UNIVERSITY OF MARYLAND

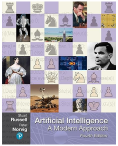
MSML610: Advanced Machine Learning

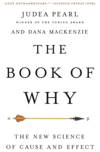
8.2: Causal Inference

Instructor: Dr. GP Saggese - gsaggese@umd.edu

References:

- AIMA (Artificial Intelligence: a Modern Approach)
- Pearl et al., The Book of Why, 2017





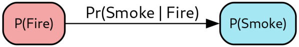
SCIENCE ACADEMY

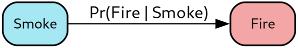
- ***Causal Networks***
    - Causal DAGs
    - Structural Causal Model
    - Variables
    - Type of Variables in Causal AI
    - Types of Paths in Causal AI
    - Intervention and Counterfactuals
    - Randomized Controlled Trial
    - Back-door Adjustment
    - Front-door Adjustment
    - Do-Calculus

- Causal Networks
  - ***Causal DAGs***
  - Structural Causal Model
  - Variables
  - Type of Variables in Causal AI
  - Types of Paths in Causal AI
  - Intervention and Counterfactuals
  - Randomized Controlled Trial
  - Back-door Adjustment
  - Front-door Adjustment
  - Do-Calculus

(Non-Causal) Bayesian Networks

- **Bayesian networks** represent a joint distribution function
  - The direction of the arrow represent *conditional dependence* (not causality)
  - $A \to B$ requires to estimate $\Pr(A|B)$
- **Many possible Bayesian networks** with same nodes, different edges to explain the same phenomenon
- **Example**
  - A Bayesian network with *Fire* and *Smoke*, which are dependent
  - $Fire \to Smoke$
    - Need $\Pr(Fire)$ and $\Pr(Smoke|Fire)$ to compute $\Pr(Fire, Smoke)$
  - $Smoke \to Fire$
    - Need $\Pr(Smoke)$ and $\Pr(Fire|Smoke)$





- **Different Bayesian networks**:
  - Are equivalent and convey the same information
  - Have different difficulties to be estimated
- There is an **asymmetry in nature**
  - Extinguishing fire stops smoke
  - Clearing smoke doesn't affect fire

Causal (Bayesian) Networks

- **Causal networks are Bayesian networks with only causal edges**
  - Use judgment based on nature instead of just statistics
  - E.g., you need to go from
    - *"Are random variables Smoke and Fire correlated?"* to
    - *"What causes what, Smoke or Fire?"*
- **"Dependency in nature"** is like assignment in programming
  - E.g., nature assigns *Smoke* based on *Fire*:
    - ✅ $Smoke := f(Fire)$
    - ❌ $Fire := f(Smoke)$
- **Structural equations** describe "assignment mechanism" in causal graphs

$$X_i := f(X_j) \iff X_j \to X_i$$

Causal DAG

- **Causal DAG**
  - *Directed*: Arrows show cause $\rightarrow$ effect
  - *Acyclic*: No feedback loops
    - Causal relationships assume temporal order: cause before effect
    - A cycle implies a variable is both cause and effect of itself
- **Benefits**
  - DAGs makes explicit *causal* links
  - Support explainable AI models
  - Stability in conditional probability estimation
  - Reason about interventions and counterfactuals
- **Limitations**
  - Requires domain knowledge for structure
  - Assumes all relevant variables included (no hidden confounders)

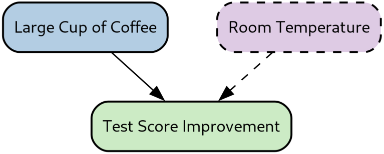
Causal Edges are Stable

- **Causal edges reflect stable relationship**
  - *Mechanistic stability*
    - Causal relationships show system function, not just behavior in one dataset
    - E.g., "*Temperature $\rightarrow$ ice melting rate*" holds true in Alaska and Arizona
  - *Invariance under interventions*
    - If X causes Y, intervening on X affects Y consistently, despite confounders or context changes
  - *Easier estimation through causal modeling*
    - Identifying causal direction focuses estimation on effect size (e.g., regression of Y on X under intervention)

- **Example**: study *Exercise $\rightarrow$ Health*:
  - Correlation may differ in young or elderly populations
  - Causal effect remains stable, as physiological mechanism doesn't change

Causal DAG: Example

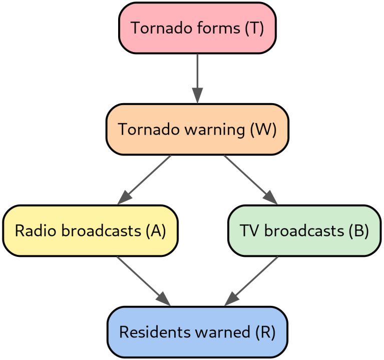
- **Explanatory variables**
  - You can manipulate or observe when changes are applied
  - E.g., *"does a large cup of coffee before an exam help with a test?"*
- **Outcome variables**
  - Result of the action
  - E.g., *"by how much did the score test improve?"*



- **Unobserved variables**
  - Not seen or more difficult to account
  - E.g., *"temperature of the room makes students sleepy and less alert"*

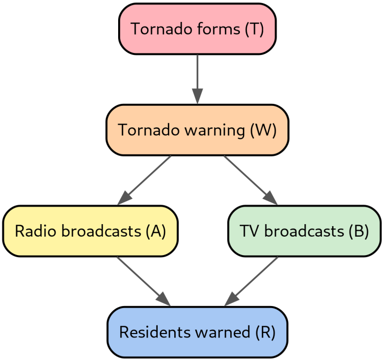
Example: Tornado Warning

- **Example**: Severe weather alert to the public
- **Causal diagram**:
  - T: Tornado forms
  - W: National Weather Service issues tornado warning
  - A, B: Radio and TV broadcast emergency
  - R: Residents warned
- **Assumptions**
  - Channels broadcast only on official order
  - Delivery systems never fail
  - If either channel activates, residents warned



Tornado Warning: Level 1 (Association)

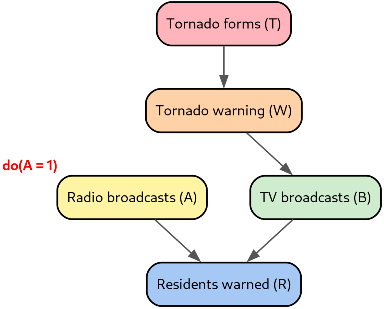
- **Use causal graph to understand how one fact informs another**
  - Use predicate logic to answer
- **If residents are warned, was a warning issued?**
  - Does $R \implies W$?
  - Yes
- **If radio A broadcast, did TV B also broadcast?**
  - Yes
  - True even though A doesn't cause B
  - Correlated through common cause W (confounder)



SCIENCE ACADEMY

Tornado Warning: Level 2 (Intervention)

- **If you make radio broadcast, will residents be warned?**
  - Yes
  - Breaks "rules of nature" as A broadcasts only if W is issued
    - Like removing edge $W \to A$ in graph
    - Set A to *true*: $do(A = 1)$
  - Expect B did *not* broadcast
    - Forcing A shouldn't affect B (no $A \to B$ edge)

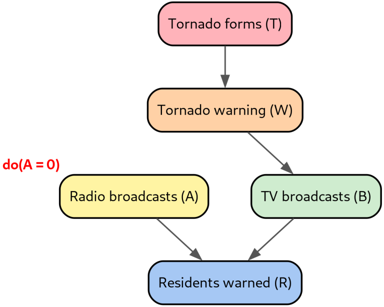


- **Key difference between "seeing" and "doing"**
  - If *see* A broadcast, conclude B broadcast too
  - If *make* A broadcast, expect B didn't
- **Observational data wouldn't reveal causal structure**
  - Either all variables T, W, A, B, R are true or all are false
  - Correlations alone are uninformative

SCIENCE ACADEMY

Tornado Warning: Level 3 (Counterfactual)

- **Assume that residents are warned $R = 1$**
  - You know A and B broadcast because
  - ... a warning W was issued
  - ... after a tornado formed T
- **Would residents be warned if A had not broadcast?**
  - Remove the edge into A
  - Set $do(A = 0)$
  - TV B still broadcasts (due to W), and $R = (A \vee B)$ remains true
  - Yes, $R = 1$



- Causal Networks
  - Causal DAGs
  - ***Structural Causal Model***
  - Variables
  - Type of Variables in Causal AI
  - Types of Paths in Causal AI
  - Intervention and Counterfactuals
  - Randomized Controlled Trial
  - Back-door Adjustment
  - Front-door Adjustment
  - Do-Calculus

Structural Causal Model

- A **Structural Causal Model** (SCM) translates a causal DAG into mathematical equations to define how variables interact
- **Structure of SCMs**
  - *Variables* $X_1, X_2, ..., X_n$ represent quantities in the system
  - *Equations* model each variable as a function of its direct causes
  - Formally, X_i is modeled as:

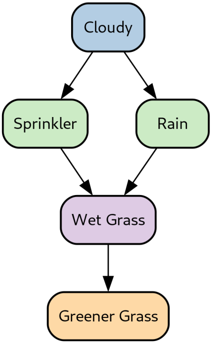
$$X_i = f_i(Parents(X_i), \varepsilon_i)$$

  where:
    - $Parents(X_i)$ are direct causes of X_i
    - ε_i is an exogenous (external, unobserved) noise term
- **Properties**
  - Same properties of causal networks
    - Explain causal relationships between variables
    - Provide a foundation for causal reasoning and simulation
    - ...
  - Quantify effect
- Used in econometrics and genetics for a long time (even before theory of causality)

Structural Causal Model: Sprinkler Example

- **Structural equations** for this causal DAG:

$$\begin{cases} C := f_C(\varepsilon_C) \\ R := f_R(C, \varepsilon_R) \\ S := f_S(C, \varepsilon_S) \\ W := f_W(R, S, \varepsilon_W) \\ G := f_G(W, \varepsilon_G) \end{cases}$$



- **Unmodeled variables** ε_x represent error terms
  - E.g., ε_W is another source of wetness (e.g., *MorningDew*) besides *Sprinkler* and *Rain*
  - Assume unmodeled variables are exogenous, independent, with a certain distribution (prior)

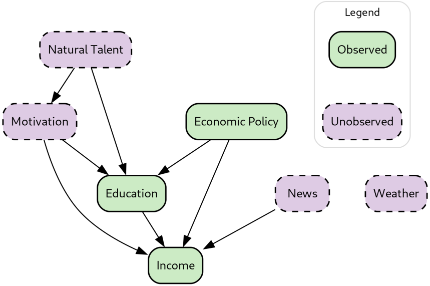
- Express **joint distribution** of all variables as a product of conditional distributions using causal DAG topology:

$$\Pr(C, R, S, W, G) = \Pr(W|R, S) \Pr(G|W) \Pr(S|C) \Pr(R|C) \Pr(C)$$

- Causal Networks
  - Causal DAGs
  - Structural Causal Model
  - ***Variables***
  - Type of Variables in Causal AI
  - Types of Paths in Causal AI
  - Intervention and Counterfactuals
  - Randomized Controlled Trial
  - Back-door Adjustment
  - Front-door Adjustment
  - Do-Calculus

Observed Vs. Unobserved Variables

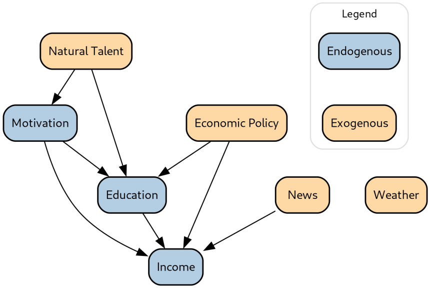
- **Observed variables**
  - Aka "measurable" or "visible"
  - Variables directly measured or collected in a dataset
  - E.g.,
    - Education
    - Income



- **Unobserved variables**
  - Aka "latent" or "hidden"
  - Exist but not measured or included in data
  - E.g.,
    - Natural talent
    - Motivation
  - Ignoring unobserved variables leads to incorrect conclusions
    - E.g., *IceCreamSales* $\leftarrow$ *Temperature* $\rightarrow$ *DrowningRates*

Endogenous Vs. Exogenous Variables

- **Endogenous variables**
  - Values determined *within* the model
    - Dependent on other variables in the system
  - Represent system's internal behavior and outcomes
  - E.g.,
    - Motivation
    - Income

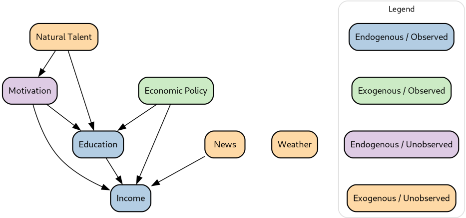


- **Exogenous variables**
  - Originate *outside* the system being modeled
    - Not caused by other variables in the model
  - Represent background conditions or external shocks
  - E.g.,
    - Natural talent
    - Economic policy
    - Weather
    - News

SCIENCE ACADEMY

Endo / Exogenous, Observed / Unobserved Vars

- **Typically**
  - *Exogenous / unobserved variables*: capture randomness or unknown external factors
  - *Exogenous / observed variables*: potential intervention factors
  - *Endogenous / observed variables*: focus for prediction and intervention



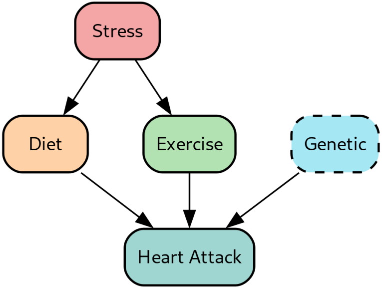
Variable Type	Observability	Example
Endogenous	Observed	Income
Exogenous	Observed	Education
Endogenous	Unobserved	Motivation
Exogenous	Unobserved	Natural Talent

Building a Causal DAG

- **Causal models** visually represent complex environments and relationships
  - Nodes are like "nouns":
    - E.g., "price", "sales", "revenue", "birth weight", "gestation period"
    - Variables can be endogenous/exogenous and observed/unobserved
  - Relationships between variables are "verbs":
    - Parents, children (direct relationships)
    - Descendants, ancestors (along the path)
    - Neighbors
- **Modeling as a Communication Tool**:
  - Shared language bridges gaps between technical and non-technical team members
- **Iterative Refinement**:
  - Continuously update models with new variables and insights

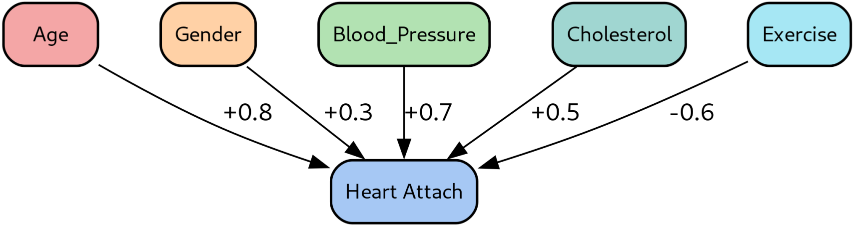
Heart Attack: Example

- Research question: *What's the relationship between stress and heart attacks*?
- **Build a causal DAG**
  - *Stress* is the treatment
  - *Heart attack* is the outcome
  - Stress is *not a direct cause* of heart attack
    - E.g., a stressed person tend to have poor eating habits and tends not to exercise
  - *Genetics* is unobserved



Weights

- Assign weights to paths to represent causal strength
- Sign indicates direction

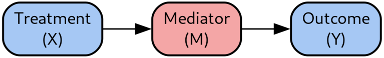


- **How to estimate sign and weight**
  - Estimate using correlation
  - Use priors and then estimate using Bayesian approach

- Causal Networks
  - Causal DAGs
  - Structural Causal Model
  - Variables
  - ***Type of Variables in Causal AI***
  - Types of Paths in Causal AI
  - Intervention and Counterfactuals
  - Randomized Controlled Trial
  - Back-door Adjustment
  - Front-door Adjustment
  - Do-Calculus

Mediator Variable

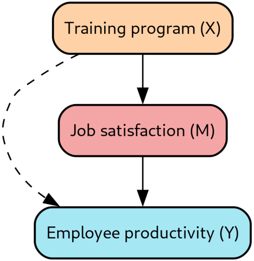
- A **mediator variable** M
  - Is an intermediate variable that *transmits* the causal effect from X (treatment) to Y (outcome)
  - Lies **on the causal path** between X and Y
  - Captures the **mechanism or process** through which X influences Y



SCIENCE ACADEMY

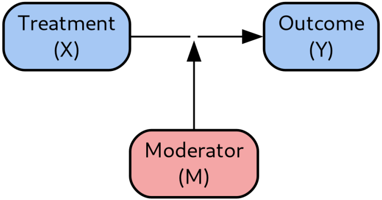
Mediator Variable: Example

- **Research question**: *"Does a training program increase employee productivity?"*
- Causal effect may be indirect, through a **mediator**
  - Training might not immediately boost productivity
  - Could enhance job satisfaction, raising productivity
- **Causal interpretation**
  - X: Training Program (cause)
  - M: Job Satisfaction (mediator)
  - Y: Employee Productivity (effect)
  - Path: $X \to M \to Y$
- **Direct vs. Indirect effects**
  - *Indirect effect*: X affects Y through M
  - *Direct effect*: X affects Y not through M
  - Controlling for M separates effects, clarifying training impact



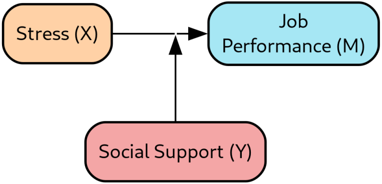
Moderator Variable

- A **moderator variable** M
  - Changes the *strength* or *direction* of the relationship between an independent variable (X) and a dependent variable (Y)
  - Is not part of the causal chain but conditions the relationship



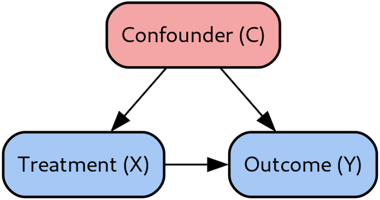
Moderator Variable: Example

- **Research question**: *"Study relationship between stress and job performance"*
- **Social support** M as a moderator
  - High social support weakens stress's negative effect on performance
  - Low social support strengthens stress's negative effect on performance



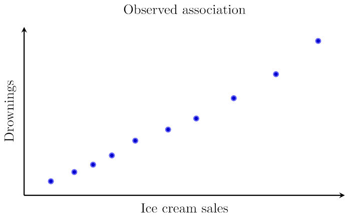
Confounder Variable

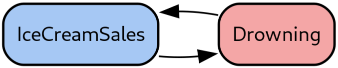
- A **confounder** C
  - Affects both treatment (cause) and outcome (effect)
  - Creates misleading association, if not controlled



Confounder Variable: Example

- *IceCreamSales* and *Drowning* **move together**
  - Correlation-based model claims association
  - Is it true?
  - You can always find an explanation (e.g., from ChatGPT)
    - Eating ice cream may distract children or guardians:
    - Cold food shock reflex causes hyperventilation in water
    - Sugar spike $\rightarrow$ hyperactivity near pools
- **How to use this relationship?**
  - Ban ice cream to prevent drowning?
  - Ice cream maker increase drowning to boost sales?

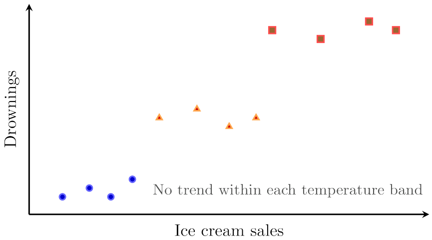


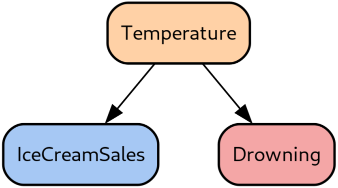


Confounder Variable: Example

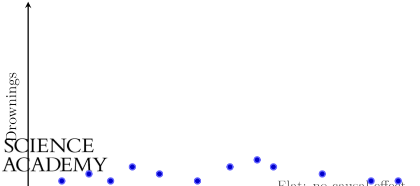
- In reality, **no cause-effect** between *IceCreamSales* and *Drowning*
  - *Temperature* is a confounder
- In fact, when control for temperature (in regression or intervention), association disappears

After controlling for temperature



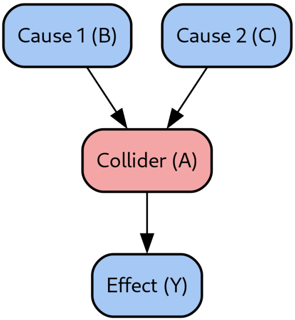


No relationship once temperature is controlled



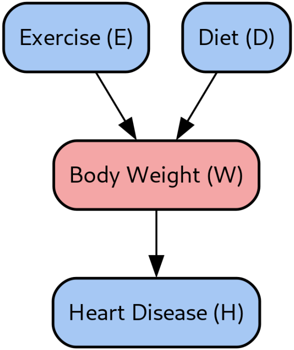
Collider

- A **collider** A
  - Is a variable influenced by multiple variables B, C
  - Complicates understanding relationships between variables B, C and those it influences, Y



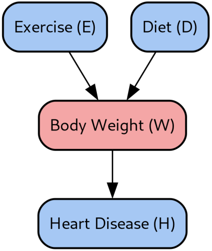
Collider: Examples

- Study the relationship between *Exercise* and *HeartDisease*
  - *Diet* and *Exercise* influence *BodyWeight*
  - *BodyWeight* influences *HeartDisease*
  - *BodyWeight* is a collider



Collider Bias

- Aka "Berkson's paradox"
- **Conditioning on a collider** can introduce a spurious association between its parents by *"opening a path that is blocked"*
- **Example**
  - *Diet* (D)
  - *Exercise* (E)
  - *BodyWeight* (W)
  - *HeartDisease* (D)

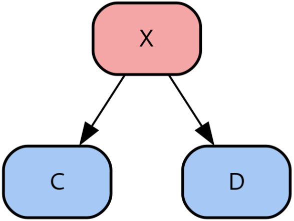


- **Without conditioning on W**
  - E and D are independent
    - E.g., knowing exercise level E doesn't inform about diet D, and vice versa
  - Collider W blocks association between E and D
- **After conditioning on W**
  - E.g., individuals with specific body weight
  - Introduce dependency between E and D
  - With W fixed, changes in E balanced by changes in D, inducing spurious correlation between E and D
  - In Bayesian network it was called *"explaining away"*

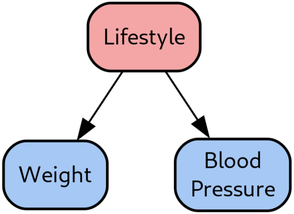
- Causal Networks
  - Causal DAGs
  - Structural Causal Model
  - Variables
  - Type of Variables in Causal AI
  - ***Types of Paths in Causal AI***
  - Intervention and Counterfactuals
  - Randomized Controlled Trial
  - Back-door Adjustment
  - Front-door Adjustment
  - Do-Calculus

Fork Structure

- A **fork** $D \leftarrow X \rightarrow C$ occurs when a single variable causally influences two or more variables
  - X is a **confounder** (common cause) of C and D
  - Forks induce statistical dependence between C and D even if C and D are not causally linked
- **Conditioning** on X *blocks the path* and removes spurious correlation



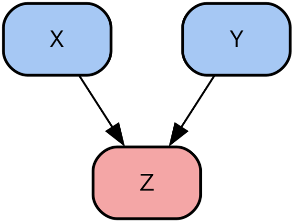
- **Example**
  - *Lifestyle* is a confounder that affects both *Weight* and *BloodPressure*
  - These outcomes may appear correlated due to shared cause

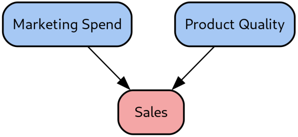


SCIENCE ACADEMY

Inverted Fork

- An **inverted fork** occurs when two or more arrows converge on a common node
  - **Colliders** block associations unless the collider or its descendants are conditioned on
- **Conditioning on a collider** *opens a path*, inducing spurious correlations
  - This is the basis of selection bias
- **Example**
  - Sales influenced by multiple independent causes
  - *MarketingSpend* and *ProductQuality* both influence *Sales*
  - Conditioning on *Sales* can induce false dependence between *MarketingSpend* and *ProductQuality*



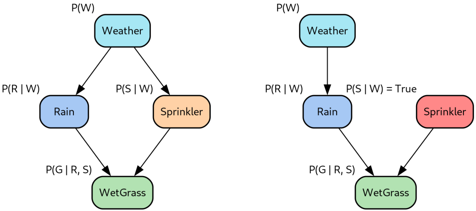


- Causal Networks
  - Causal DAGs
  - Structural Causal Model
  - Variables
  - Type of Variables in Causal AI
  - Types of Paths in Causal AI
  - ***Intervention and Counterfactuals***
  - Randomized Controlled Trial
  - Back-door Adjustment
  - Front-door Adjustment
  - Do-Calculus

Interventions in Causal Networks

- **Causal Bayesian Networks** represent cause–effect relations between variables
  - E.g., *Rain* $\rightarrow$ *WetGrass*
- **Interventions**
  - *Intervention* means setting a variable to a fixed value, overriding its causal mechanism

- E.g., *"Turning the sprinkler on manually"* regardless of cloudiness
  - Replace equation $S = f_S(C, U_S)$ with $S = \text{true}$
  - Causal link from *Cloudy* to *Sprinkler* is *cut*, forming a new "mutilated" model



Original graph *Mutilated graph*

SCIENCE ACADEMY

Interventions in Causal Networks

- **The do-operator**
  - Denoted as $\mathrm{do}(X = x)$
  - Represents performing an action that *sets* X to x, not *observing* $X = x$
  - $do(X_j = x_j^k)$ removes $\Pr(x_j|\mathrm{parents}(X_j))$ from the product and gives a new joint distribution:

$$P_{X_j = x_j^k}(x_1, \ldots, x_n) = \begin{cases} \prod_{i \neq j} \Pr(x_i|\mathrm{parents}(X_i)) & \text{if } X_j = x_j^k \\ 0 & \text{otherwise} \end{cases}$$

- **Difference between observation and intervention**

$$\Pr(Y|do(X = x) \neq \Pr(Y|X)$$

  - Observing $S = \mathrm{true}$ *provides information* about its causes
    - E.g., information about weather (and the Markov blanket)
  - Intervening with $\mathrm{do}(S = \mathrm{true})$ *breaks* those causal dependencies
    - E.g., it doesn't inform about anything

SCIENCE ACADEMY

Intervention

- **Estimate causal effect** of X_j on X_i with adjustment formula:

$$\Pr(X_i = x_i | \mathrm{do}(X_j = x_j^k)) = \sum_{\mathrm{parents}(X_j)} \Pr(x_i | x_j^k, \mathrm{parents}(X_j)) \Pr(\mathrm{parents}(X_j))$$

- **Example**
  - In the Sprinkler model $do(S = \text{true})$, gives the new distribution:

$$\Pr(c, r, w, g | do(S = true)) = \Pr(c) \Pr(r|c) \Pr(w|r, S = true) \Pr(g|w)$$

    - Only descendants of *Sprinkler* (i.e., *WetGrass*) change
    - *Weather* and *Rain* remain unaffected

- **Intuition**
  - Do-operator isolates *causal effects* by simulating external manipulation
  - Essential for answering "what if" questions: *What happens if you intervene and change X?*

Counterfactuals

- A **counterfactual** describes what would have happened in the past under a different scenario
  - *"What would the outcome have been, if X had been different?"*
- **Business examples**
  - *"What if we had two suppliers instead of one? Would we have fewer delays?"*
  - *"Would customers be more satisfied if we shipped products in one week instead of three?"*
  - That's what businesses want, but they can't get it from correlation-based models!
- **Causal reasoning**
  - Goes beyond correlation and association
  - Requires a causal model to simulate alternate realities
  - E.g.,
    - Actual: *"A student received tutoring and scored 85%"*
    - Counterfactual: *"What if the student didn't receive tutoring?"*
    - Causal model estimates the alternative outcome (e.g., 70%)
- **Challenges**
  - Requires strong assumptions and accurate models
  - Difficult to validate directly since counterfactuals are unobservable

Causal Discovery

- **Definition**
  - Causal discovery learns causal network structure from data
  - Identify which variables directly cause others (learn causal directions, not just correlations)
- **Approaches to causal discovery**
  - **Search-based methods**
    - Start with an empty or initial model and iteratively modify it (add, reverse, or delete links)
    - Evaluate each candidate network based on fit to data (e.g., likelihood)
    - Use search strategies while ensuring the network remains acyclic
  - **Constraint-based methods**
    - Infer causal directions from conditional independence tests among variables
    - If X and Y are independent given Z, this constrains possible arrows
- **Dealing with complexity**
  - Possible network structures grow super-exponentially with the number of variables
  - Complexity penalties to avoid overfitting
- **Causality connection**
  - Causal discovery bridges Bayesian learning and causal inference
  - Under certain assumptions, infer causality from observational data not only from experiments

- Causal Networks
  - Causal DAGs
  - Structural Causal Model
  - Variables
  - Type of Variables in Causal AI
  - Types of Paths in Causal AI
  - Intervention and Counterfactuals
  - ***Randomized Controlled Trial***
  - Back-door Adjustment
  - Front-door Adjustment
  - Do-Calculus

What is a Randomized Controlled Trial?

- **RCTs estimate causal effects** by comparing treatment and control groups
  - Randomly assign treatment, not chosen by subjects
    - E.g., assign new drug vs placebo by flipping a coin
  - Ensure groups are statistically equivalent except for treatment
  - Isolate treatment effect
- **Goal**: estimate $\Pr(Y|do(X))$
  - Randomization simulates do-operator by removing incoming arrows to X
  - Eliminate (known and unknown) confounding paths from background variables
  - Turn observational data into experimental data
    - Use $\Pr(Y|X)$ to measure $\Pr(Y|do(X))$
  - Allow causal inference without knowing the causal graph
- **Pros**
  - Implement intervention in a principled, unbiased way
  - Foundation for scientific experimentation and evidence-based policy
  - Gold standard of causal inference (when feasible)

Randomized Controlled Trial: Example

- **Research question**: *"Does offering an after-school tutoring program increase the probability that a student passes the end-of-term exam?"*
- **Population**: eligible students in a district with proper sample size n
- **Treatment and control**
  - $X = 1$ (treatment): student is offered/assigned to tutoring
  - $X = 0$ (control): student is not offered tutoring
- **Assignment mechanism (RCT)**
  - Students are randomly assigned to $X \in \{0, 1\}$
  - Randomization ensures, in expectation, balance on prior GPA, motivation, parental income, etc
- **Outcome**
  - $\Pr(Y \mid do(X)) = \Pr(Y \mid X)$
  - $Y_{X=x} = I\{\text{pass exam}\}$ measured at term's end for treatment vs control
  - Measure and report $Y_{X=1} - Y_{X=0}$

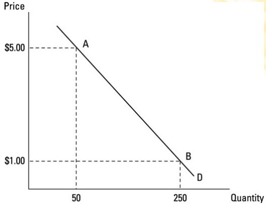
Randomized Controlled Trial: Limits

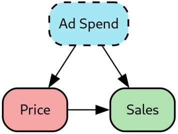
- **May be unethical**
  - E.g., assigning harmful treatment
  - E.g., you want to verify if asbestos causes cancer
- **Can be expensive or impractical**
- **It doesn't always work**
  - Non-compliance: some participants may not follow assigned treatment
  - Attrition: dropout rates may differ between groups
  - May not generalize to broader populations
  - Requires careful implementation and monitoring
- **Blind RCT**: participants don't know which group they're in (e.g., placebo)
- **Double-blind RCTs**: participants and investigators/clinicians don't know assignments

- Causal Networks
  - Causal DAGs
  - Structural Causal Model
  - Variables
  - Type of Variables in Causal AI
  - Types of Paths in Causal AI
  - Intervention and Counterfactuals
  - Randomized Controlled Trial
  - ***Back-door Adjustment***
  - Front-door Adjustment
  - Do-Calculus

Back-Door Paths: Example

- **Example**
  - A company wants to understand the causal effect of *price* on *sales*
  - Advertising spend *AdSpend* is a **confounder** since it can affect both:
    - The *price* the company can set
    - E.g., the cost increases to cover advertisement costs and the product is perceived as more valuable
    - The *sales* (directly)
- The **back-door path** is *Price* → *AdSpend* → *Sales*





- The company **needs to control** for *AdSpend* to estimate the causal effect of *Price* on *Sales* by:
  1. Using *AdSpent* as covariate in the regression
  2. Designing experiment holding *AdSpend* constant or randomized
  3. Using back-door criterion

The Back-Door Adjustment

- **Hypotheses**
  - You have a (correct) causal graph
  - You block the back-door paths (needs definition!) that satisfy the back-door criterion (needs definitions!)
- **Thesis**
  - The *adjustment formula* holds

$$\Pr(Y \mid do(X)) = \sum_{z} \Pr(Y \mid X, Z = z) \Pr(Z = z)$$

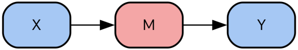
- **Consequences**
  - It allows you an intervention (level 2 of the causality ladder) only using observational data (level 1 of the causality ladder)
    - Correlation implies causation
  - This is an alternative to randomized controlled experiments
  - **Mind blown!**

Back-Door Criterion: Overview

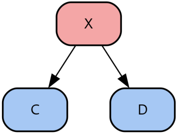
- A **back-door path** is any path from X to Y starting with an arrow into X and ending pointing into Y
  - E.g., $X \leftrightarrow \ldots \leftrightarrow \ldots \rightarrow Y$
  - Arrow direction doesn't matter
  - Unblocked paths create spurious associations
  - *Intuition*: Paths make X and Y look related even if changing X doesn't change Y
- **Back-door criterion**: A set of variables Z satisfies the criterion relative to $X \leftrightarrow \ldots Z \ldots \rightarrow Y$ if:
  - No variable in Z is a descendant of X
  - Z blocks every path from X to Y starting with an arrow into X
  - *Intuition*: Block variables that interfere with the effect you're estimating and remove paths where X and Y connect through common causes
- Condition on variables satisfying the back-door criterion to estimate a causal effect without intervention
  - This guy won the Nobel prize for CS for this! **Mind blown!**

Chains, Forks, and Colliders

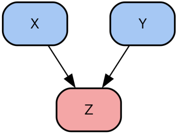
- In a **chain** $X \to M \to Y$
  - Conditioning on M blocks causal effect
  - Mediators must remain unconditioned
  - 💀 *Do not do it*!



- In a **fork** $X \leftarrow Z \to Y$
  - Conditioning on Z removes confounding
  - 😊 *Need to do it*!



- In a **collider** $X \to M \leftarrow Y$
  - Conditioning on M introduces bias
  - Colliders must remain unconditioned
  - 💀 *do not do it*!



Common Mistakes

- The back-door criterion tells you all and only what you need to condition on (i.e., block) to:
  - Transform observation in intervention
  - Say that "correlation IS causation"
- Before this, the solution that researchers used was to "condition on everything"
  - **This is incorrect!**
- **Common mistakes**
  - Conditioning on a descendant of X can bias the estimate
  - Controlling for too many variables can open colliders and introduce bias
  - Forgetting to block all back-door paths
  - Using variables that lie on the causal path (blocks the mediator of the effect)
  - Ignoring unobserved confounders: can make causal effect unidentifiable
- **1000s of papers and their conclusions are wrong!**
  - In medicine, economics, social science people have used observational study incorrectly

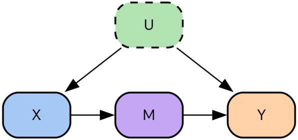
When Back-Door Adjustment Fails

- **Back-door is simple but not universally applicable**
  - No set of observable variables satisfies the back-door criterion
    - In order to condition on a variable it needs to be observable
    - E.g., an unobserved or unknown confounders
  - You need to know a good approximation of the true causal graph
    - Very easy to omit variables
- **Alternatives**:
  - *Front-door criterion*: uses mediators
  - *Instrumental variables*: uses external variation
  - *Do-calculus*: symbolic transformations to eliminate *do*()

- Causal Networks
    - Causal DAGs
    - Structural Causal Model
    - Variables
    - Type of Variables in Causal AI
    - Types of Paths in Causal AI
    - Intervention and Counterfactuals
    - Randomized Controlled Trial
    - Back-door Adjustment
    - ***Front-door Adjustment***
    - Do-Calculus

Front-Door Adjustment in Causal Inference

- **Front-door criterion** identifies causal effects with unobserved confounders
  - Applies when a **mediator variable** transmits all causal influence from treatment to outcome
- Assume the causal graph looks like:
  - X: treatment or cause
  - M: mediator
  - Y: outcome
  - U: unobserved confounder



- **Hypotheses**:
  1. *All directed paths* from X to Y go through M
  2. *No unobserved confounder* affects X and M
  3. *All backdoor paths* from M to Y are blocked by X
- **Thesis**: estimate the causal effect $P(Y|do(X))$ despite unobserved U

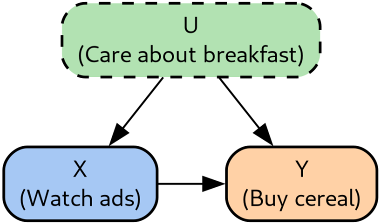
$$P(Y \mid do(X)) = \sum_m P(M \mid X) \sum_{x'} P(Y \mid M, X')P(X')$$

  - *Intuition*: estimate observed link $X \to M$ and $M \to Y$

SCIENCE ACADEMY

Cereal and Ads: Example

- **Research question**: *"Does watching ads (X) make people buy more cereal (Y)?"*
- You might think **yes, of course!**, I say **no, not so fast!**
- **Hidden factor**: *"Parents who care about breakfast (U)"* might:
  - Let kids watch more TV to have them eat breakfast and see ads
  - Buy more cereal anyway



- Hidden factor U confounds "Watch ads" and "Buy cereals"
  - Correlation exists even if ads don't cause it
  - Observing X and Y without controlling for U leads to spurious association
  - Same of "ice cream" and "drowning"
- A spurious relationship is **terrible for the business**!
  - It means you spend money on ads and that doesn't matter
  - Google and Facebook are worth $3T and it's all predicated on "Watch ads" $\implies$ "Buy stuff"

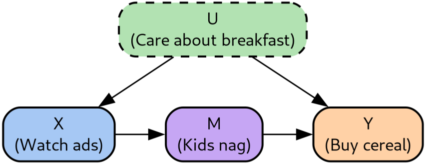
Cereal and Ads: Solutions

- **Strategy 1: Back-door adjustment**
  - If you *know* and can *measure* U "how much parents care about breakfast", include U as a control variable in analysis
  - *Intuition*:
    - Compare families with *the same* breakfast attitudes (U fixed)
    - See if ads (X) still change cereal buying (Y)
- **Strategy 2: Use randomization**
  - Randomized experiments break link between X and U
    - Randomly show ads to some families and not others
    - Randomization ignores parental breakfast attitude; differences in buying come from ads
  - This is why controlled experiments are gold standard for causal inference
- **Strategy 3: Front-door Adjustment**

Cereal and Ads: Finding a Mediator

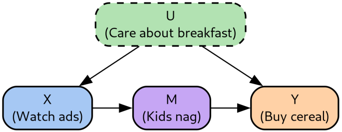
- Imagine ads work by *"making kids ask for cereal"* (aka "nagging") M
  - This is a true advertisement strategy!
  - At the convenience store the candies are at the bottom of the desk
- There is a **mediator** variable

- So the **causal chain** is:
  - Ads (X) $\rightarrow$ ...
  - Kids Nagging (M) $\rightarrow$ ...
  - Parents Buy Cereal (Y)
- The hidden factor *"parents that care about breakfast"* U:
  - Affects how much cereal gets bought
  - Doesn't affect how much kids nag (only ads do that)



When Front-Door Works

- **Is the front-door criterion verified** for *"kids' nagging" M*?
    - Influence of ads on buying goes through nagging ($X \to M \to Y$)
    - No hidden confounders affect both ads and nagging (TV schedule is random, not linked to parents' breakfast attitudes)
    - All confounding between nagging and buying is blocked by controlling for ads



- **Yes! The front-door criterion is verified**
- Instead of doing an intervention $do(X)$, just observe!

1. Observe how often ads make kids nag, $\Pr(M|X)$
2. Observe nagging changes buying, $\Pr(Y|M, X')$
3. Combine both to estimate what happens if you *force* more ads

$$\Pr(Y|do(X)) = \sum_m \Pr(M|X) \sum_{x'} \Pr(Y|M, X') \Pr(X')$$

- *Intuition*: "How ads cause nagging" $\times$ "How nagging causes buying"

SCIENCE ACADEMY

- Causal Networks
  - Causal DAGs
  - Structural Causal Model
  - Variables
  - Type of Variables in Causal AI
  - Types of Paths in Causal AI
  - Intervention and Counterfactuals
  - Randomized Controlled Trial
  - Back-door Adjustment
  - Front-door Adjustment
  - ***Do-Calculus***

Do-Calculus

- **Do-calculus** is a formal system for reasoning about causal effects in graphical models (Judea Pearl, 2000)
- **Problem**
  - You care about **causal effects** like:

    $$\Pr(Y|do(X = x))$$

    - I.e., distribution of Y if you intervene and set X to x, breaking causal links into X
  - You have **observational data** like:

    $$\Pr(Y|X = x)$$

  - In general $\Pr(Y|X) \neq \Pr(Y|do(X = x))$ due to "correlation is not causation" (confounding, ...)
- **Solution**: do-calculus provides algebraic rules to transform intervention expressions (do-operator, i.e.,, $do(X = x)$) into expressions computable from observational data, given certain conditions

The Rules of Do-Calculus

- Do-calculus provides **three transformation rules** for manipulating expressions involving $do()$:

  1. **Insertion/Deletion of Observations:** If $Y \perp Z \mid X, W$ in $G_{\overline{X}}$ (where incoming edges to X are removed), then:

  $$\Pr(Y \mid do(X), Z, W) = \Pr(Y \mid do(X), W)$$

  2. **Action/Observation Exchange:** If $Y \perp Z \mid X, W$ in $G_{\overline{X},\underline{Z}}$ (incoming edges to X removed, outgoing from Z removed), then:

  $$\Pr(Y \mid do(X), do(Z), W) = \Pr(Y \mid do(X), Z, W)$$

  3. **Insertion/Deletion of Actions:** If $Y \perp Z \mid X, W$ in $G_{\overline{X},\overline{Z(W)}}$ (incoming edges to X and to Z excluding those from W removed), then:

  $$\Pr(Y \mid do(X), do(Z), W) = \Pr(Y \mid do(X), W)$$

- These rules allow the systematic reduction of expressions involving $do()$ into observational terms, if the causal graph permits

SCIENCE ACADEMY

Back/Front-door Adjustments and Do-calculus

- The **back-door** and **front-door** criteria are **specific applications** of do-calculus
  - They are simpler, graphical conditions that allow $P(Y \mid do(X))$ to be expressed using observational probabilities
- *Back-door adjustment*: If a set of variables Z blocks all back-door paths from X to Y (paths that go into X), then:

$$\Pr(Y|do(X)) = \sum_{z} \Pr(Y|X,Z)\Pr(Z)$$

- *Front-door adjustment*: If there exists a variable Z such that
  1. Z is affected by X
  2. Z affects Y
  3. All back-door paths from X to Z are blocked
  4. All back-door paths from Z to Y are blocked by X then:

$$\Pr(Y|do(X)) = \sum_{z} \Pr(Z|X) \sum_{x'} \Pr(Y|Z,X')\Pr(X')$$

SCIENCE ACADEMY