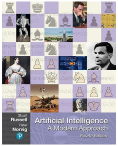
MSML610: Advanced Machine Learning

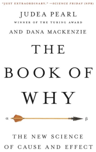
8.2: Causal Inference

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

References:

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



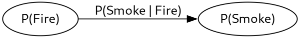


SCIENCE ACADEMY

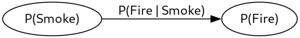
- ***Causal Networks***
  - Variables
  - Type of Variables in Causal AI
  - Paths

(Non-Causal) Bayesian Networks

- **Bayesian networks** represent a joint distribution function
  - The direction of the arrow represent **conditional dependence** (not causality)
    - $A \rightarrow B$ requires to estimate $\Pr(A|B)$
- There are **many possible** *edges* and *node ordering* for the same Bayesian network



- E.g., a Bayesian network with *Fire* and *Smoke*, which are dependent
  - $Fire \rightarrow Smoke$
    - Need $\Pr(Fire)$ and $\Pr(Smoke|Fire)$ to compute $\Pr(Fire, Smoke)$
  - $Smoke \rightarrow Fire$
  - Need $\Pr(Smoke)$ and $\Pr(Fire|Smoke)$



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

SCIENCE ACADEMY

Causal (Bayesian) Networks

- **Causal networks** are Bayesian networks forbidding **non-causal edges**
- Use judgment based on nature instead of just statistics
  - E.g., 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

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

SCIENCE ACADEMY

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 encode *causal* links
  - Reason about interventions and counterfactuals
  - Support explainable AI models
  - Stability in conditional probability estimation
- **Limitations**
  - Requires domain knowledge for structure
  - Assumes all relevant variables included (no hidden confounders)

Causal Edges are Stable

- **Causal edge** $X \to Y$ shows direct causal influence of X on Y, holding other variables constant
  - Captures how manipulating X changes Y, not just their covariance
- Causal edges reflect **stable relationship**
  - Mechanistic stability
    - Causal relationships show system function, not just behavior in one dataset
    - E.g., "Temperature $\to$ 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)
  - Reduced sensitivity to sampling and omitted variables
    - Correlations may change with confounder addition or removal
    - True causal edge persists, stable across model specifications
- **Example**: study *Exercise* $\to$ *Health*:
  - Correlation may differ in young or elderly populations
  - Causal effect remains stable, as physiological mechanism doesn't change

Structural Causal Model

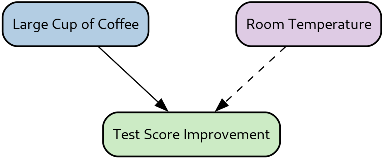
- A **Structural Causal Model** (SCM) translates a causal DAG into mathematical equations
  - DAGs show structure (variables and arrows)
  - SCMs use equations to define how variables interact
- **Structure of SCMs**
  - Variables $X_1, X_2, ..., X_n$ represent quantities in the system
  - Equations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

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

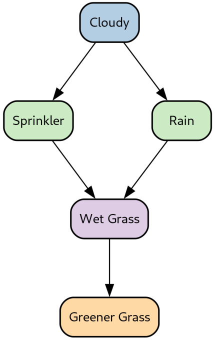
    where:
    - $Parents(X_i)$ are direct causes of X_i
    - ε_i is an exogenous (external, unobserved) noise term
- **Properties**
  - Explain causal relationships between variables
  - Provide a foundation for causal reasoning and simulation
  - Describe how the world works, not just variable correlations

Structural Causal Model: 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"*



Structural Causal Model: Sprinkler Example



- Structural equations for this model:

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

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

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

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

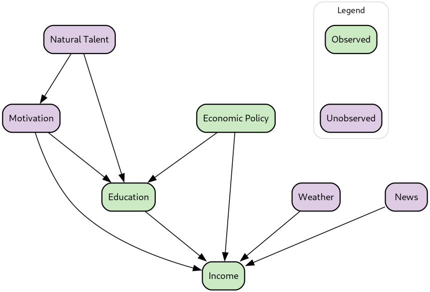
SCIENCE ACADEMY

- Causal Networks
    - ***Variables***
    - Type of Variables in Causal AI
    - Paths

SCIENCE ACADEMY

Observed Vs. Unobserved Variables

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

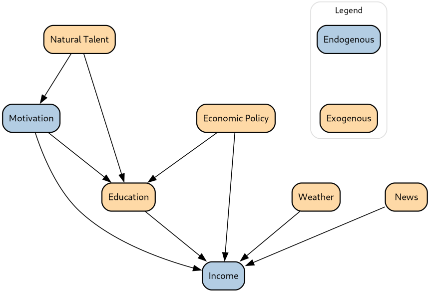


- **Unobserved variables**
  - Aka "latent" or "hidden"
  - Exist but not measured or included in data
  - E.g.,
    - Natural talent
    - Motivation
    - Company culture
  - Ignoring unobserved variables distorts causal relationships
    - Observed: *IceCreamSales* and *DrowningRates*
    - Unobserved: *Temperature*
    - Misleading conclusion: *IceCream* causes *Drowning*

SCIENCE ACADEMY

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

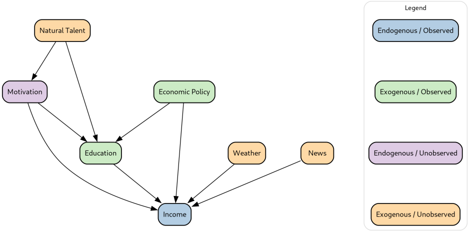
SCIENCE ACADEMY

Endo / Exogenous, Observed / Unobserved Vars

- In **Structural Causal Models**

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

  where:
  - X_i: endogenous
  - ε_i: exogenous noise
- **Typically**
  - *Endogenous variables*: focus for prediction and intervention
  - *Exogenous variables*: capture randomness or unknown external factors



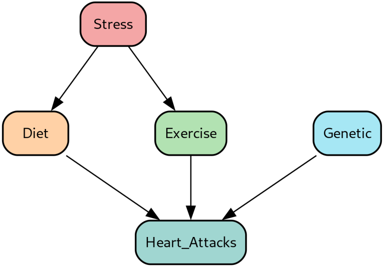
Variable Type	Observability	Example
Endogenous	Observed	Income
Exogenous	Observed	Education
Endogenous	Unobserved	Motivation
Exogenous	Unobserved	Natural Talent

Building a Causal DAG

- **Causal models** visually represent complex environments and relationships
- Nodes are like "nouns" in the model:
  - E.g., "price", "sales", "revenue", "birth weight", "gestation period"
  - Variables can be endogenous/exogenous and observed/unobserved
  - Complex relationships between variables:
    - Parents, children (direct relationships)
    - Descendants, ancestors (along the path)
    - Neighbors
- **Iterative Refinement**:
  - Continuously update models with new variables and insights
- **Modeling as a Communication Tool**:
  - Shared language bridges gaps between technical and non-technical team members
- **Unobservable Variables**:
  - Include variables not empirically observed but known to exist
  - E.g., trust or competitor activity modeled despite lack of direct data

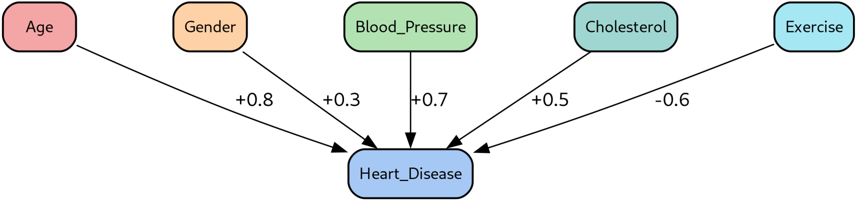
Heart Attack: Example

- What's the relationship between stress and heart attacks?
  - 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



Weights

- Assign weights to paths to represent causal strength
  - Estimate weights using statistical methods
- Sign indicates direction



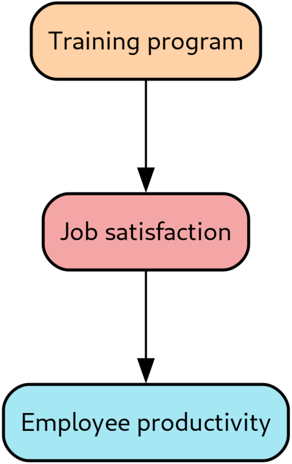
- Causal Networks
  - Variables
  - ***Type of Variables in Causal AI***
  - Paths

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

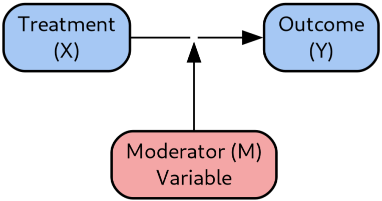
Mediator Variable: Example

- **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



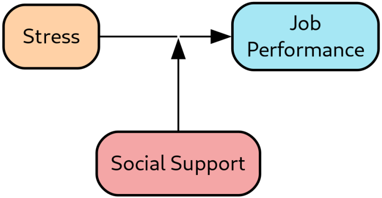
Moderator Variable

- 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



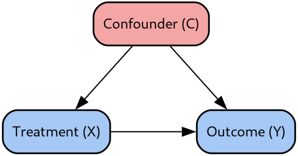
Moderator Variable: Example

- **Research question**: *"Study relationship between stress X and job performance Y"*
- **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

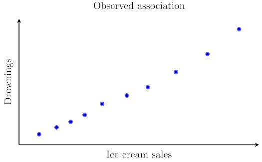
- A **confounder** C
  - Affects both treatment (cause) and outcome
  - Creates misleading association if not controlled

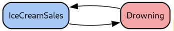


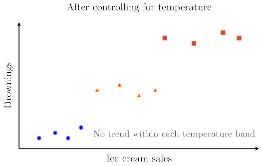
SCIENCE ACADEMY

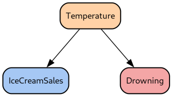
Confounder Variable: Example

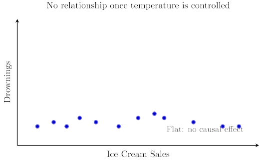
- *IceCreamSales* and *Drowning* move together
    - Correlation-based model claims association, but how to use this relationship?
    - Ban ice cream to prevent drowning?
    - Ice cream maker increase drowning to boost sales?
- No cause-effect between *IceCreamSales* and *Drowning*
    - *Temperature* is a confounder
    - Control for season in regression or intervention, association disappears





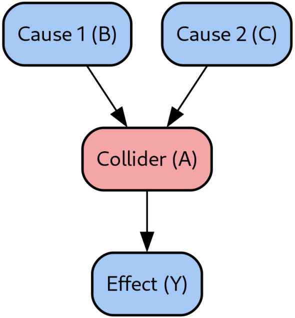






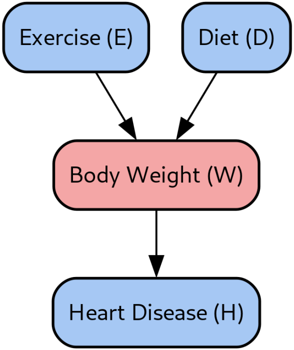
Collider

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



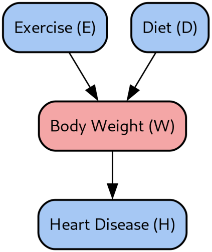
Collider: Examples

- 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"*
- Consider the variables:
  - *Diet* (D)
  - *Exercise* (E)
  - *BodyWeight* (W)
  - *HeartDisease* (D)

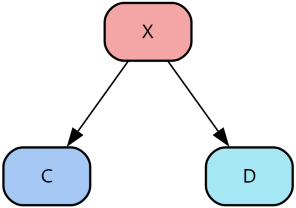


- **Without conditioning on W**
  - E and D are independent
    - E.g., knowing exercise level E doesn't inform about diet D, and vice versa
  - Collider W blocks association between E and D
- **After conditioning on W**
  - E.g., individuals with specific body weight
  - Introduce dependency between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

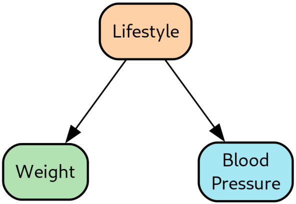
- Causal Networks
  - Variables
  - Type of Variables in Causal AI
  - ***Paths***

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

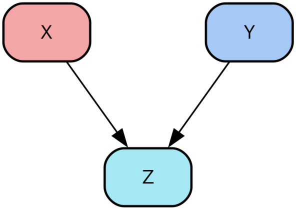


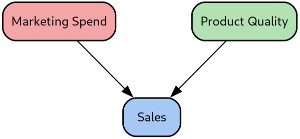
- **Example**
  - *Lifestyle* is a confounder that affects both *Weight* and *BloodPressure*
  - These outcomes may appear correlated due to shared cause



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*

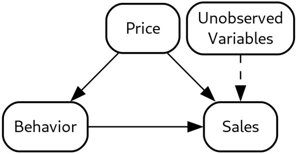




SCIENCE
ACADEMY

Path connecting unobserved variables

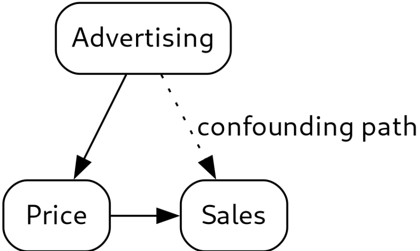
- **Unobserved variables** affect the model but we don't have a direct measure of it
- E.g., consider the causal DAG
  - A retailer does market research, expecting *Price* to influence *Sales* in a predictable way
  - A retailer sets the *Price* of a new product based on market research
  - The retailer can observe and measure *Behavior*, e.g.,
    - Discounts
    - Promotional campaign
  - There are unobserved vars that influence the model, e.g.,
    - Social media buzz
    - Word-of-mouth recommendation



SCIENCE ACADEMY

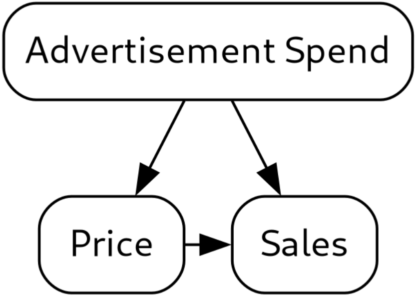
Front-door Paths in Causal Inference

- A front-door path reveals causal influence through an observable mediator
  - The causal effect flows: $A \to P \to S$
- Requirements for identifiability:
  - All confounders of $A \to P$ and $P \to S$ are observed and controlled
  - There are no back-door paths from A to S through unobserved variables
- Enables causal estimation when back-door adjustment is infeasible
- Example:
  - Advertising impacts sales through customer perception of price
  - A: Advertising, P: Price perception, S: Sales
- Pearl's front-door criterion provides a formal method for adjustment
  - Estimate $P(P|A)$, $P(S|P, A)$, and $P(A)$ from data to compute causal effect



Back-Door Paths

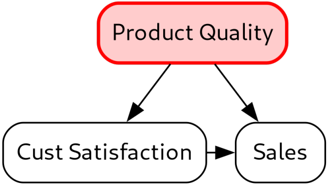
- A company wants to understand the causal effect of price on sales



- `Price` $\rightarrow$ `Sales` is the front-door path
- A confounder is `Advertising spend` 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 goes from `Sales` to `Price` via `Advertising spend`

Frontdoor and Backdoor Paths

- Question: *Will increasing our customer satisfaction increase our sales?*
  - Assume that the Causal DAG is



  - **Front-door path** (i.e., a direct causal relationship): *CustomerSatisfaction* $\rightarrow$ *Sales*
  - **Backdoor path**: *ProductQuality* is a common cause (confounder) of both *CustomerSatisfaction* and *Sales*
  - To analyze the relationship between customer satisfaction and sales, we need to:
    - Control for *ProductQuality* to close the backdoor path
    - Eliminate the confounding effect
  - In reality there are more confounding effects (e.g., price)

Counterfactuals

- A **counterfactual** describes what would have happened under a different scenario
  - *"What would the outcome have been if X had been different?"*
  - *"If kangaroos had no tails, they would topple over"*
  - *"What if we had two suppliers of our product, rather than one? Would we have more sales?"*
  - *"Would customers be more satisfied if we could ship products in one week, rather than three weeks?"*
- **Causal reasoning**:
  - Goes beyond correlation and association
  - Requires a causal model (like an SCM) 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