UNIVERSITY OF MARYLAND

MSML610: Advanced Machine Learning

Causal Inference

Instructor: GP Saggese, PhD - `gsaggese@umd.edu`

References:

- Easy:
  - Hurwitz, Thompson: Causal Artificial Intelligence: The Next Step in Effective Business AI, 2024
- Medium / Difficult
  - AIMA
  - Facuce

Causal AI

Why Causal AI?

- Causal AI
  - **Why Causal AI?**
  - Concepts in Causal AI
  - Variables
  - Paths
  - The Ladder of Causation
  - Correlation vs causation models
- Business processes around data modeling

Big data and traditional AI

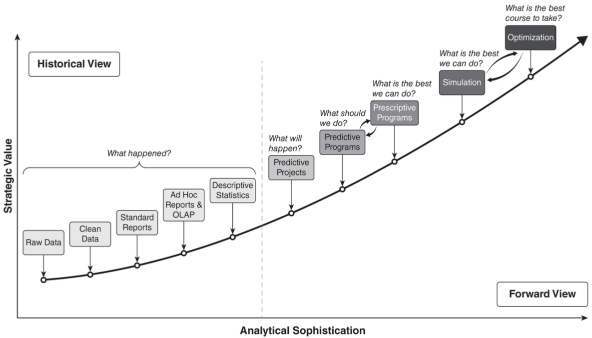
- For the past 10 years, the focus of AI and advanced analytics has been:
  - Organize and analyze massive amounts of data
  - Answer difficult problems
- Problem with traditional AI
  - Makes prediction based on observed correlations
  - Can't tell why a particular outcome occurred
- AI in decision making
  - Organizations need to understand the impact of decision making
  - E.g., what happens if a product price is reduced by 10%?
    - Will more customers buy?
    - If revenue suddenly decreases, what to do?
    - Why are customers leaving? Is it because of quality issue? Is it because an emerging competitor?

What are Data Analytics?

- Dashboards
  - Visual displays of key metrics for quick insights
  - E.g., a finance department dashboard showing quarterly revenue and expenses
- Historical reports
  - Detailed examination of past performance
  - E.g., monthly sales reports for the past fiscal year
- Collections of data
  - Aggregated and organized data sets for analysis
  - E.g., customer purchase histories stored in a CRM system
- Models
  - Statistical or mathematical representations to forecast or explain phenomena
  - E.g., a predictive model to anticipate customer churn based on behavioral data
- Descriptive statistics
  - Summary metrics like mean, median, mode, and standard deviation
  - E.g., calculating the average sales per quarter to understand performance trends

Data Analytical Sophistical Model

- Different analytics have different strategic value and sophistication



Business Question
What happened?
What will happen?
What should we do?
What's the best we can dc

Explainability

- Regulators require that if you are making decisions using ML, you should be able to defend the results of your analysis
  - E.g., decide who to hire, set up a policy
  - Organizations can:
    - Be sanctioned or fined by regulatory authorities
    - Face backlash from customers and activists
- Neural networks are "black boxes"
  - Humans can't understand how inputs are combined into a conclusion
  - Data managers cannot explain to shareholders why certain decisions were made
  - Lack of explainability and identification of bias
    - E.g., using age, race, sex as a feature can introduce bias
- Explainable AI (XAI) allow users to:
  - Comprehend
  - Trust the results by the machine

Correlation is Not Causation!

- Correlation is a statistical method for understanding relationships between data sets
  - Pros
    - Can use past outcomes to predict future outcomes by finding patterns and anomalies in data
  - Cons
    - Doesn't explain the cause of the results
    - Two variables may move together due to coincidence or a hidden factor
- Causation means changing one variable actively influences the other, which cannot be concluded from correlation alone
- Only humans can understand cause and effects in the context of data
  - Data does not understand causes and effects
  - Humans have to identify variables and relationships based on their understanding of the issues
  - Without causation, one can't make intelligent decisions
- AI needs to augment the power of humans to understand the world

Causal AI

- Causal AI solves the previous problems
- Understands the why
  - Determines cause-and-effect between variables
  - E.g., determining whether a marketing campaign increased sales
- Identify interventions
  - Identifies variables and interventions to change outcomes
  - E.g., determining which lifestyle changes can reduce blood pressure
- Predicting counterfactuals
  - Hypothesizes what could happen under different circumstances
  - E.g., estimating student performance if they had attended a different school
- Avoiding bias
  - Traditional AI is biased by training data or ignored variables
  - Causal AI helps avoid bias and ensures fairness by accounting for confounding variables
- Improving decision-making

Concepts in Causal AI

- Causal AI
  - Why Causal AI?
  - **Concepts in Causal AI**
  - Variables
  - Paths
  - The Ladder of Causation
  - Correlation vs causation models
- Business processes around data modeling

Causal DAG

- A **Causal DAG** is a Directed Acyclic Graph that represents causal relationships between variables
  - Directed: Arrows show direction of cause $\rightarrow$ effect
  - Acyclic: No feedback loops
    - Causal relationships assume a temporal order: cause happens before effect
    - A cycle would imply a variable is both a cause and effect of itself, creating a paradox
- **Structure**:
  - Nodes represent variables (e.g., income, education)
  - Edges represent causal influences (e.g., education $\rightarrow$ income)
  - E.g., education casually affects income through job skill

$$\textit{Education} \rightarrow \textit{JobSkill} \rightarrow \textit{Income}$$

- **Benefits**
  - DAGs encode *causal* rather than *associative* links
  - Enables reasoning about interventions and counterfactuals
  - Supports fair, explainable AI models
- **Limitations**:
  - Requires domain knowledge to specify structure
  - Assumes all relevant variables are included (no hidden confounders)

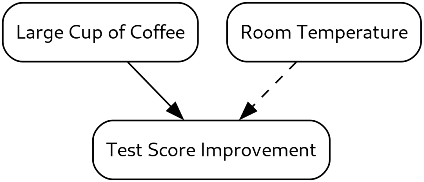
Structural Causal Model

- A Structural Causal Model (SCM) translates a graphical model / DAG into mathematical equations
  - DAGs show structure (variables and arrows)
  - SCMs add equations that define how variables interact
- **Structure**
  - Variables: $X_1, X_2, ..., X_n$ represent quantities in the system
  - Structural equations: $X_i = f_i(PA_i, U_i)$
    - f_i: function describing how X_i is determined
    - PA_i: parent variables (direct causes)
    - U_i: exogenous (external, unobserved) variables
- SCMs can be used to:
  - Explain causal relationships between variables
  - Make predictions on how relationships change if conditions changed

Structural Causal Model: example

- SCM expresses the relationship between the state of the world and how the variables interact

- 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 (independent variables)
  - E.g., *"by how much did the score test improve?"*
- Unobserved variables = not seen and more difficult to account
  - E.g., *"temperature of the room, which makes students sleepy and less alert"*



Variables

- Causal AI
  - Why Causal AI?
  - Concepts in Causal AI
  - **Variables**
  - Paths
  - The Ladder of Causation
  - Correlation vs causation models
- Business processes around data modeling

Observed vs. Unobserved Variables

- **Observed variables**
  - Aka “measurable” or “visible” variables
  - Variables directly measured or collected in a dataset
  - E.g., age, income, blood pressure, product price
- **Unobserved variables**
  - Aka “latent” or “hidden” variables
  - Variables that exist but are not measured or included in the data
  - E.g., patient’s stress level, trust in a brand, company culture
  - Unobserved variables, when ignored, can distort causal relationships creating spurious correlations or biased results, e.g.,
    - Observed: `Ice cream sales` and `Drowning rates`
    - Unobserved: `Temperature` (if not recorded)
    - Misleading conclusion without unobserved variable: ice cream causes drowning

Exogenous vs. Endogenous Variables

- **Endogenous variables**
  - Variables whose values are determined *within* the model
  - Dependent on other variables in the system
  - Represent the system's internal behavior and outcomes
  - E.g., in a model of `education` $\rightarrow$ `income`, `income` is endogenous since it is affected by education level
- **Exogenous variables**
  - Variables that originate *outside* the system being modeled
  - Not caused by other variables in the model
  - Often represent background conditions or external shocks
  - E.g., natural talent, economic policy, weather

Endogenous / Exogenous vs. Observed / Unobserved Variables

- In Structural Causal Models (SCMs) $X_i = f_i(PA_i, U_i)$
  - X_i is endogenous
  - PA_i: its parent variables (causes within the model)
  - U_i: exogenous noise term (outside causes)
  - Exogenous variables: capture randomness or unknown external factors
  - Endogenous variables: focus for prediction and intervention

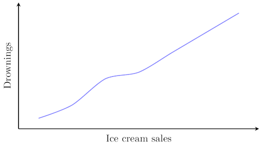
Variable Type	**Observability**	**Example**
Endogenous	Observed	Sales
Exogenous	Observed	Marketing Budget
Endogenous	Unobserved	Motivation
Exogenous	Unobserved	Macroeconomic shocks

Counterfactuals

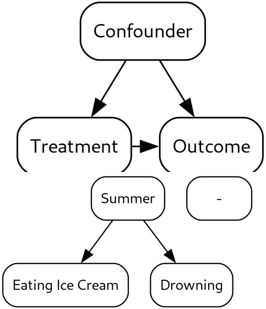
- A **counterfactual** describes what would have happened under a different scenario
  - Expressed as: *"What would the outcome have been if X had been different?"*
  - *"If kangaroos had no tails, they would topple over"*
  - *"What if we had two suppliers of our product, rather than one? Would we have 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
  - Example:
    - 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

Confounder variable

- A confounder is a variable in a causal graph that influences multiple variables
- A confounder can lead to spurious associations

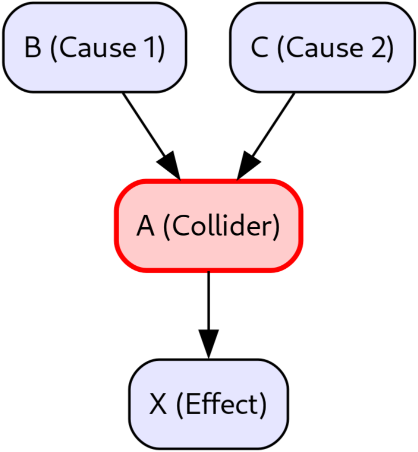


- "Eating ice cream" and "Drowning" are associated
- There is no cause-effect, since "Summer" is a confounder



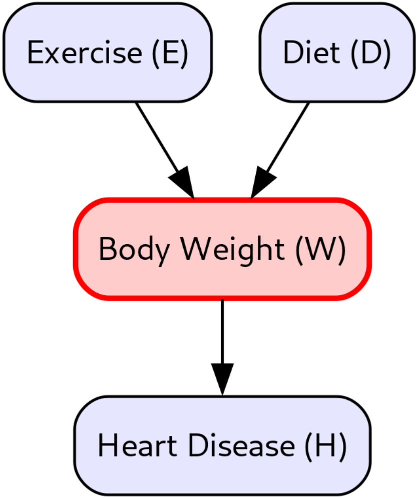
Collider

- A **collider** is a variable A with incoming edges from variables B, C in a causal DAG (i.e., influenced by multiple variables)
- A collider complicates understanding relationships between variables B, C and those it influences, X



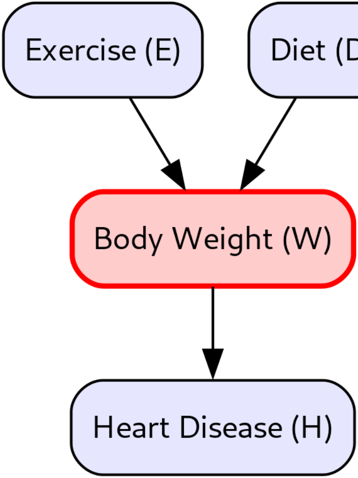
Collider: examples

- Study the relationship between exercise E and heart disease H
  - Diet D and exercise E influence body weight W
  - Body weight W influences heart disease H
  - Body weight W 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"
- Without conditioning on W
  - E and D are independent of each other
  - I.e., knowing someone's exercise level E doesn't give information about diet D, and vice versa
  - The collider W blocks any association between E and D
- After conditioning on W (e.g., looking for individuals with specific body weight)
  - We introduce a dependency between E and D
  - Since W is fixed, any change in E must be balanced by a change in D to maintain the same body weight, inducing a spurious correlation between E and D

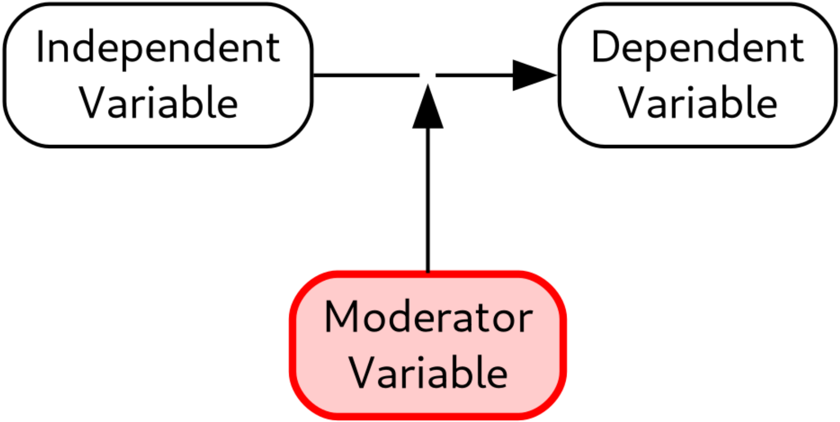


Mediator variable

- A **mediator variable** lies on the causal path between a treatment and an outcome
- E.g., when studying the effect of a training program (X) on employee productivity (Y), job satisfaction could be a mediator variable

Moderator variable

- **Moderator variable** affects the strength or direction between two other variables



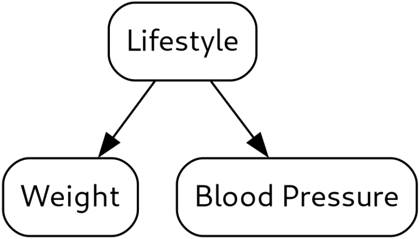
- E.g., when studying the relationship between stress X and job performance Y the level of social support an individual receives M could be a moderator

Paths

- Causal AI
  - Why Causal AI?
  - Concepts in Causal AI
  - Variables
  - **Paths**
  - The Ladder of Causation
  - Correlation vs causation models
- Business processes around data modeling

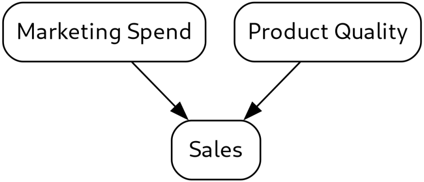
Fork

- Multiple arrows fork away from a node
- E.g.,
  - `Lifestyle` influences both `Weight` and `Blood Pressure`



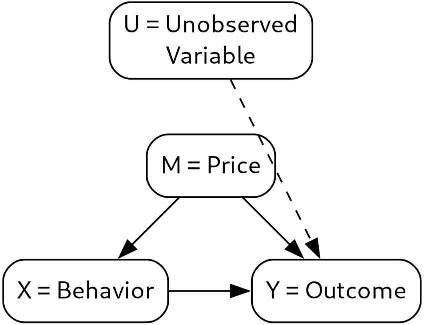
Inverted fork

- Two more paths converge on a node (aka “collider”)
- E.g.,
  - `Marketing spend` and `Product quality` both influence `Sales`



Path connecting an unobserved var

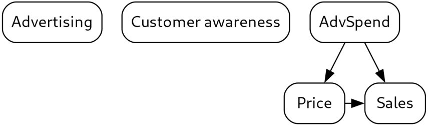
- Unobserved vars affect the model but we don't have a direct measure of it



- E.g., consider the causal DAG
  - A retailer sets the `Price` of a new product based on market research, expecting price to influence `Sales` in a predictable way
  - `Behavior`: the retailer can observe and measure discounts, promotional campaign, etc
  - Unobserved vars can be `Social media buzz` or `Word-of-mouth recommendations`

Front-door paths

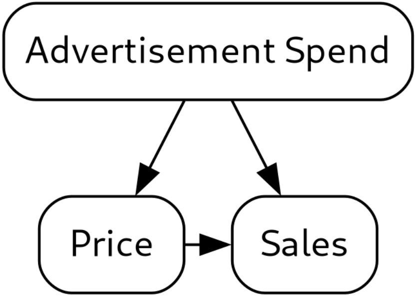
- An independent variable influences an outcome through an intermediate variable
- The path is observable and can be used to identify the causal effect



  - All confounders are observed and controlled for
  - There are no unobserved variables

Back-door paths

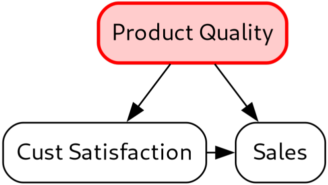
- A company wants to understand the causal effect of price on sales



- `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)

The Ladder of Causation

- Causal AI
  - Why Causal AI?
  - Concepts in Causal AI
  - Variables
  - Paths
  - **The Ladder of Causation**
  - Correlation vs causation models
- Business processes around data modeling

Causal AI vs Traditional AI

- *"The next revolution of data science is the science of interpreting reality, not of summarizing data"* (Judea Pearl)
- The current approach of AI uses statistics and ML to analyze data, identifying patterns and anomalies to make predictions
    - Models depend on the quality of the data
    - Biased or unclean data $\implies$ poor model

The Ladder of Causation

- Pearl provided a 3-layer framework for understanding causality

Level	**Symbol**	**Typical Activity**
1. Association	$\Pr(Y\|X)$	Seeing
2. Intervention	$\Pr(Y\|do(X), Z)$	Doing, Intervening
3. Counterfactuals	$\Pr(Y_X\|x', y')$	Imagining, Retrospection

Rung 1: Association

- $\Pr(Y|X)$: how would seeing X change our belief in Y?
- It is just "passive observation" to determine if two things are related
  - Traditional AI and ML is based on this
- E.g.,
  - "The tree has green leaves during spring"
  - "What does a symptom tell me about a disease?"
  - "What does a survey tell us about the election results?"

Rung 2: Intervention

- $\Pr(Y|do(X), Z)$: what happens to Y if we do X?
- Not just passively observing but understanding the impact of change
  - Association is purely observational and is not model based
  - E.g., "tree has green leaves" vs "spring makes the tree leaves turn green"
  - Interventions are about "doing something" and require a causal model
- E.g.,
  - "Why did the headache go away?"
    - "Because the pain reliever" or "Because you ate food after skipping lunch"
  - "What if I take aspirin, will my headache be cured?"
  - "What if we ban sodas?"

Level 3: Counterfactuals

- $\Pr(Y_X|x', y')$: was X that caused Y?
- Determine what would occur if a new condition were applied to a situation
  - "Imaging what will happen if facts were different"
  - Counterfactuals establish causal relationships
  - Predicting an outcome is the highest form of reasoning, since it requires to understand relationships between cause and effect
- E.g.,
  - "What if I had acted differently?"
  - "What if I had not been drinking sodas for 2 years?"
  - Scientific experiments: "What if we give a child an adult dose of a drug?"
  - Litigation: "What would the jury conclude?"
  - Marketing: "Why did my marketing campaign fail to generate sales?"

Correlation vs causation models

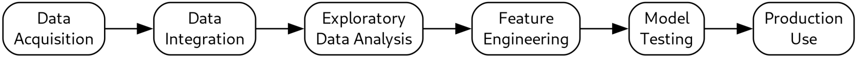
- Causal AI
    - Why Causal AI?
    - Concepts in Causal AI
    - Variables
    - Paths
    - The Ladder of Causation
    - **Correlation vs causation models**
- Business processes around data modeling

Correlation vs causation model

- Correlation = identify how variables are related to each other
- Causality = determine whether one variable causes another variable
- Both:
  - Accept inputs and transform them
  - Identify how variables are related to each other
- Correlation-based AI is best when there is abundant historical and observational data
- Causal-based AI first creates a business-focused model before integrating data

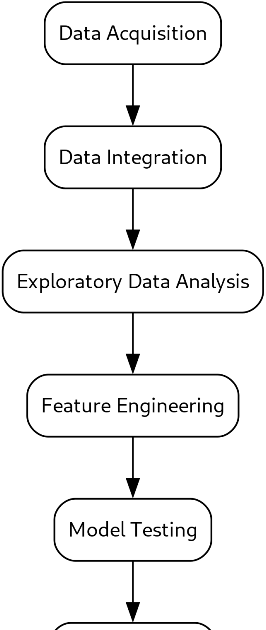
Correlation-based model process

- Correlation-based AI is "data first"
- The more data collected the better
- Many AI projects fail because
  - Cultural and organizational issues
  - Models are opaque
  - Lack explainability
  - Spurious correlations



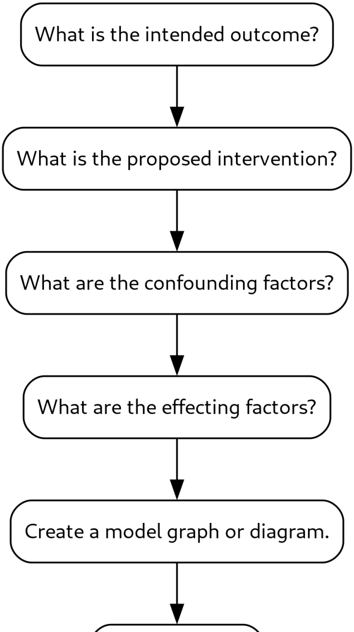
Correlation-based model process

- Correlation-based AI is “data first”
- The more data collected the better
- Many AI projects fail because
  - Cultural and organizational issues
  - Models are opaque
  - Lack explainability
  - Spurious correlations



Causation-based model process

- Causal-based AI is “model first”
- Understand business question before ingest and transform the data

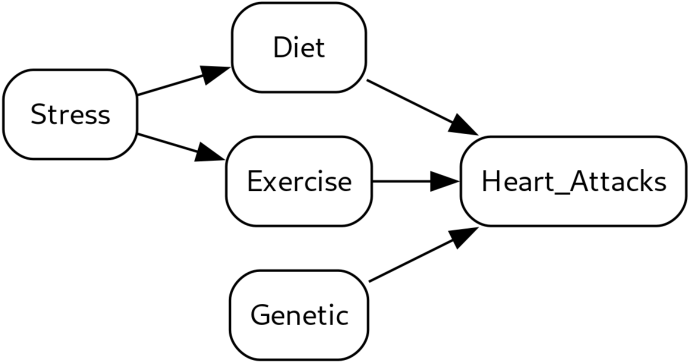


Building a DAG

- **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**:
  - Models are continuously updated with new variables and insights
- **Modeling as a Communication Tool**:
  - A shared language that bridges gaps between technical and non-technical team members
- **Unobservable Variables**:
  - Supports inclusion of variables not empirically observed but known to exist
  - E.g., trust or competitor activity can be modeled despite lack of direct data

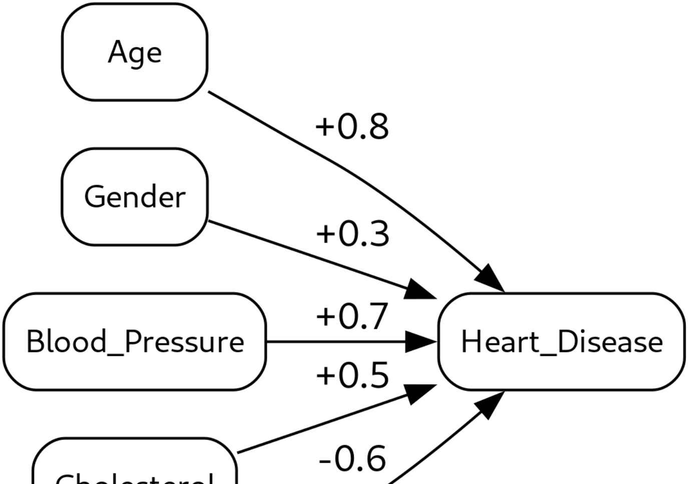
Heart attack: example

- 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

- Weights can be assigned to paths to represent the strength of the causal relationship
  - Weights can be estimated using statistical methods
- Sign represents the direction



Business processes around data modeling

- Causal AI
- **Business processes around data modeling**
  - Modeling processes
  - Roles

Modeling processes

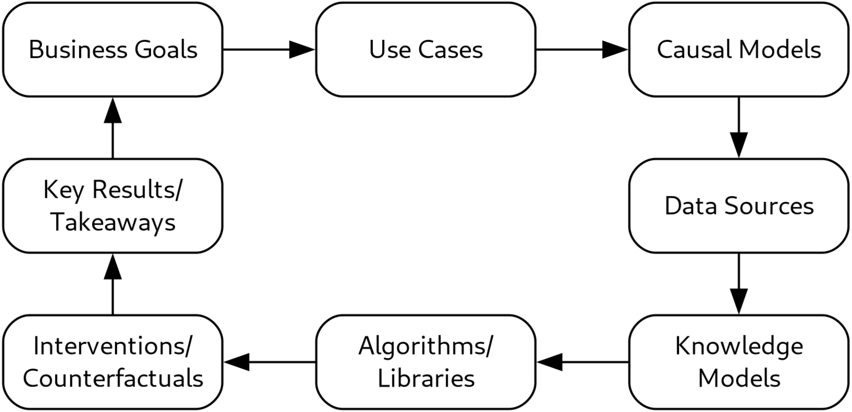
- Causal AI
- Business processes around data modeling
    - **Modeling processes**
    - Roles

Digital Transformation

- Integration of Digital Technology
    - Embed digital tools (AI, cloud, IoT, automation) into all business areas to enhance efficiency and value delivery
- Cultural & Organizational Change
    - Encourage innovation, agility, and a data-driven mindset to adapt to new digital workflows and business models
- Customer-Centric Approach
    - Use digital solutions (e.g., personalized experiences, AI-driven insights) to enhance customer engagement and satisfaction
- Process Automation & Optimization
    - Streamline operations through AI, robots, and analytics to reduce costs and improve decision-making
- Data-Driven Decision Making
    - Leverage big data, machine learning, and real-time analytics to make smarter, faster, and more strategic business decisions

Causal modeling process

- The overall modeling process looks like:



Step 1: What are the intended outcomes?

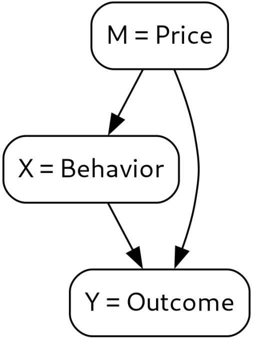
- What is the process/environment we are interested in analyzing?
- What will happen if a course of action is (or not) taken?
- What outcomes are positive, negative, unacceptable, optimal?
- What are the possible / feasible interventions?
- What confounding factors might be correlated with outcomes and treatments?
- What factors exist but we cannot accurately measure?
- What related data sets can be combine / leverage?

Step 2: What are the proposed interventions?

- We will make reference to a use case of customer marketing
- Can we introduce a new product?
- Should we buy one or more competitors?
- Does bundling multiple products improve sales?
- Does bundling multiple products inhibit long-term sales?
- Should advertisement focus on quality of our product vs other options?
- Should we divest the product line?
- Should we discontinue the product?
- Should we add more variations of the same product?

Marketing example: price intervention

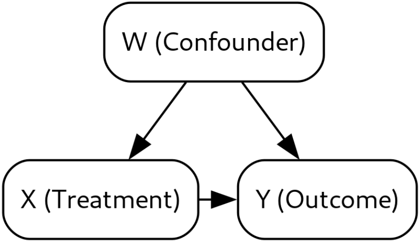
- Assume price is our intervention and the Causal DAG is:



- What happens to sales when we change the price often?
- What pricing interventions are optimal?
  - Should we increase the price and how much?
  - Should we decrease the price and how much?
  - Should price change in one-time or over time?
- Should we adopt a dynamic pricing model?
- Should we develop individual pricing model for each customer?

Step 3: What are the confounding factors?

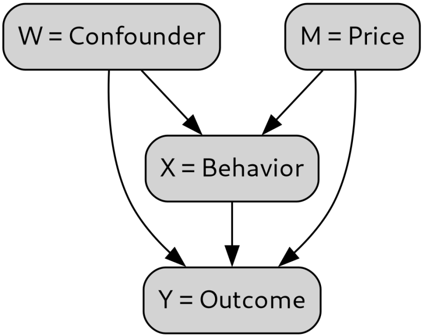
- There can be a variable W that affects both X and Y



- E.g.,
  - Competitive offers
  - Distance to store
  - Amount of product
  - Time to consume product
- A confounder can:
  - Make it difficult to understand the relationship between variables
  - Mute or inflate a relationship

Marketing example: effect of confounder

- E.g.,
  - Intervention = a marketing campaign to sell winter jackets
  - A confounding variable can be “running the campaign in the middle of the winter, after customers have already purchased their jackets”
  - A confounding var can be “a warm winter”



Step 4: What are the factors creating the effects and changes?

- Total causal effect
  - Effect of all factors in the environment or model that modify the outcome
- Direct effect
  - Effect introduced through an intervention
- Indirect effect
  - Effect introduced by environment or it is a byproduct of the intervention in a way that was not planned

Step 5: Build Causal DAG

- Causal models
  - Simplify complex systems without losing key relationships
  - Focus on essential variables and their interactions
- Visual models (e.g., DAGs) help abstract complexity into interpretable formats
  - Highlight direction and strength of influence between variables
- Simplicity
  - Aids communication between technical and non-technical stakeholders
  - Promotes shared understanding and collaborative refinement
  - Reduces cognitive overload by excluding irrelevant details or noise
  - Guides data collection by identifying the most impactful variables
  - Supports hypothesis testing through counterfactual and intervention scenarios
- Balance is key
  - Too much simplicity loses insight
  - Too much complexity loses clarity

Step 7: Data Acquisition and Integration

- You can use the data collected for correlation-based ML
- Data collection can be done specifically for causal AI
  - Treating, conditioning, transforming data

Step 8: Model Modification

- Once the DAG is designed, use software packages to build models
  - Refine the initial DAG and causal model to reduce bias and improve reliability
  - Clarify variables as confounders, mediators, or outcomes
- Avoid common pitfalls:
  - Do not control for mediators or effects, which can distort results
  - Control for direct and indirect confounders to prevent biased estimates
- Implementation tools:
  - Use libraries to operationalize models
  - Test models against technical and business objectives
- d-separation:
  - Identifies conditional independence relationships in a DAG
  - Determines necessary controls to isolate causal effects
  - Based on Judea Pearl's definition: independence = separation in the graph
  - Prevent unintentional inclusion of bias
  - Ensure causal assumptions align with data and domain logic
  - Improve model interpretability and predictive power
- Goal: Ensure final models are technically valid and business-relevant before proceeding to data transformation and testing stages

Step 10: Data Transformation

- Prepare the data to match the refined causal model
  - Clean, normalize, and align data with model assumptions
- Transformations include:
  - Mapping observed variables to nodes in the DAG
  - Encoding categorical variables appropriately
  - Handling missing or unobserved data (e.g., imputation or exclusion)
  - Normalizing or scaling values to align with model expectations
- Control for bias and confounding:
  - Apply methods like propensity score matching or stratification
  - Exclude or adjust for variables that introduce bias per d-separation insights
- Goal: Ensure the data structure supports causal estimation
  - Consistent with assumptions made in model refinement
  - Aligned with theoretical model
  - Fit for downstream tasks like estimation, inference, and simulation

Step 11: Preparing for Deployment in Business

- Operationalize the causal model within a business context
  - Transition from experimentation to integration with decision-making processes
  - Validate the model against real-world business data and outcomes
  - Ensure stakeholders understand and trust the causal logic and assumptions
- Model packaging:
  - Develop user-friendly interfaces or dashboards for business users
  - Automate data pipelines for timely updates and monitoring
  - Embed the model within decision-support tools or policy engines
- Governance and monitoring:
  - Establish metrics for performance tracking and drift detection
  - Create feedback loops to refine and improve models post-deployment
- Documentation and training:
  - Provide clear model documentation for auditors and users
  - Train stakeholders on interpreting causal results and making informed decisions
- Goal: A deployable causal AI solution that supports strategic decisions and delivers measurable business value

Roles

- Causal AI
- Business processes around data modeling
  - Modeling processes
  - **Roles**

Why ML / AI projects fail?

- AI projects fail because they approach problems only from a ML perspective
  - Data scientists:
    - Use data to create models
    - Work in isolation from business users and internal data teams
  - Black-box models unable to produce solutions to real-world problems

How to make ML / AI project succeed

1. Create a hybrid team
    - Organizations are complex in structure and offerings
    - A single group lacks the knowledge / skills to tackle difficult problems
    - Need an hybrid team:
        - Represents all aspects of the business problem
        - Uses a collaborative framework
        - Communicates with a single language (e.g., through DAGs)
        - Team size depends on company size and project complexity
2. Meet regularly to ensure project continuity
3. Find an executive sponsor for the project
    - Someone who understands the project's goals and potential
4. Initial pilot
    - Small team for a targeted problem
    - Demonstrate the merit of the AI approach

Roles in hybrid teams

Role	Responsibilities
Business Strategists	Align modeling with business goals
	Sponsor projects
	Communicate insights to stakeholders
Subject-Matter Experts	Provide domain expertise
	Identify relevant variables and assumptions
	Validate DAGs
Data Experts	Source and clean data
	Map data to model variables
	Handle missing values
Data Scientists	Construct and validate DAGs
	Apply causal inference methods
	Simulate decisions
Software Developers	Build tools and interfaces
	Create data pipelines
IT Professionals	Provide infrastructure and governance
	Ensure model execution
	Integrate with enterprise systems
Project Managers	Coordinate collaboration and timelines
	Manage documentation
	Ensure alignment with strategic goals

Steps for a hybrid team project

- **Establish a Phased, Collaborative Approach**
  - Align strategic goals with technical efforts
  - Emphasize early stakeholder engagement and shared ownership
- **Strategic Kickoff Meeting**
  - Unite business, technical, and operational roles
  - Clarify problems, outcomes, responsibilities, success metrics
- **Define Team Goals**
  - Use SMART objectives aligned with business strategy
  - Focus on outcomes and supported decisions, not tools
- **Target a Project**
  - Choose a bounded, feasible, high-value use case
  - Prioritize early wins for trust and momentum
- **Define the Hypothesis**
  - Translate business problems into testable causal assumptions
  - Build a preliminary DAG with experts' input
- **Incremental Model Development**
  - Build the model in small, reviewable stages
  - Iterate with regular feedback, refining scope and variables
- **Embrace Iteration and Continuous Refinement**
  - Keep progress collaborative and aligned with business needs
  - Add complexity gradually to manage risk and enhance understanding

The importance of explainability

- Managers rely on AI systems to automate decision-making
  - Decisions rely on complex algorithms and data
- Understanding AI-based models is growing in importance
  - How do ML models make decisions?
  - How can they be trusted?
  - Are they biased?
- Management often faces demands to prove code validity
  - Loss of trust, regulation violations, fines, additional development costs, lawsuits
  - E.g., a false negative in a medical screening for cancer
- Well-designed AI systems must foster trust, transparency, and user confidence
- **Humans in the loop**
  - AI systems lack true reasoning and contextual understanding
  - Human involvement ensures interpretation and context are considered

Techniques for interpretability

- Local Interpretable Model-agnostic Explanations (LIME)
  - Focuses on a single prediction (local fidelity)
  - Approximates the model locally with an interpretable model
  - Perturbs input data and observes changes in predictions
- Partial Dependence Plots (PDP)
  - Show the marginal effect of one or two features on the predicted outcome
  - Vary the value of one feature while keeping others constant
  - Plot feature values against the average predicted outcome
- Individual Conditional Expectation (ICE)
  - Show the relationship between a feature and the prediction for individual instances
- SHapley Additive exPlanations (SHAP)
  - Quantify the contribution of each feature to a specific prediction

Causal AI in interpretable AI

- Causal AI helps understand causes, effects, and potential solutions
  - Uses causal graphical models to present variables, relationships, and strengths
  - Counterfactual analysis predicts outcomes of different actions or policies before deployment
  - Model output is understandable to humans and non-experts
  - Removing confounding variables prevents skewed causal estimates due to hidden influences
  - Hybrid teams (technical + domain experts) enhance context awareness and reduce blind spots

The iterative causal AI pipeline

- Similar to the DevOps infinity loop, the causal AI pipeline has 8 stages, which are circular and iterative

1. Define business objectives
   - Identify goals and key performance indicators (KPI)
2. Model development
   - Establish causal relationships among variables
   - Develop causal AI model using graphical models, potential outcomes, do-calculus
3. Data identification and collection
   - Gather data with causal relationships and potential confounders
4. Model validation
   - Test and validate model to ensure it provides accurate causal insights, e.g.,
     - Measure performance against KPIs
     - Assess model generalizability
5. Deployment / production
   - Integrate validated AI model into existing systems for decision-making
6. Monitor and evaluate
   - Track performance of deployed model
   - Assess impact on decision-making process
7. Update and iterate
   - Refine model based on real-time performance evaluation

Define business objectives

- Use one of the systems
  - SMART goals
  - MoSCoW method
  - Kano model
- Goals
  - Establish a shared understanding of the project's goals among stakeholders, business leaders, data scientists, and exports
  - Enhances clarity, focus, and motivation
  - Improves performance tracking and outcome evaluation
  - Facilitates effective communication and collaboration

SMART goals

- SMART is an acronym for Specific, Measurable, Achievable, Relevant, Time-bound
- **Specific**
  - Clearly defines what is to be achieved
  - E.g., "Increase website traffic by 15%"
- **Measurable**
  - Includes criteria to track progress and success
  - E.g., "Track visits using Google Analytics"
- **Achievable**
  - Realistic based on resources and constraints
  - E.g., "Based on historical growth rates, 15% is attainable"
- **Relevant**
  - Aligns with broader organizational or personal objectives
  - E.g., "Traffic increase supports lead generation goals"
- **Time-bound**
  - Specifies a deadline or timeframe
  - E.g., "Achieve by end of Q3"

MoSCoW method

- Framework for prioritization of features to:
    - Focus on critical features
    - Ensure stakeholders are aligned on priorities
    - Support agile and iterative development by delivering value in phases

Priority	Description	Example
Must-have	Essential requirements	A login system for a banking app
Should-have	Important but not critical features	Multi-factor authentication to enhance security
Could-have	Enhance user experience but are not essential	Dark mode for a mobile app
Won't-have	Excluded features due to low priority	AI-powered chat assistant

Kano model

- Framework for product development and customer satisfaction analysis to prioritize features

Category	Description	Examples
Must-have	Expected features; absence causes dissatisfaction	Keyboard and screen on a laptop Calling and texting on a cellphone
Performance	The better, the more satisfied customers are	Faster battery charging Longer battery life
Delighters	Unexpected features that wow customers and build loyalty	Free cloud storage with a laptop Novel camera
Indifferent	Features that do not affect customer satisfaction	5 power adapter options for laptop Color of phone internals
Reverse features	Features liked by some but disliked by others	Touchscreen instead of touchpad Excessive automation

Decision matrix

- Tool to evaluate options based on criteria for objective comparison
- **Goals**
  - Eliminate biases by quantifying decisions
  - Clarify priorities
  - Enable comparison between options
- **Approach**
  1. List alternatives
     - E.g., laptop A, laptop B
  2. Define criteria
     - E.g., price, performance, battery life
  3. Assign weight to each criterion
  4. Score each option
     - Rate 1-5 for each criterion
  5. Calculate weighted score
     - Weighted sum of scores for each alternative
  6. Compare totals
     - E.g., choose the alternative with the highest score

Key Performance Indicators (KPIs): General Overview

- **What are KPIs?**
  - Quantitative metrics that reflect how well an individual, team, or organization is achieving key objectives
- **Purpose of KPIs**
  - Provide focus for strategic and operational improvement
  - Serve as benchmarks for progress and performance evaluation
- **Categories of KPIs**
  - **Financial KPIs**: Revenue, profit margin, cost per acquisition
  - **Customer KPIs**: Customer satisfaction, churn rate
  - **Process KPIs**: Efficiency, cycle time, error rate
  - **People KPIs**: Employee engagement, turnover rate, training hours
- **Characteristics of Effective KPIs**
  - Specific and clearly defined
  - Measurable with reliable data
  - Achievable and realistic
  - Relevant to the business goals
  - Time-bound for periodic evaluation
- **Developing KPIs**
  - Align with strategic goals
  - Involve key stakeholders in the design process

The Importance of Synthetic Data in Causal AI

- Synthetic data is artificially generated to simulate real-world data properties
  - Created using algorithms or simulations based on real data distributions
- **Key Benefits**
  - Enhances data availability when real data is limited, biased, or sensitive
  - Enables experimentation without risking privacy breaches
  - Protects identities by not using real personal data
  - Helps comply with data privacy regulations like GDPR and HIPAA
  - Allows rebalancing of underrepresented groups or scenarios
  - Facilitates fairer model training and evaluation
- **Use in Causal Modeling**
  - Simulates counterfactual scenarios to test causal hypotheses
  - Assists in validating causal assumptions and interventions
- **Challenges**
  - Ensuring synthetic data accurately reflects key patterns and relationships
  - Avoiding distortion of causal dependencies or overfitting to synthetic features
- **Best Practices**
  - Combine with real data for validation
  - Use domain knowledge to guide synthetic data generation and use

Digital twin

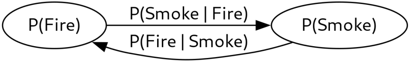
- A digital twin is a virtual representation of a physical system or process created using data and software
  - Used to simulate and identify potential problems under different conditions
  - Used to generate synthetic data

The future of Causal AI

- Causal AI:
  - Is moving out of academia into the commercial world
  - Is a departure from the 2000 approach of a purely data-driven AI systems
  - Is a reflection of the reality of how humans think, analyze, and make decisions and how the real world works
- Causal, traditional AI, deep learning, and generative techniques will merge

Causal networks

- Bayesian networks represent the joint distribution function, independently of how the nodes are ordered



- E.g., a Bayesian network with *Fire* and *Smoke*, which are dependent
  - We can specify it as $Fire \rightarrow Smoke$
    - We need $\Pr(Fire)$ and then $\Pr(Smoke|Fire)$, to compute $\Pr(Fire, Smoke)$
  - The same distribution can also be represented as $Smoke \rightarrow Fire$
    - Then we need $\Pr(Smoke)$ and $\Pr(Fire|Smoke)$
  - These two networks are equivalent and convey the same information

- There is an **asymmetry in nature** that we need to capture
  - Extinguishing the fire will stop the smoke
  - Clearing the smoke won't affect the fire

Causal (Bayesian) networks

- Causal networks are Bayesian networks that forbids any variable ordering that is not causal
- Instead of using probabilistic reasoning (e.g., "are *Smoke* and *Fire* variables correlated?"), we use judgement rooted in how nature works (e.g., "what causes what, *Smoke* or *Fire*?")
- "Dependency in nature" is similar to assignment in programming language
  - E.g., it's like if nature assigns a value to *Smoke* based on the value of *Fire*:

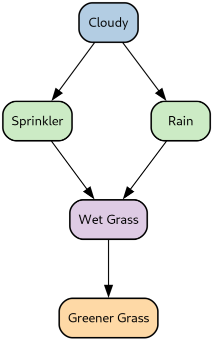
$$Smoke = f(Fire)$$

- A structural equation describes a stable mechanism in nature that remain invariant to changes in the environment and measurements

$$x_i = f(x_j) \iff X_j \to X_i$$

- We will see that "intervention" action affect a causal network only locally

Structural equation: sprinkler example



- Consider the Sprinkler example
- The joint distribution of the five variables can be expressed as a product of conditional distributions according to the causal DAG topology

$$\Pr(c, r, s, w, g) = \Pr(c) \Pr(r|c) \Pr(s|c) \Pr(w|r, s) \Pr($$

- The structural equations for this model look like:

$$\begin{aligned} C &= f_C(U_C) \\ R &= f_R(C, U_R) \\ S &= f_S(C, U_S) \\ W &= f_W(R, S, U_W) \\ G &= f_G(W, U_G) \end{aligned}$$

- The unmodeled variables U-variables represent error terms
  - E.g., U_W is another source of wetness besides *Sprinkler* and *Rain* (e.g., *MorningDew*)

Intervention in structural equations and joint probability

- The structural equations:
  - Model the system
  - Allow to predict how interventions will affect the system, which is not easily represented by the joint distribution
- The intervention $do(X_j = x_j)$ causes $X_j = f_j(Parents(X_j), U_J)$ to become $X_j = x_j$ in the structural equations
  - The causal network is "mutilated" by removing the edge
  - Then the joint distribution can be derived from the causal network
- Multiple interventions are the superposition of the individual interventions
  - In practice we can

Adjustment Formula in Causal Networks

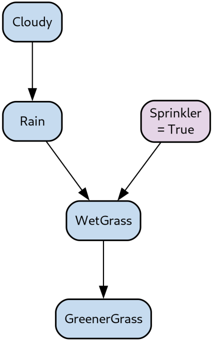
- Estimate the effect of an intervention $do(X_j = x_{jk})$ on another variable X_i in a causal network
- **The Basic Adjustment Formula**
  - Derived from the modified joint distribution after intervention:

$$\Pr(X_i = x_i | do(X_j = x_{jk})) = \sum_{\text{parents}(X_j)} \Pr(x_i | x_{jk}, \text{parents}(X_j)) \Pr(\text{parents}(X_j))$$

  - Omits the causal mechanism for X_j and treats it as fixed
- **Interpretation**
  - Weighted average of the effect of X_j and its parents on X_i
  - Weights are given by the prior probabilities of the parent values
- **Causal Inference via Graph Surgery**
  - "Mutilate" the network by cutting incoming edges to X_j
  - Replace X_j's structural equation with $X_j = x_{jk}$
- **E.g., (Sprinkler Network)**
  - To evaluate $\Pr(GreenerGrass | do(Sprinkler = true))$:

Intervention and Structural Equations: Sprinkler Example

- We "intervene" and turn the sprinkler on
- This is modeled in Do-calculus as $do(Sprinkler = T)$
- Now the sprinkler variable s is not dependent on whether it's a cloudy day c
- The causal network is "mutilated"



- The structural equations after the intervention become:

$$
\begin{aligned}
C &= f_C(U_C) \\
R &= f_R(C, U_R) \\
S &= True \\
W &= f_W(R, S, U_W) \\
G &= f_G(W, U_G)
\end{aligned}
$$

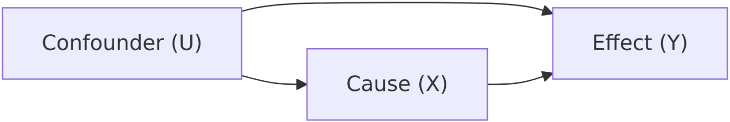
- Then the $\Pr(s|c) = 1$ and $\Pr(w|r, s) = \Pr(w|r, s = T)$ and the joint probability becomes:

Intervention vs observation

- Intervention "breaks" the normal causal link between *Weather* and *Sprinkler*
  - From the causal graph we know that there is no influence of *Sprinkler* on *Weather*
- There is a difference between $do(Sprinkler = T)$ (intervention) and $Sprinkler = T$ (observation)
  - If we observe that the sprinkler is on:
    - We can deduce that it's less likely that the weather is cloudy
  - If we turn on the sprinkler:
    - The weather is not affected and so the probability of cloudy is unaffected
-
- A causal network can predict the effect of an intervention using the adjustment formula
- The problem is that it requires accurate knowledge of the conditional distributions of the model
- E.g., in the Sprinkler example, why does someone turn on the sprinkler? Maybe they check the weather, but how do they make their decision?

Problem

- In experimental settings, randomization ensures that X is not confounded so that we can directly estimate its causal effect on Y
- Problem:
  - In observational data, we don't have randomization and confounding variables can create spurious correlations between X and Y
  - Consider the causal graph



  where U is a common cause (confounder) of both X and Y
- Solution:
  - Controlling for a set of Z that satisfies the backdoor criterion, we can estimate the causal effect of variable X on Y
  - This allows to "simulate" a randomized experiment
-
- We want to estimate the effect of *Sprinkler* on *GreenerGrass*

Backdoor path

- Z blocks all backdoor paths from X to Y
- A backdoor path is any path from X to Y that starts with an arrow pointing into X. These paths create confounding relationships that can bias the estimate of X's effect on Y

Backdoor criteria: condition

- Variables Z satisfy the backdoor criterion for X and Y in a causal network if:
  1. No element of Z is a descendant of X
     - Z do not capture the effect of X on Y through any causal pathway
  2. Z "blocks" every path from X to Y
- The idea is to estimate the causal effect of X on Y without confounding relationships, by controlling variables Z satisfying the backdoor criterion
- Then we can use non-experimental (i.e., observational data), assigning X randomly