

Notes for Causal Inference

MoeKid101

1 Causal Model

Def Exogenous Variables: The set of variables that are determined outside the model and act as the driving force of the model.

Def Endogenous Variables: The set of variables that are influenced by other variables inside the model.

Def Cause / Direct Cause: X is a cause of Y if X is a direct cause of Y or the cause of any cause of Y .

A **structural causal model** (SCM) has two sets of variables, exogenous variables U , endogenous variables V and a set of functions f used to determine the values of endogenous variables based on other variables (when variables in U are determined, f gives all values in V). Each SCM is associated with a **graphical causal model** represented by a DAG $G = (\mathcal{V}, \mathcal{E})$ where $(X, Y) \in \mathcal{E}$ represents direct causality.

Thm The Rule of Product Decomposition: In a graphical causal model G , $\Pr[x_1, \dots, x_n] = \prod_i \Pr[x_i | pa_i]$ where X_1, \dots, X_n are variable nodes.

An unspecified graphical causal model means we know only the topological structure of the graph, but are ignorant of the specific functions f . With only the information given by unspecific graph we can know many causal relationships between variables.

- X and Y are dependent if $(X, Y) \in \mathcal{E}$ or $(Y, X) \in \mathcal{E}$.
- X and Y are *likely* dependent if there exists a path between X and Y .
- X and Y are conditionally independent given Z if (1) there is only one path between X and Y ; (2) Z intercepts the path.
- Y and Z are independent conditional on X if (1) X is a common cause of Y and Z ; (2) there exists only one path between Y and Z in the undirected graph.
- X and Y are unconditionally independent but conditionally dependent given Z or any descendants of Z if (1) Z is the collision node between X and Y ; (2) there exists only one path between X and Y in the undirected graph.

With these basic rules dealing with chains, forks and colliders, we can use **d -separation** to measure in any graph whether variable nodes are connected (or dependent, termed as **d -connected** whose contrary is **d -separated**). Notice that being d -connected doesn't necessarily mean dependent, but being d -separated means independent.

Def Blocking: For DAG G and its undirected version \bar{G} , a path p in \bar{G} is blocked by a set of nodes Z iff.

- (1) p contains a chain $A \rightarrow B \rightarrow C$ or a fork $A \leftarrow B \rightarrow C$ and $B \in Z$;
- (2) p contains a collider $A \rightarrow B \leftarrow C$ and none of B and its descendants are in Z .

Def d -Separated: X and Y are d -separated conditional on Z (and thus conditionally independent) if Z blocks every path between X and Y in \bar{G} .

2 Intervention

Usually the goal of statistical studies are to study the effect of interventions (through randomized controlled experiments), but such experiments aren't always practical. Intervening on a variable $\Pr[Y = y | do(X = x)]$ and conditioning on a variable $\Pr[Y = y | X = x]$ are different in that intervening changes the system, through which other variables are changed, however in conditioning, we are simply narrowing the focus to a subset of the system. Specifically, intervention changes the graphical causal model while conditioning doesn't.

Thm The Causal Effect Rule: Given graph G where pa is the set of parents of X , then

$$\Pr[Y = y | do(X = x)] = \sum_z \Pr[Y = y | X = x, pa = z] \Pr[pa = z]$$

Proof: Notice that when intervention $X = x$ is introduced, the only changes in the causal graph are the deletion of paths from elements of pa to X , indicating unchanged conditional probabilities in other parts of the graph. Let $\Pr[\dots]$ be the probabilities in the

original graph, and $\mathbf{Pr}_m[\dots]$ be the probabilities in the intervened graph. Then

$$\begin{aligned}\mathbf{Pr}[Y = y|do(X = x)] &= \mathbf{Pr}_m[Y = y|X = x] = \sum_z \mathbf{Pr}_m[Y = y, \mathbf{pa} = z|X = x] \\ &= \sum_z \mathbf{Pr}_m[Y = y|X = x, \mathbf{pa} = z] \mathbf{Pr}_m[\mathbf{pa} = z|X = x] = \sum_z \mathbf{Pr}[Y = y|X = x, \mathbf{pa} = z] \mathbf{Pr}[\mathbf{pa} = z]\end{aligned}$$

Q.E.D.

Such result helps us generalize the intervention to multiple variables. Recall that $\mathbf{Pr}[X_1, \dots, X_n] = \prod_i \mathbf{Pr}[X_i|\mathbf{pa}_i]$ where $\mathcal{V} = \{X_1, \dots, X_n\}$. When we intervene a set of variables $S \subset \mathcal{V}$ to be s_i , we have $\mathbf{Pr}_m[X_i|\mathbf{pa}_i] = \begin{cases} \mathbf{Pr}[X_i|\mathbf{pa}_i], & X_i \notin S \\ 1, & X_i \in S \wedge x_i = s_i \\ 0, & X_i \in S \wedge x_i \neq s_i \end{cases}$ therefore, as long as the values of intervened variables coincide with intervention values, the product decomposition rule is stated as $\mathbf{Pr}[X_1, \dots, X_n|do(S = s)] = \prod_{i \notin S} \mathbf{Pr}[X_i|\mathbf{pa}_i]$.

Generally, problem arises from the fact that not every represented variable in the causal graph is measurable (from the perspective of reality). Therefore we might rely on other sets of variables than the parents of X to measure the causal effect of X on another variable Y .

Def The Backdoor Criterion: Given ordered (X, Y) in directed G with undirected \overline{G} , a set of variables Z satisfies backdoor criterion relative to (X, Y) if no node in Z is a descendant of X in G , and Z blocks every path in \overline{G} between X and Y that contains an arrow into X . (Intuitively, Z blocks all spurious paths (those with arrows into X) and creates no new spurious paths (by not including descendants of X).)

Thm : If Z satisfies backdoor criterion relative to (X, Y) , then $\mathbf{Pr}[Y = y|do(X = x)] = \sum_z \mathbf{Pr}[Y = y|X = x, Z = z] \mathbf{Pr}[Z = z]$.