Causal Inference with Linear Regression: A Modern Approach

By CausAI

CausAI

Instructor

- Jamilla Cooiman
- Cum Laude BSc and MSc Econometrics & Data Science
- Founder of CausAI

CausAI

Linear Regression

Linear Regression is very well-understood, right...?

Can you answer:

- *If our Linear Regression coefficients are unbiased, then what exactly are they unbiased for?*
- *To obtain unbiased coefficient estimates in our model, is it necessary to include all variables that are part of the DGP for the outcome variable?*
- *Under what exact conditions can Linear Regression coefficients be interpreted causally, and under what conditions can they not?*

CausAI

Linear Regression

Linear Regression is very well-understood, right...?

Can you answer:

- *If regression coefficients do represent causal effects, how should we interpret them when our model includes quadratic, cubic, or interaction terms?*
- *What if the "No Omitted Variable Bias" assumption is violated?*

CausAI

Course Goals

- Understand when and how Linear Regression can be used for Causal Inference
- A modern approach using Judea Pearl's Causal Inference framework
- First part of a 2-part course series on Causal Inference with Linear Regression

CausAI

Course Content

- **Module 1:** Basics of Causal Inference
- **Module 2:** Mechanics of Linear Regression with OLS
- **Module 3:** Structural Causal Models, its parameters, and when Linear Regression can uncover Structural Causal Parameters
- **Module 4:** Robustness Tests and Sensitivity Analysis
- Clarify common misconceptions, discuss high-quality Causal Inference sources, coding examples & exercises

CausAI

Who is this course for?

- Everyone with basic Probability, Linear Algebra and Statistics knowledge
- It's beneficial if you know Python
- No prior knowledge on Causal Inference is required

CausAI

Why Causal Inference?

- Mainstream AI Methods don't help us understand the impact of different decisions on future outcomes
- They rely on patterns/associations in data to make predictions
- E.g. An airline company might predict higher demand during holidays because demand and holidays are positively associated
- E.g. A grocery store might predict higher sales of a tropical drink on a sunny day

CausAI

Why Causal Inference?

- But we don’t just want to predict outcomes, we want to **change** them
- Airline company: we don’t just want to predict demand, we want to **increase** demand
- Grocery store: we don’t just want to predict sales, we want to **increase** sales
- Can association-focused models help here?

CausAI

Why Causal Inference?

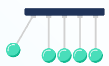
- Airline company: higher prices are associated with higher demand, does this mean increasing the prices will increase demand?
- No… positive association exists because holiday increases both
- *Association/Correlation is not Causation*
- We can’t tweak variable values in association-focused models and see how the outcome changes as a result to measure causal effects

CausAI

Why Causal Inference?

- Most business questions are about the effect of decisions on outcomes
- *What’s the effect of marketing strategy X on sales?*
- *How will launching a loyalty program affect customer retention?*
- *What impact will introducing a new product line have on revenue?*
- *How will offering free delivery influence customer satisfaction?*

We need Causal Inference!



CausAI

Individual Treatment Effect

- Suppose you own a company that has multiple retail stores, and one of the stores struggles with high stockout rates
- You hire a consultant to optimize inventory management, and the stockout rate decreases from 20% to 8% over the next quarter
- You want to know: if you hadn’t hired the consultant, would the stockout rate of the store still have improved?

CausAI

Individual Treatment Effect

- Let T denote the treatment variable, where $T = 1$ means you hire the consultant, and $T = 0$ means you don't
- Let Y be the outcome variable, representing the stockout rate (e.g. $Y = 0.12$ means the stockout rate is 12% after the quarter, $Y = 0.15$ means the stockout rate is 15% after the quarter etc.)
- Denote by $Y(1)$ the stockout rate Y of the store if you hire the consultant, and $Y(0)$ the stockout rate if you don't. Here, $Y(t)$, where $t = 0$ or 1 are called potential outcomes

The quantity of interest is (ITE) $Y(1) - Y(0)$

CausAI

Individual Treatment Effect

- Let T denote the variable representing different treatment options
- A *unit* is the basic entity being studied at a specific point in time (stores, countries, households,...)
- For each unit and treatment pair, the outcome that results from applying that treatment is the potential outcome. We define this as $Y_i(t)$
- Each $Y_i(t)$ is a realization of random variable $Y(t)$, where $Y(t)$ can differ for different units

CausAI

Individual Treatment Effect

The *Individual Treatment Effect* for unit i is defined as (binary treatment)

$$Y_i(1) - Y_i(0)$$

- All potential outcomes have the potential to be observed, but only one is actually observed for a unit
- The observed outcome is called the *factual outcome*, unobserved potential outcomes are called *counterfactual outcomes*
- Since we can never observe counterfactuals, we can never compute ITEs exactly
- It's hard to obtain reliable estimates for ITE's...

CausAI

Average Treatment Effect

Given binary treatment T, the *Average Treatment Effect* is defined as

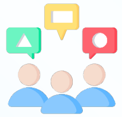
$$ATE = \mathbb{E}[Y(1) - Y(0)] = \mathbb{E}[Y(1)] - \mathbb{E}[Y(0)]$$

- We no longer look at the treatment effect for one specific unit, but rather at the average effect for an entire population
- In the store example: we look at the effect hiring a consultant on average has on the stockout rates of all stores in the chain
- The ATE is the "best guess" for the treatment effect of a randomly selected unit from the population
- It hides individual differences...

CausAI

Average Treatment Effect

- When the treatment effect differs among units, we say that the treatment effects are *heterogeneous*
- The ATE can't capture heterogeneity which results in information loss, but the ITE can't be estimated reliably...
- A midway approach: Conditional Average Treatment Effects!



Conditional Average Treatment Effect

Given binary treatment T, the *Conditional ATE* is defined as

$$CATE(x) = \mathbb{E}[Y(1) - Y(0)|X = x] = \mathbb{E}[Y(1)|X = x] - \mathbb{E}[Y(0)|X = x]$$

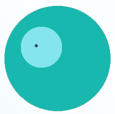
where X represents a set of features describing the population units

- The $CATE(x)$ is just an ATE in a subgroup of the population, defined by characteristics $X = x$
- Since $CATE(x)$ can differ for different values of X, it can capture heterogeneity

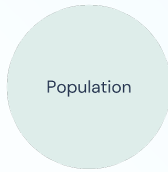
Treatment Effects

In our store example:

- The ATE focuses on populations, so it calculates the average effect of hiring a consultant on all of our stores
- The $CATE(x)$ focuses on subpopulations, so it calculates the average effect of hiring a consultant on stores with similar characteristics $X = x$
- The ITE focuses on individual units, so it calculates the effect of hiring a consultant for one specific store

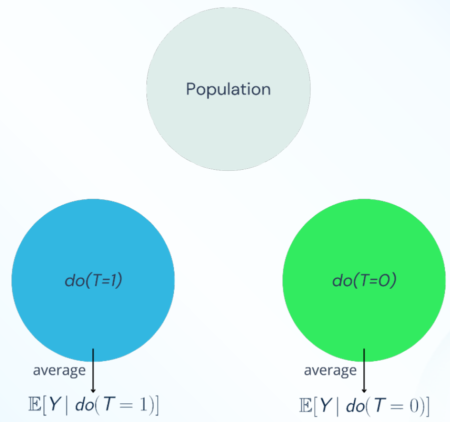


$$ATE = \mathbb{E}[Y(1)] - \mathbb{E}[Y(0)]$$



$$ATE = \mathbb{E}[Y(1)] - \mathbb{E}[Y(0)]$$

$$ATE = \mathbb{E}[Y(1)] - \mathbb{E}[Y(0)]$$



Average Treatment Effect

The ATE can be defined in do-operator notation as (binary treatment)

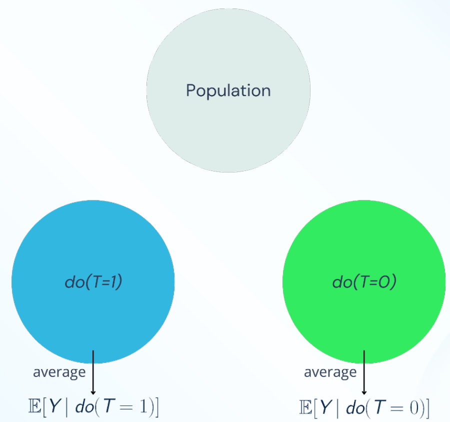
$$ATE = \mathbb{E}[Y \mid do(T=1)] - \mathbb{E}[Y \mid do(T=0)]$$

This looks similar to familiar statistical expression

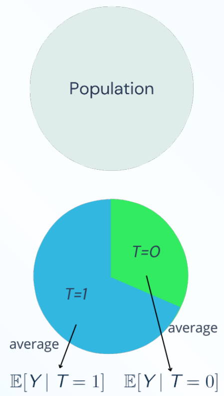
$$\mathbb{E}[Y \mid T=1] - \mathbb{E}[Y \mid T=0]$$

But they are not the same...

Average Treatment Effect

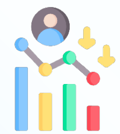


Average Treatment Effect



Example

- Imagine a company that offers a paid subscription service providing additional benefits to its customers
- The company wants to reduce churn (the rate at which customers cancel their subscriptions) by sending out discounts
- But before they do, they want to understand the effect they can expect of sending a discount on the churn rate



Example

	CustomerID	ShoppingFrequency	DiscountSent	Churn
0	1	medium	0	1
1	2	medium	1	0
2	3	low	1	1
3	4	medium	1	0
4	5	high	0	0
...	...	...	...	...
49995	49996	medium	0	1
49996	49997	medium	0	1
49997	49998	low	1	1
49998	49999	low	0	0
49999	50000	medium	1	1

50000 rows × 4 columns

Example

	CustomerID	ShoppingFrequency	DiscountSent	Churn
1	2	medium	1	0
2	3	low	1	1
3	4	medium	1	0
7	8	low	1	1
8	9	low	1	0
...	...	...	...	...
49990	49991	medium	1	0
49991	49992	high	1	0
49993	49994	medium	1	1
49997	49998	low	1	1
49999	50000	medium	1	1

26476 rows × 4 columns

$$\mathbb{E}[Churn \mid DiscountSent = 1]$$

	CustomerID	ShoppingFrequency	DiscountSent	Churn
0	1	medium	0	1
4	5	high	0	0
5	6	high	0	0
6	7	low	0	0
11	12	medium	0	0
...	...	...	...	...
49992	49993	high	0	1
49994	49995	medium	0	0
49995	49996	medium	0	1
49996	49997	medium	0	1
49998	49999	low	0	0

23524 rows × 4 columns

$$\mathbb{E}[Churn \mid DiscountSent = 0]$$

Example

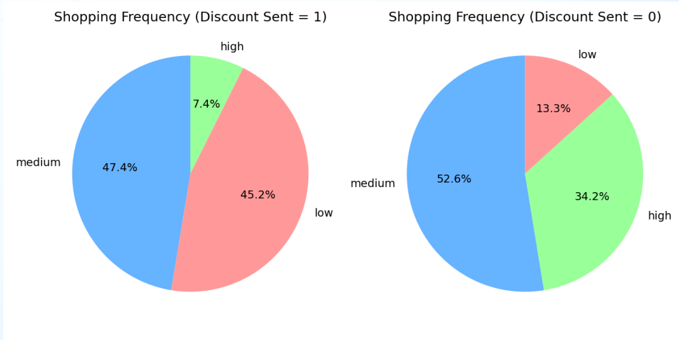
Surprisingly

$$\mathbb{E}[Churn \mid DiscountSent = 1] - \mathbb{E}[Churn \mid DiscountSent = 0]$$

is positive

CausAI

Example



CausAI

Example

- The marketing team tended to send discounts sooner to customers with lower historical shopping frequency
- But historical shopping frequency also directly affects churn probability
- For customers who got discounts we are more likely to see higher churn rates compared to those who didn't get discounts, not necessarily because of whether they got a discount or not, but rather because of differences in their historical shopping frequency (bias!)

CausAI

Example

$$\mathbb{E}[Churn \mid DiscountSent = 1]$$
$$-$$
$$\mathbb{E}[Churn \mid DiscountSent = 0]$$

is positive

CausAI

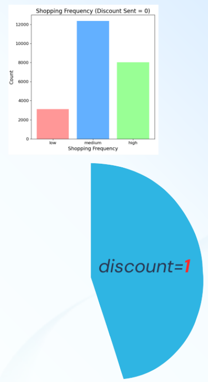
Example

Even if the customers with discount would have actually had discount = 0, we would still be observe a nonzero difference in average outcomes

CausAI

Example

Even if the customers without discount would have actually had discount = 1, we would still be observe a nonzero difference in average outcomes



CausAI

Example

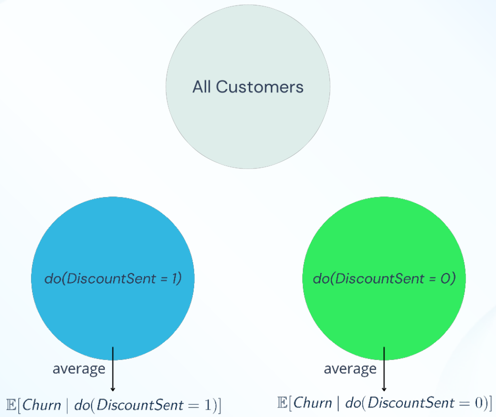
$$ATE = \mathbb{E}[Churn \mid do(DiscountSent = 1)] - \mathbb{E}[Churn \mid do(DiscountSent = 0)]$$

$$\neq$$

$$\mathbb{E}[Churn \mid DiscountSent = 1] - \mathbb{E}[Churn \mid DiscountSent = 0]$$

CausAI

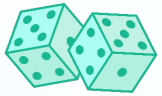
Example



CausAI

Example

- To make the associational difference equal to the ATE, we need that any characteristics that affect the outcome are equally distributed among the groups with different treatments
- Most intuitive way to ensure this: randomize the treatment



Example

The only difference between the groups on average is the difference in treatment (discount sent yes/no)

Example

If discounts are randomized, then

$$ATE = \mathbb{E}[Churn \mid do(DiscountSent = 1)] - \mathbb{E}[Churn \mid do(DiscountSent = 0)]$$

$$=$$

$$\mathbb{E}[Churn \mid DiscountSent = 1] - \mathbb{E}[Churn \mid DiscountSent = 0]$$

Ignorability

The *Ignorability* condition states that
$Y(1), Y(0) \perp\!\!\!\perp T$

$$\mathbb{E}[Y(0)|T = 0] = \mathbb{E}[Y(0)|T = 1] = \mathbb{E}[Y(0)]$$
$$\mathbb{E}[Y(1)|T = 0] = \mathbb{E}[Y(1)|T = 1] = \mathbb{E}[Y(1)]$$

It doesn't matter which group received which treatment, if the groups were switched we would get the exact same result

The groups are *exchangeable*

Ignorability

Under Ignorability (and SUTVA) we can rewrite

$$
\begin{aligned}
ATE &= \mathbb{E}[Y \mid do(T=1)] - \mathbb{E}[Y \mid do(T=0)] \\
&= \mathbb{E}[Y(1)] - \mathbb{E}[Y(0)] \\
&= \mathbb{E}[Y(1) \mid T=1] - \mathbb{E}[Y(0) \mid T=0] \\
&= \mathbb{E}[Y \mid T=1] - \mathbb{E}[Y \mid T=0]
\end{aligned}
$$

We can use observational data to estimate/measure causal effects!

Unfortunately, ignorability is unlikely to hold in observational studies...

CausAI

Conditional Ignorability

The *Conditional Ignorability* condition states that
$Y(1), Y(0) \perp\!\!\!\perp T \mid Z$

The potential outcomes are independent of the treatment assignment, conditional on a set of variables Z

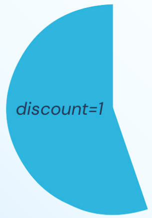
CausAI

Example (revisited)

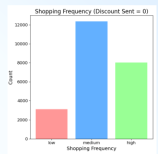
- We couldn't directly compare churn rates between customers who received a discount and those who didn't because they also differed in their historical shopping frequency (which also affects churn probabilities)
- But what if we would compare churn rates between customers who received vs didn't receive a discount **only among those with the same shopping frequency**?
- That's what *conditional ignorability* is about: If we look at **sub**populations where Z attains a fixed value, then the treatment groups are comparable or exchangeable

CausAI

Example

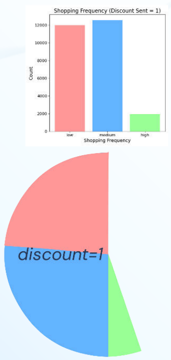


$$
\mathbb{E}[Churn \mid DiscountSent = 1] - \mathbb{E}[Churn \mid DiscountSent = 0] \neq ATE
$$



CausAI

Example



Conditional ignorability given *ShoppingFrequency* means it's valid to compute subgroup *ATE*'s by comparing averages among customers with the same shopping frequency

The Adjustment Formula

The *Adjustment Formula* states that under Conditional Ignorability, we can rewrite

$$\begin{aligned} ATE &= \mathbb{E}[Y \mid do(T=1)] - \mathbb{E}[Y \mid do(T=0)] \\ &= \mathbb{E}_Z[\mathbb{E}[Y \mid T=1, Z] - \mathbb{E}[Y \mid T=0, Z]] \end{aligned}$$

$$\begin{aligned} CATE(x) &= \mathbb{E}[Y \mid do(T=1), X=x] - \mathbb{E}[Y \mid do(T=0), X=x] \\ &= \mathbb{E}_Z[\mathbb{E}[Y \mid T=1, X=x, Z] - \mathbb{E}[Y \mid T=0, X=x, Z]] \end{aligned}$$

Example (revisited)

- *ShoppingFrequency* produced a positive bias in the association measured between *DiscountSent* and *Churn*
- At the same time, it could be that *DiscountSent* has a causal effect on *Churn*, which produces a (negative) association between them

Total Association = Causal Association + Non-causal Association (Bias)

- By conditioning on *ShoppingFrequency*, we 'removed' the bias part:

Total Association = Causal Association

Example (revisited)

$$\mathbb{E}[Churn \mid DiscountSent = 1, ShoppingFrequency = z]$$
$$-$$
$$\mathbb{E}[Churn | DiscountSent = 0, ShoppingFrequency = z]$$

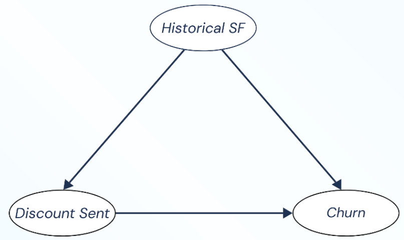
gives us the *ATE* of a subgroup with historical shopping frequency $z, z \in \{low, medium, high\}$, because conditional on *ShoppingFrequency*, any association between *DiscountSent* and *Churn* is of *causal nature*

Causal Graphs

- *Causal Directed Acyclic Graphs* (Causal Graphs) are visual representations of the existence of cause-and-effect relationships between variables
- *Nodes* represent variables
- *Edges* representing causal relationships between variables
- An edge from node X to Y means changing X could potentially directly effect Y
- The absence of an edge from node X to Y means changing X does not directly affect Y
- Causal DAGs are *acyclic*: there's no sequence of edges that forms a loop

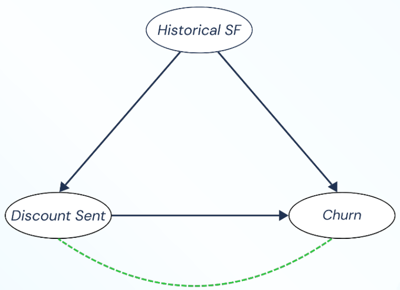
CausAI

Causal Graphs



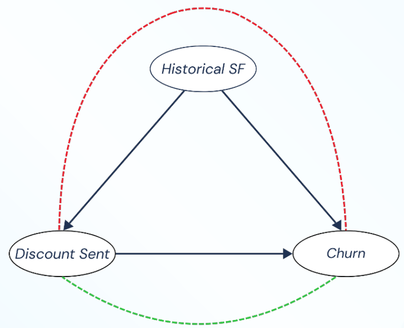
CausAI

Causal Graphs



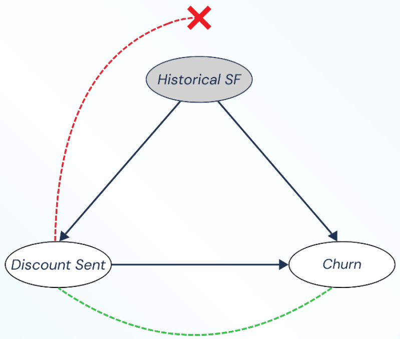
CausAI

Causal Graphs



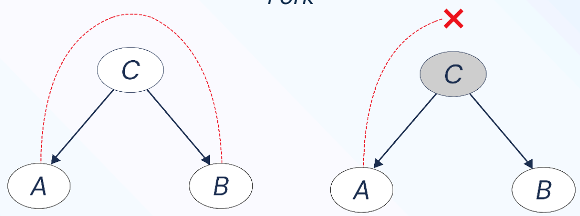
CausAI

Causal Graphs



Graph Patterns

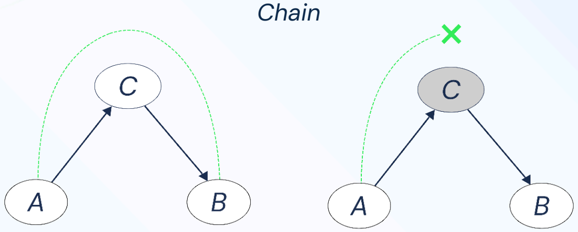
Fork



If we don't condition on *C*, *A* and *B* are dependent (bias),
if we do, *A* and *B* are independent

Graph Patterns

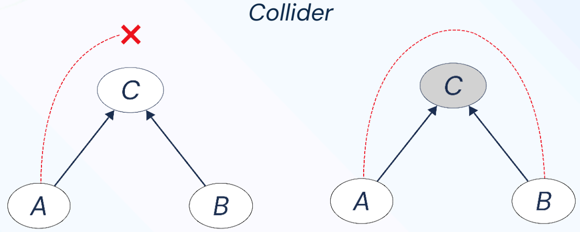
Chain



If we don't condition on *C*, *A* and *B* are dependent (causal),
if we do, *A* and *B* are independent

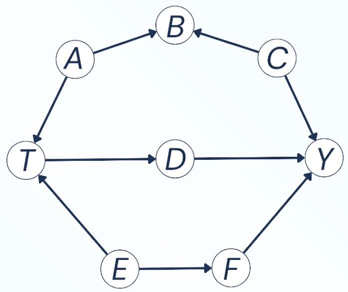
Graph Patterns

Collider



If we don't condition on *C*, *A* and *B* are independent,
if we do, *A* and *B* become dependent (bias)

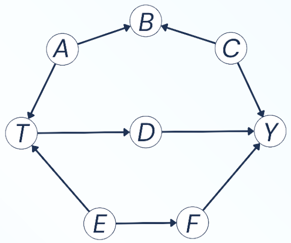
Blocking Paths



Blocking Paths

- A *path* in a Causal Graph is a sequences of edges that connect two nodes, regardless of the direction of the edges
- A *causal path* between T and Y is a directed path where every edge points in the same direction, starting with an arrow out of T and ending with an arrow into Y
- A *non-causal path* is any other path connecting T and Y

Blocking Paths



There are 3 paths connecting T and Y

$$T \leftarrow A \rightarrow B \leftarrow C \rightarrow Y, \qquad T \rightarrow D \rightarrow Y, \qquad T \leftarrow E \rightarrow F \rightarrow Y$$

Blocking Paths

- Each path in a Causal Graph can produce some kind of *association flow* between T and Y
- Along causal paths, either a causal association or no association is produced
- Along a non-causal path, either a biased association or no association is produced
- When there is no flow going along a path, we say that a path is *blocked*
- We can use the properties of forks, chains and colliders to block/open paths

Blocking Paths

A path between nodes T and Y is *blocked* by a (potentially empty) conditioning set Z if either of the following is true:

- Along the path, there is a chain $\cdots \to W \to \cdots$ or a fork $\cdots \leftarrow W \to \cdots$ where $W \in Z$
- Along the path, there is a collider $\cdots \to W \leftarrow \cdots$ that is not conditioned on and none of its descendants are conditioned on ($W \notin Z$, $de(W) \notin Z$)

When all paths between T and Y are blocked by conditioning on a set Z, we say that Z *d-separates* T and Y in the graph

CausAI

Blocking Paths

Path 1: $T \leftarrow A \to B \leftarrow C \to Y$

Contains patterns $T \leftarrow A \to B$, $A \to B \leftarrow C$, $B \leftarrow C \to Y$

Condition 2 of the definition of blocking paths holds by not conditioning on anything, and so the path is blocked (due to collider B)

CausAI

Blocking Paths

Path 2: $T \to D \to Y$

Contains simple chain pattern $T \to D \to Y$

If we don't condition on anything, this path isn't blocked, since both condition 1 and 2 are not satisfied. The path produces a causal association

CausAI

Blocking Paths

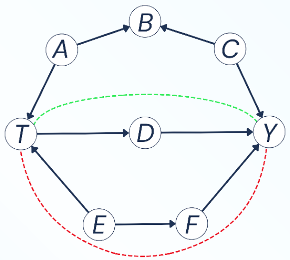
Path 3: $T \leftarrow E \to F \to Y$

Contains patterns $T \leftarrow E \to F$, $E \to F \to Y$

If we don't condition on anything, this path isn't blocked, since both condition 1 and 2 are not satisfied. The path produces a non-causal association

CausAI

Blocking Paths



If we don't condition on anything, the total association between T and Y consists of a causal and non-causal part

CausAI

Backdoor Criterion

The *backdoor criterion* is a method that helps identify a set of variables Z that, if conditioned on, ensure conditional ignorability w.r.t. T and Y

Formally:

Given a pair of variables (T, Y) in a Causal DAG, a set of variables Z satisfies the backdoor criterion relative to (T, Y) if Z blocks every path between T and Y that contains an arrow into T and no node in Z is a descendant of T

CausAI

Backdoor Criterion

"Z blocks every path between T and Y that contains an arrow into T"

Causal paths must start with an outgoing arrow from T, and so any path with an ingoing arrow into T can only produce a biased association flow and must be blocked

"No node in Z is a descendant of T"

Any path between T and Y with an outgoing arrow of T will always only contain either causal flow or no flow if we don't condition on anything

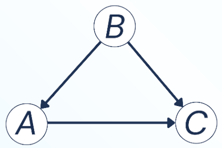
CausAI

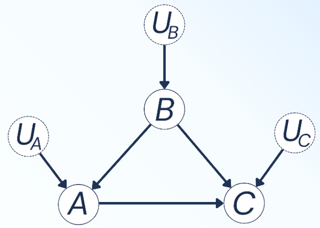
Backdoor Criterion

- If a set of variables Z satisfies the backdoor criterion relative to (T, Y), we have that conditional on Z, all non-causal paths between T and Y are blocked and all causal paths are kept open
- We then have conditional ignorability given Z and we can use the adjustment formula to rewrite the ATE or $CATE$ into purely statistical quantities

CausAI

Double-Headed Arrows

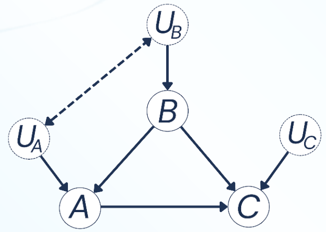


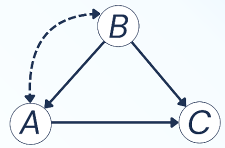


U_A, U_B, U_C represent omitted causes of A, B and C, respectively

By drawing the graph this way, we implicitly assume that U_A, U_B, U_C are independent of each other

Double-Headed Arrows

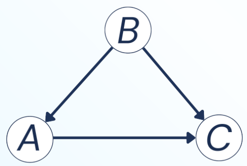




If we believe that two unobserved factors are dependent, we draw a double-headed arrow between them

Double-Headed Arrows

How does the existence of double-headed arrows change the way in which we analyze paths in a graph?

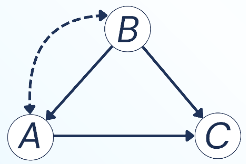


Paths between A and C:

$A \leftarrow B \rightarrow C$ (blocked by set $\{B\}$)
$A \rightarrow C$ (unblocked causal path)

Double-Headed Arrows

How does the existence of double-headed arrows change the way in which we analyze paths in a graph?

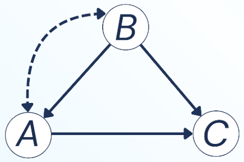


Paths between A and C:

$A \leftarrow B \rightarrow C$ (blocked by set $\{B\}$)
$A \rightarrow C$ (unblocked causal path)
$A \leftrightarrow B \rightarrow C$

Double-Headed Arrows

How does the existence of double-headed arrows change the way in which we analyze paths in a graph?



Paths between A and C:

$A \leftarrow B \rightarrow C$ (blocked by set $\{B\}$)
$A \rightarrow C$ (unblocked causal path)
$A \leftarrow \cdots \rightarrow B \rightarrow C$ (blocked by set $\{B\}$)

Double-Headed Arrows

In general, if there is a path with $\leftrightarrow$, just replace each $\leftrightarrow$ by $\leftarrow \cdots \rightarrow$ and apply blocking path techniques like usual

For example, rewrite a path

$$A \rightarrow B \rightarrow C \leftrightarrow D \leftrightarrow E \rightarrow F$$

as

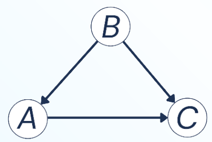
$$A \rightarrow B \rightarrow C \leftarrow \cdots \rightarrow D \leftarrow \cdots \rightarrow E \rightarrow F$$

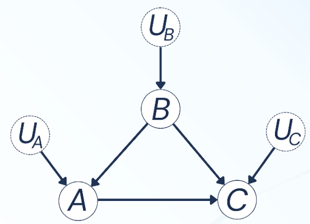
Do-operator in Causal Graphs

- $do(X = x)$ means that we enforce the value of $X = x$ over a population
- The natural mechanism that determines the value of X is no longer relevant. It is completely replaced by our intervention $do(X = x)$
- We can show this in a Causal Graph by removing all incoming arrows into X

Do-operator in Causal Graphs

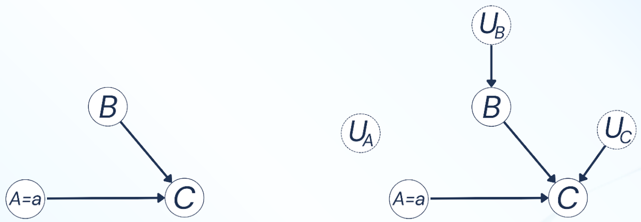
Causal Graph corresponding to natural mechanisms





Do-operator in Causal Graphs

Causal Graph corresponding to mechanisms after intervention $do(A = a)$



CausAI

Continuous Treatments

The *Average Treatment Effect* of going from treatment level $T = t_1$ to $T = t_2$ is defined as

$$ATE(t_1, t_2) = \mathbb{E}[Y \mid do(T = t_2)] - \mathbb{E}[Y \mid do(T = t_1)]$$

For example, the average effect on sales of increasing the price of some product from 10$ to 15$ is defined as

$$ATE(10, 15) = \mathbb{E}[sales \mid do(price = 15)] - \mathbb{E}[sales \mid do(price = 10)]$$

CausAI

Continuous Treatments

The *Conditional Average Treatment Effect* of going from treatment level $T = t_1$ to $T = t_2$ is defined as

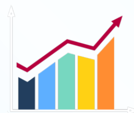
$$CATE(x, t_1, t_2) = \mathbb{E}[Y \mid do(T = t_2), X = x] - \mathbb{E}[Y \mid do(T = t_1), X = x]$$

where X represents a set of features describing the population units

CausAI

Continuous Treatments

- With continuous treatments, the treatment effect can be different for any different pair (t_1, t_2)
- This is known as *non-linearity* in treatment effects
- For example, the effect of increasing a credit limit on people's spending behaviour (stronger effect at lower levels compared to at higher levels)



CausAI

Continuous Treatments

The *Conditional Ignorability assumption* under continuous treatment is no longer defined as $Y(0), Y(1) \perp\!\!\!\perp T \mid Z$, but instead as

$$Y(t) \perp\!\!\!\perp T \mid Z$$

where $Y(t)$ is the potential outcome under treatment $T = t$

The intuition remains the same: It still means that conditional on the variables Z, any group of units with different treatment values only differ on average in their treatment

CausAI

Continuous Treatments

The *Adjustment Formula* for *ATE* and *CATE* of going from treatment level $T = t_1$ to $T = t_2$ is defined as

$$\begin{aligned} ATE(t_1, t_2) &= \mathbb{E}[Y \mid do(T = t_2)] - \mathbb{E}[Y \mid do(T = t_1)] \\ &= \mathbb{E}_Z[\mathbb{E}[Y \mid T = t_2, Z] - \mathbb{E}[Y \mid T = t_1, Z]] \end{aligned}$$

$$\begin{aligned} CATE(x, t_1, t_2) &= \mathbb{E}[Y \mid do(T = t_2)] - \mathbb{E}[Y \mid do(T = t_1)] \\ &= \mathbb{E}_Z[\mathbb{E}[Y \mid T = t_2, X = x, Z] - \mathbb{E}[Y \mid T = t_1, X = x, Z]] \end{aligned}$$

CausAI

Why we need models

- In the coding example, we approximated $\mathbb{E}[Y \mid T = t, Z]$ for different values of Z using subgroup averages
- This doesn't work for continuous treatments, or with many adjustment variables
- Imagine trying to find many other units with the exact same value $T = 0.285$ and $Z = 0.987$
- But even if we don't have exact matches, we might find other data points with *similar* values for the treatment and adjustment variables
- We rely on models to find such 'similar' data points, understand the relationships between the variables, and approximate $\mathbb{E}[Y \mid T = t, Z]$

CausAI

Recap

- the *ATE* and *CATE* compare outcomes under different treatment conditions at a certain point in time of a specific (sub)population
- Under conditional ignorability, we can express causal quantities like the *ATE* and *CATE* in terms of purely statistical or observational quantities
- To find an adjustment set Z to achieve conditional ignorability, we can use Causal Graphs and the Backdoor Criterion
- Double-headed arrows in Causal Graphs imply dependence due to omitted common cause(s)

CausAI