Identifiability and the Backdoor Criterion

Statistics · Module 11 — Causality

Node 11.3

Position in Knowledge Graph

System: Statistics. Structural Domain: Causality (Module 11). Node 11.3.

Identifiability is the question that precedes all estimation: can the causal effect be expressed as a functional of the observed data distribution? This node develops the backdoor criterion (the most commonly applicable graphical condition), the front-door criterion (for unmeasured confounding settings), and the do-calculus (the complete set of rules for converting interventional distributions to observational ones). It connects the potential outcomes framework (Node 11.1) to the DAG framework (Node 11.2) and enables both experimental (Node 11.4) and observational (Node 11.5) estimation.

Formal Definition

Let $\mathcal{G}$ be a causal DAG over (T, Y, X), where T is treatment, Y is outcome, and X is a vector of observed covariates. The target is

$$\tau = \mathbb{E}[Y \mid do(T = 1)] - \mathbb{E}[Y \mid do(T = 0)].$$

A causal quantity $Q = Q(p(\cdot \mid do(\cdot)))$ is **identified** from the observational distribution $p(Y, T, X)$ if it can be uniquely computed from $p(Y, T, X)$ for all structural causal models compatible with $\mathcal{G}$.

Non-identifiability arises when two SCMs generate the same observational distribution but imply different values of Q.

Intuition

The causal effect of T on Y is the change in Y when we physically set T to different values (an intervention). But in observational data, T was not set by the researcher—it was determined by a potentially confounded process. Identifiability asks whether the confounding can be removed by conditioning on observed variables. The backdoor criterion gives a graphical test: if we can block all "backdoor paths" (non-causal paths that create spurious association) by conditioning on observed covariates, the causal effect equals the covariate-adjusted observational contrast.

Structural Role

Dependencies: Requires the potential outcomes framework (Node 11.1) for the estimands and DAGs (Node 11.2) for the graphical language. **Enables:** Experimental design (Node 11.4) as the case where identifiability holds by design; observational estimation methods (Node 11.5) which require identification as a precondition.

Identifiability analysis determines whether causal effects can be estimated at all, prior to any estimation procedure. It is the logical prerequisite for all of Nodes 11.4–11.6.

Mathematical Development

Backdoor Paths

A **backdoor path** from T to Y is any path that begins with an arrow into T (i.e., the first edge on the path is $\cdot \to T$). Backdoor paths represent non-causal associations between T and Y—they transmit confounding.

The Backdoor Criterion

Definition. A set of variables X satisfies the **backdoor criterion** relative to (T, Y) in $\mathcal{G}$ if:

1. No variable in X is a descendant of T.
2. X d-separates T from Y in the graph $\mathcal{G}_{\underline{T}}$ (the graph with all arrows out of T removed), or equivalently, X blocks every backdoor path from T to Y.

Theorem: Backdoor Adjustment Formula

Theorem (Pearl, 1993). If X satisfies the backdoor criterion relative to (T, Y), then

$$p(Y = y \mid do(T = t)) = \int p(Y = y \mid T = t, X = x)\, p(X = x)\, dx.$$

Proof sketch. In the mutilated graph $\mathcal{G}_{\overline{T}}$ (arrows into T removed, representing $do(T = t)$):

$$p(y \mid do(t)) = \int p(y \mid t, x, \mathcal{G}_{\overline{T}})\, p(x \mid \mathcal{G}_{\overline{T}})\, dx.$$

By condition 1, X is not a descendant of T, so removing arrows into T does not affect the distribution of X: $p(x \mid \mathcal{G}_{\overline{T}}) = p(x)$.

By condition 2, X blocks all backdoor paths, so in $\mathcal{G}_{\overline{T}}$, $Y \perp\!\!\!\perp$ (removed parents of T) $\mid T, X$. This means the conditional $p(y \mid t, x)$ is the same in the original and mutilated graphs: $p(y \mid t, x, \mathcal{G}_{\overline{T}}) = p(y \mid t, x)$.

Combining: $p(y \mid do(t)) = \int p(y \mid t, x)\, p(x)\, dx$.

The **ATE** is therefore:

$$\tau = \mathbb{E}_X\big[\mathbb{E}[Y \mid T = 1, X] - \mathbb{E}[Y \mid T = 0, X]\big].$$

This is the **g-formula** or **adjustment formula**.

Connection to Ignorability

The backdoor criterion in the DAG framework is equivalent to ignorability in the potential outcomes framework:

$$X \text{ satisfies backdoor} \iff \{Y(0), Y(1)\} \perp\!\!\!\perp T \mid X.$$

The DAG makes the structural reason explicit: X blocks all non-causal paths. The potential outcomes framework states the consequence: treatment is as-good-as-random conditional on X.

The Front-Door Criterion

When no measured set satisfies the backdoor criterion (unmeasured confounding between T and Y), identification may still be possible via a **mediator** M.

Definition. A set M satisfies the **front-door criterion** relative to (T, Y) if:

1. T blocks all backdoor paths from M to Y: there are no unblocked backdoor paths from M to Y that do not go through T.
2. M intercepts all directed paths from T to Y: every causal path from T to Y goes through M.
3. There are no backdoor paths from T to M: the effect of T on M is identified without adjustment.

Front-Door Adjustment Formula:

$$p(Y = y \mid do(T = t)) = \sum_m p(M = m \mid T = t) \sum_{t'} p(Y = y \mid M = m, T = t')\, p(T = t').$$

Derivation. Apply the do-calculus in two steps:

Step 1: $p(y \mid do(t)) = \sum_m p(y \mid do(t), do(m))\, p(m \mid do(t))$. Since all $T \to Y$ paths go through M, $p(y \mid do(t), do(m)) = p(y \mid do(m))$.

Step 2: $p(m \mid do(t)) = p(m \mid t)$ (no backdoor paths from T to M).

Step 3: $p(y \mid do(m)) = \sum_{t'} p(y \mid m, t')\, p(t')$ (applying the backdoor adjustment with T blocking backdoor paths from M to Y).

Combining: $p(y \mid do(t)) = \sum_m p(m \mid t) \sum_{t'} p(y \mid m, t')\, p(t')$.

Pearl's do-Calculus

The **do-calculus** consists of three rules for manipulating expressions involving $do(\,\cdot\,)$. Let X, Y, Z, W be disjoint sets of variables in a causal DAG $\mathcal{G}$.

Rule 1 (Insertion/deletion of observations):

$$p(y \mid do(x), z, w) = p(y \mid do(x), w) \quad \text{if } Y \perp_{\mathcal{G}_{\overline{X}}} Z \mid W.$$

Rule 2 (Action/observation exchange):

$$p(y \mid do(x), do(z), w) = p(y \mid do(x), z, w) \quad \text{if } Y \perp_{\mathcal{G}_{\overline{X}\underline{Z}}} Z \mid W.$$

Rule 3 (Insertion/deletion of actions):

$$p(y \mid do(x), do(z), w) = p(y \mid do(x), w) \quad \text{if } Y \perp_{\mathcal{G}_{\overline{X}\,\overline{Z(S)}}} Z \mid W,$$

where $Z(S)$ is the set of nodes in Z that are not ancestors of any node in W in $\mathcal{G}_{\overline{X}}$.

Completeness Theorem (Huang–Valtorta, Shpitser–Pearl, 2006). An interventional distribution $p(y \mid do(x))$ is identifiable from observational data if and only if it can be reduced to an observational expression using the three rules of the do-calculus.

Worked Examples

Example 1: Backdoor Adjustment

DAG: $U \to T$, $U \to Y$, $T \to Y$, $X \to U$, $X \to T$. Observed: T, Y, X. Unobserved: U.

Does X satisfy the backdoor criterion for (T, Y)?

Backdoor paths from T to Y: $T \leftarrow U \to Y$ and $T \leftarrow X \to U \to Y$. The path $T \leftarrow U \to Y$: is it blocked by X? The fork $T \leftarrow U \to Y$ requires $U \in X$ to block, but U is unobserved. The path $T \leftarrow X \to U \to Y$: the fork at X is blocked by conditioning on X.

So X blocks one backdoor path but not $T \leftarrow U \to Y$. The backdoor criterion is **not satisfied** by X alone. We cannot identify the causal effect of T on Y by adjusting for X only; the unmeasured confounder U creates a non-blockable backdoor path.

Example 2: Front-Door Criterion Application

DAG: $U \to T$, $U \to Y$, $T \to M \to Y$. All of T, M, Y observed; U unobserved. All directed paths from T to Y go through M. No backdoor paths from T to M (the only candidate $T \leftarrow U \to \cdots \to M$ does not exist since $U \not\to M$). Backdoor paths from M to Y: $M \leftarrow T \leftarrow U \to Y$, blocked by T.

The front-door criterion is satisfied by M. The causal effect is:

$$\mathbb{E}[Y \mid do(T = t)] = \sum_m P(M = m \mid T = t) \sum_{t'} \mathbb{E}[Y \mid M = m, T = t'] \, P(T = t').$$

Suppose $T, M \in \{0, 1\}$, $P(M = 1 \mid T = 1) = 0.8$, $P(M = 1 \mid T = 0) = 0.2$, $P(T = 1) = 0.5$, and the conditional means of Y are: $\mathbb{E}[Y \mid M = 1, T = 1] = 10$, $\mathbb{E}[Y \mid M = 1, T = 0] = 8$, $\mathbb{E}[Y \mid M = 0, T = 1] = 4$, $\mathbb{E}[Y \mid M = 0, T = 0] = 3$.

$\mathbb{E}[Y \mid do(T = 1)] = 0.8[0.5 \cdot 10 + 0.5 \cdot 8] + 0.2[0.5 \cdot 4 + 0.5 \cdot 3] = 0.8 \cdot 9 + 0.2 \cdot 3.5 = 7.2 + 0.7 = 7.9.$

$\mathbb{E}[Y \mid do(T = 0)] = 0.2 \cdot 9 + 0.8 \cdot 3.5 = 1.8 + 2.8 = 4.6.$

Causal effect: $\tau = 7.9 - 4.6 = 3.3$.

Cross-Links

AI. Causal inference in reinforcement learning requires identifying the effect of policies (interventions on action sequences) from offline data. The do-calculus generalizes to sequential settings where the "treatment" is a sequence of actions, and identification requires blocking backdoor paths at each time step.

Economics. The g-formula (adjustment formula) is the semiparametric identification result underlying the potential outcomes approach in economics. Instrumental variables (Node 11.6) handle cases where the backdoor criterion fails due to unmeasured confounders.

Linear Algebra. The adjustment formula $\mathbb{E}[Y \mid do(T = t)] = \sum_x p(y \mid t, x) p(x)$ is a marginalization operation: a matrix-vector product when X is discrete, with the transition matrix $p(y \mid t, x)$ and the marginal distribution $p(x)$ as the vector.

Quantum Mechanics. In quantum causal models, the do-operator corresponds to a quantum intervention (a completely positive trace-preserving map applied to a quantum system), and the analogue of the backdoor criterion determines when quantum causal effects are identifiable from observed quantum correlations.

Structural Compression

Invariant

A causal effect is identified if and only if all non-causal (backdoor) paths between treatment and outcome can be blocked by observed variables; the adjustment formula then expresses the causal effect as a weighted average of the observational conditional.

Mechanism

The backdoor criterion uses d-separation in the mutilated graph to ensure that $p(y \mid t, x) = p(y \mid do(t), x)$. Marginalizing over X with respect to $p(x)$ (not $p(x \mid t)$) removes the confounding. The front-door criterion chains two identification steps through a mediator. The do-calculus provides the complete algorithmic procedure for testing identifiability in arbitrary DAGs.

Cross-System Echo

The g-formula is the causal analogue of the law of total probability. In AI, offline policy evaluation requires identifying the effect of a target policy from data collected under a behavior policy—the sequential generalization of backdoor adjustment via importance weighting. In economics, the Frisch–Waugh theorem for instrumental variables is the linear-model specialization of identification under confounding.

Exercises

1. For the DAG $Z \to T \to Y$, $Z \to Y$, verify that Z satisfies the backdoor criterion for (T, Y) and write out the adjustment formula.
2. Prove the backdoor adjustment formula by showing that in the mutilated graph $\mathcal{G}_{\overline{T}}$, $p(y \mid t, x, \mathcal{G}_{\overline{T}}) = p(y \mid t, x)$ and $p(x \mid \mathcal{G}_{\overline{T}}) = p(x)$ when X satisfies the backdoor criterion.
3. Construct a DAG where X contains a descendant of T, and show that conditioning on X biases the estimate of the causal effect (an instance of "bad control").
4. Apply the front-door criterion to the smoking–tar–cancer DAG (smoking $\to$ tar deposits $\to$ cancer, with an unmeasured common cause of smoking and cancer) and derive the front-door adjustment formula.
5. Using the three rules of the do-calculus, reduce $p(y \mid do(t))$ to an observational expression for the DAG $T \to M \to Y$, $U \to T$, $U \to Y$, verifying the front-door formula.
6. Show that for the DAG $T \to Y \leftarrow U \to T$ with U unobserved and no other variables, the causal effect $\mathbb{E}[Y \mid do(T = t)]$ is not identified. Explain why the do-calculus fails.
7. Argue, structurally, why completeness of the do-calculus means that identifiability is a purely graphical question—it depends on the DAG structure and which variables are observed, not on the functional forms or parameter values of the structural equations.

Statistics · Module 11 — Causality · Node 11.3
Concepts: conditional-probability, graph-theory
Prerequisites: 11.1, 11.2
Enables: 11.4, 11.5

hbar.university — own.your.cognition