Steps in Computing Counterfactuals

There is a three-step process for computing any deterministic counterfactual:

(i) Abduction: Use evidence $E = e$ to determine the value of U.

(ii) Action: Modify the model M by removing the structural equations for the variables in X and replacing them with $X = x$, to obtain the modified model M_x.

(iii) Prediction: Use M_x and the value of U to compute the value of Y, the consequence of the counterfactual.

We can generalize to any probabilistic nonlinear system. Given an arbitrary counterfactuals of the form, $\mathbb{E}[Y_{X=x}|E = e]$, the three-step process is:

(i) Abduction: Update $P(U)$ by the evidence to obtain $P(U|E = e)$.

(ii) Action: Modify the model M by removing the structural equations for the variables in X and replacing them with $X = x$, to obtain the modified model M_x.

(iii) Prediction: Use M_x and the updated probabilities over the $U, P(U|E = e)$, to compute the expectation of Y, the consequence of the counterfactual.

3.2 Nondeterministic Counterfactuals

Theorem 3.1 (Counterfactual Interpretation of Backdoor)**.** If a set Z of variables satisfies the backdoor condition relative to (X, Y), then for all x, the counterfactual Y_x is conditionally independent of X given Z, i.e.,

$$P(Y_x|X, Z) = P(Y_|Z)$$

Note. The theorem implies that $P(Y_x = y)$ is identifiable by the adjustment formula

$$\begin{aligned} P(Y_x = y) &= \sum_z P(Y_x = y|Z = z)P(z) \\ &= \sum_z P(Y_x = y|Z = z, X = x)P(z) \quad \text{(Theorem 3.1)} \\ &= \sum_z P(Y = y|Z = z, X = x)P(z) \quad \text{(Consistency Rule)} \end{aligned}$$

Theorem 3.2. Let τ be the slope of the total effect of X on Y,

$$\tau = \mathbb{E}[Y|do(x+1)] - \mathbb{E}[Y|do(x)]$$

then for any evidence $Z = e$, we have

$$\mathbb{E}[Y_{X=x}|Z = e] = \mathbb{E}[Y|Z = e] + \tau(x - \mathbb{E}[X|Z = e])$$