CSDS 452 Causality and Machine Learning

Lecture 3: Structural Causal Model

Instructor: Jing Ma

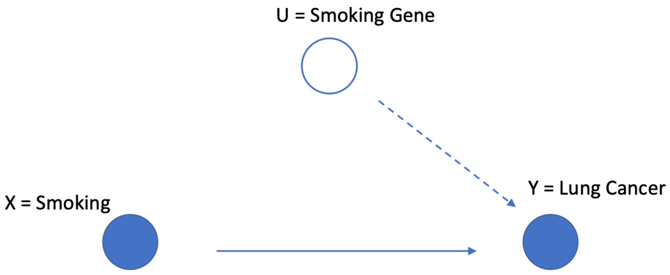
Fall 2024, CDS@CWRU

Outline

- Introduction to Graphical Models
  - Undirected graphical models
  - Directed graphical models
- Structural Causal Model
  - Causal graph
  - Structural equations
  - Intervention
  - Backdoor adjustment

Recap: Frameworks in Causal Inference

- Structural Causal Model
  - Based on graphical models
  - Causal graph + structural equations




Judea Pearl

Reference books:
- Pearl J. Causality[M]. Cambridge university press, 2009.
- Pearl J, Mackenzie D. The book of why: the new science of cause and effect[M]. Basic books, 2018.

Recap: Frameworks in Causal Inference

- Potential Outcome Framework (Neyman–Rubin causal model)
  - An approach to the statistical analysis of cause and effect based on the framework of potential outcomes

Jerzy Neyman

Donald B. Rubin

Reference Book:
Guido Imbens & Donald Rubin (2015). Causal Inference for Statistics, Social, and Biomedical Sciences: An Introduction. Cambridge: Cambridge University Press.

Recap: Identification and Estimation

- Two components in learning causality
  - (1) Identification
  - (2) Estimation, inference

Underlying Causality

Identifying assumptions ↓ ↑ Identification

Population distribution

Sampling ↓ ↑ Estimation, inference

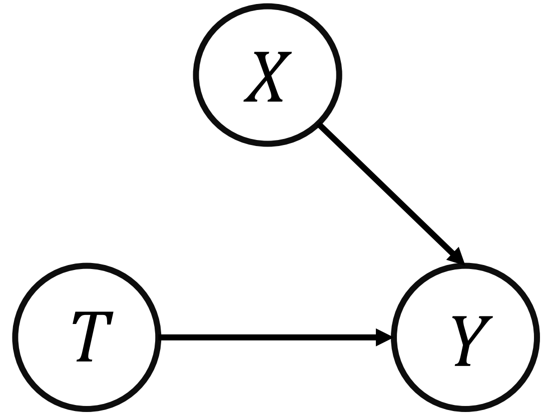
Observational data

Recap: Exchangeablibity

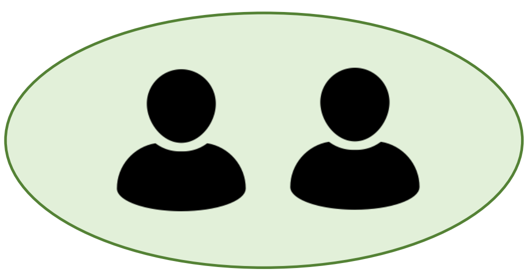
- $\left(Y(1), Y(0)\right) \perp\!\!\!\perp T$

Caution!

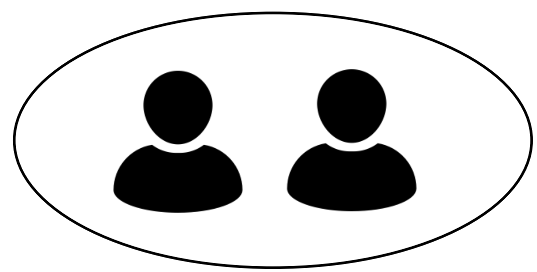
$\left(Y(1), Y(0)\right) \perp\!\!\!\perp T$ is different from $Y \perp\!\!\!\perp T$



Treatment group $T = 1$



Control group $T = 0$



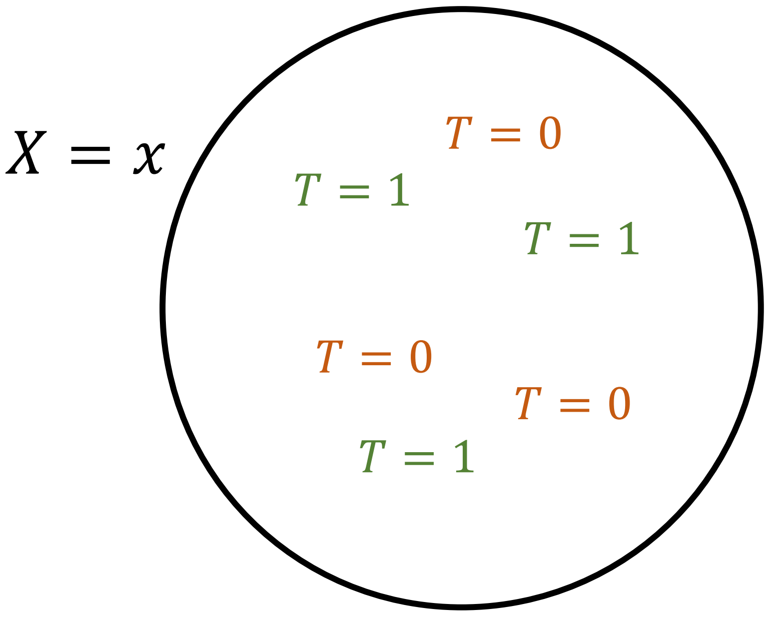
$$E[Y(1)] = E[Y(1)|T = 1] = E[Y(1)|T = 0]$$
$$E[Y(0)] = E[Y(0)|T = 1] = E[Y(0)|T = 0]$$

Treatment group and control group are comparable (“exchangeable”)

Recap: Positivity / Overlap

- For all values of $X = x$ with $P(X = x) > 0$ in the population of interest:

$$P(T = t | X = x) > 0$$



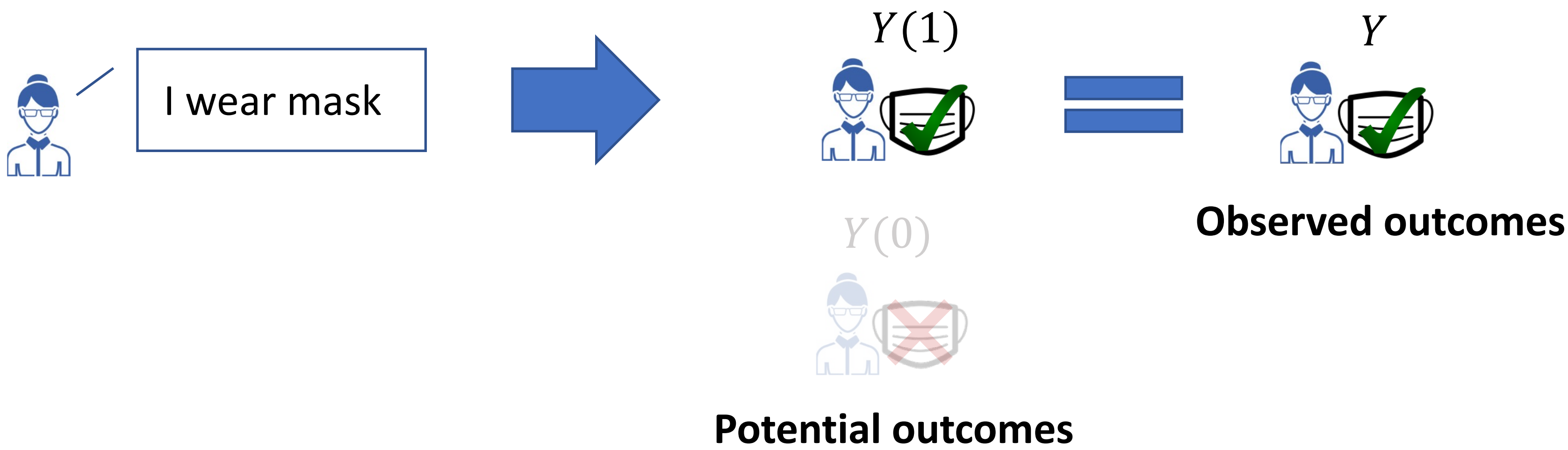
Recap: Tradeoff between Positivity and Unconfoundedness

- Conditioning on more covariates
  - higher chance of satisfying unconfoundedness
  - higher chance of violating positivity
- Example:
  - Conditioning on 1 dimension – 50% overlap
  - Conditioning on 2 dimension – 25% overlap
  - …

  Related to the Curse of dimensionality

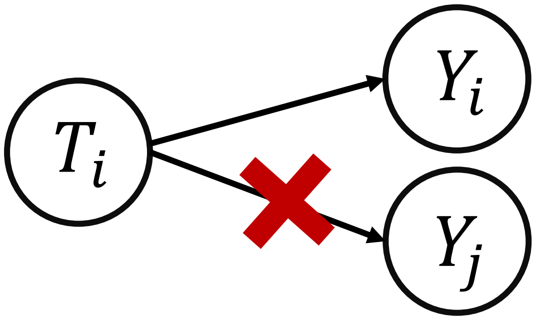
Recap: Consistency

- $Y = Y(t)$ when $T = t$



Recap: Stable Unit Treatment Value Assumption (SUTVA)

- The potential outcomes for any unit do not vary with the treatments assigned to other units.
  - No **interference**



- For each unit, there are no different forms or versions of each treatment level, which lead to different potential outcomes.
  - E.g., when treatment is "take a surgery", this surgery is operated by the same surgeon with the same procedure

Recap: Go Back to Identifiability

- ATE:

$$E[Y(1) - Y(0)]$$
$$= E[Y(1)] - E[Y(0)]$$
$$= E_X[E[Y(1)|X] - E[Y(0)|X]]$$ Law of total expectation
$$= E_X[E[Y(1)|X, T = 1] - E[Y(0)|X, T = 0]]$$ Unconfoundedness & positivity
$$= E_X[E[Y|X, T = 1] - E[Y|X, T = 0]]$$ consistency

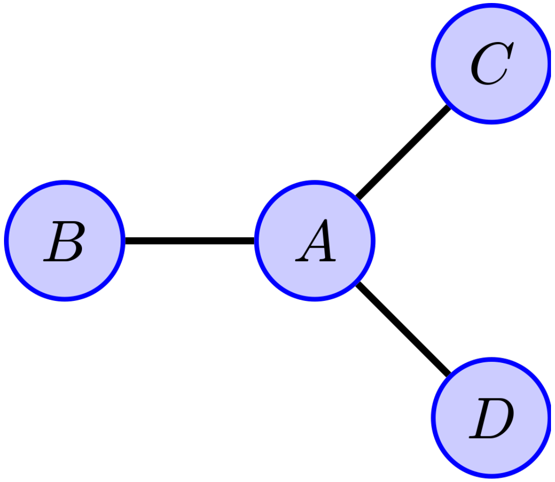
Statistical quantities → Causal quantities

Outline

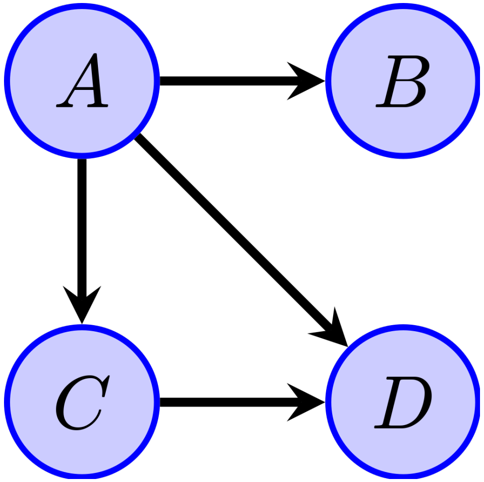
- Introduction to Graphical Models
  - Undirected graphical models
  - Directed graphical models
- Structural Causal Model
  - Causal graph
  - Structural equations
  - Intervention
  - Backdoor adjustment

Graphical Model

- A graphical model is a probabilistic model for which a graph expresses the conditional dependence structure between random variables.
- Commonly used in probability theory, statistics—particularly Bayesian statistics and ML.



An undirected graph with four vertices



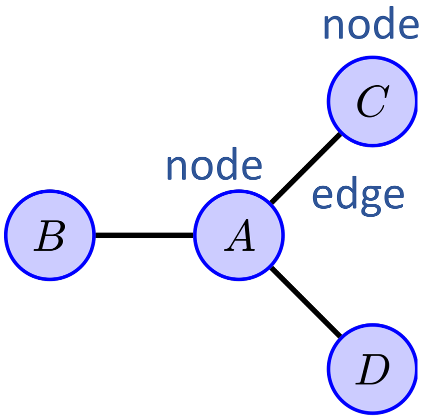
Example of a directed acyclic graph on four vertices.

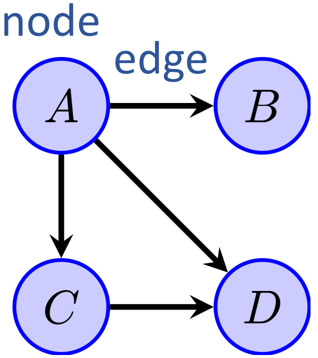
Graphical Model

- Natural tool for handling Uncertainty and Complexity
  - which occur throughout applied mathematics and engineering
- Fundamental to the idea of a graphical model is the notion of modularity
  - a complex system is built by combining simpler parts.

Basic Concepts in Graphs

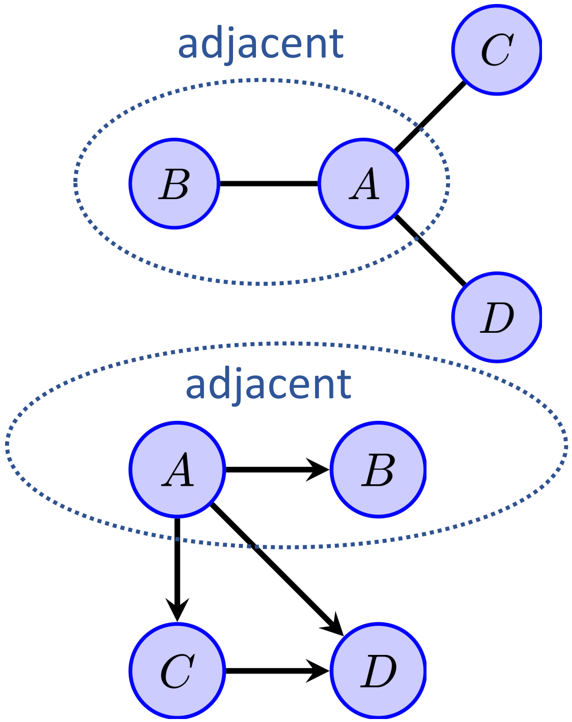
- Node (a.k.a. vertex)
- Edge (a.k.a. link)
    - Directed (arrow)
    - Undirected





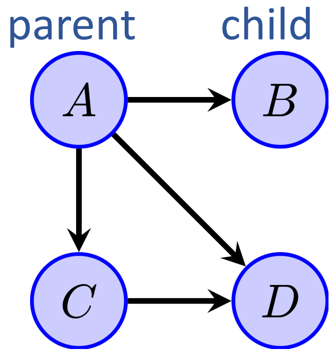
Basic Concepts in Graphs

- Node
- Edge
    - Directed (arrow)
    - Undirected
- Adjacent/Neighbor



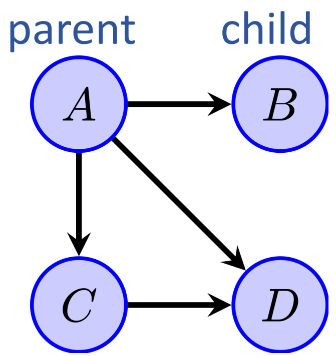
Basic Concepts in Graphs

- Node
- Edge
  - Directed (arrow)
  - Undirected
- Adjacent/Neighbor
- Parent & Child



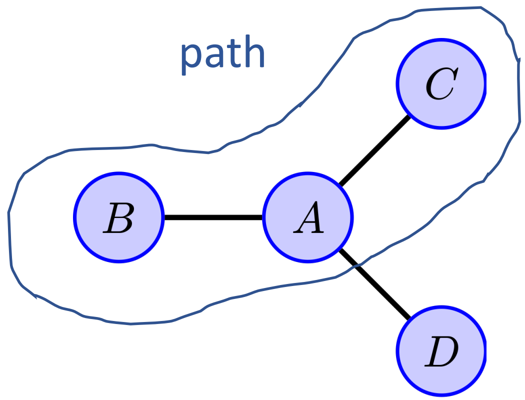
Basic Concepts in Graphs

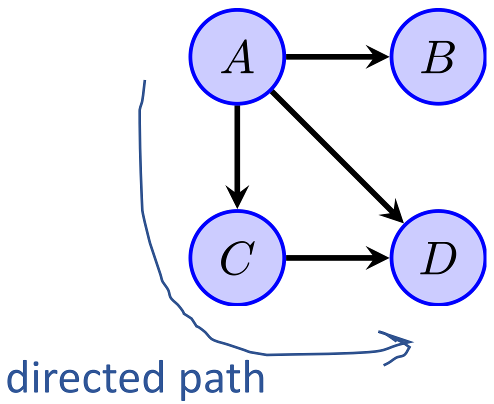
- Node
- Edge
  - Directed (arrow)
  - Undirected
- Adjacent/Neighbor
- Parent & Child
- Ancestor/Descendant
  - Parents-of-parents-of…
  - Children-of-children-of…



Basic Concepts in Graphs

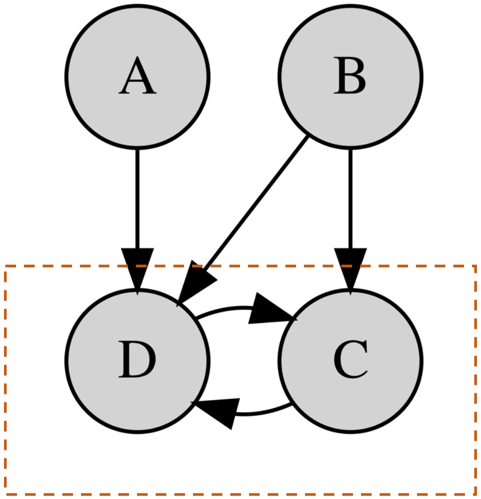
- Node
- Edge
  - Directed (arrow)
  - Undirected
- Adjacent/Neighbor
- Parent & Child
- Ancestor/Descendant
- Path





Basic Concepts in Graphs

- Node
- Edge
  - Directed (arrow)
  - Undirected
- Adjacent/Neighbor
- Parent & Child
- Ancestor/Descendant
- Path
- Circle

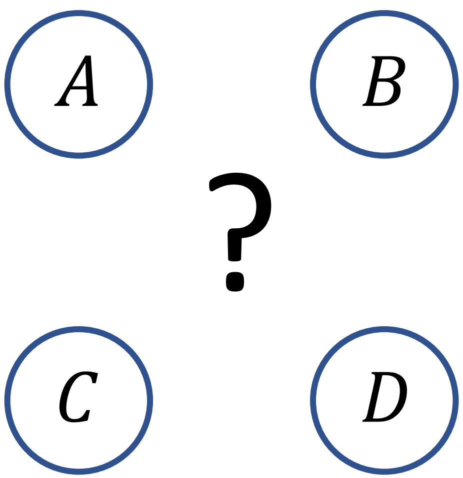


Joint Distribution

- Naïve modeling for joint distribution:

$$P(X_1, \dots, X_n) = P(X_1)P(X_2|X_1)\dots P(X_n|X_{n-1}, \dots, X_1)$$

In binary cases, how many possible combinations of values for n variables?



$$P(A, B, C, D)$$
$$= P(A)P(B|A)P(C|A, B)P(D|A, B, C)$$

Joint Distribution

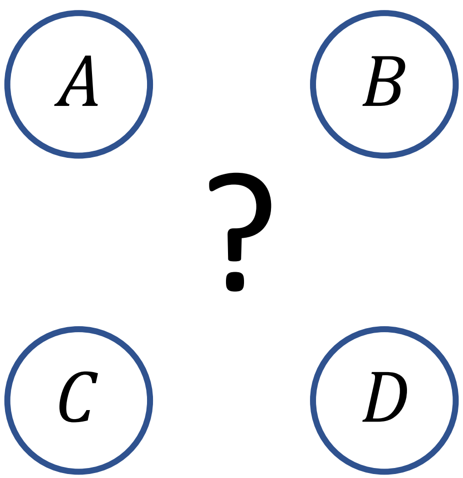
2x2x2x2=16
combinations

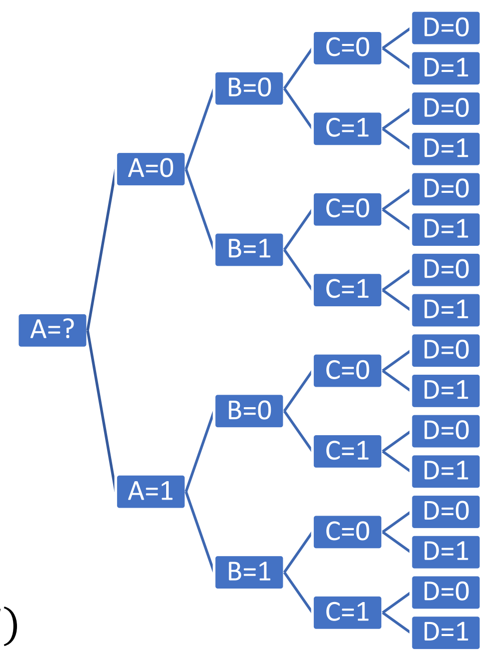
- Naïve modeling for joint distribution:

$$P(X_1, \ldots, X_n) = P(X_1)P(X_2|X_1) \ldots P($$

In binary cases, how many possible combinations of values for n variables?

index	A	B	C	D
1	0	0	0	0
2	0	0	0	1
3	0	0	1	0
4	0	1	0	0
5	1	0	0	0
6	0	0	1	1
7	0	1	0	1
8	1	0	0	1
9	0	1	1	0
10	1	0	1	0
11	1	1	0	0
12	0	1	1	1
13	1	0	1	1
14	1	1	0	1
15	1	1	1	0
16	1	1	1	1





$$P(A, B, C, D) = P(A)P(B|A)P(C|A, B)P(D|A, B, C)$$

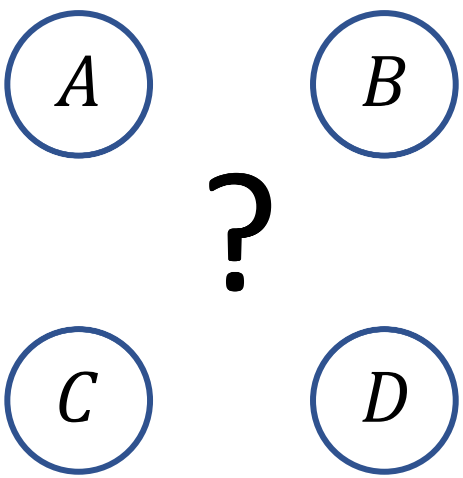
Joint Distribution

- Naïve modeling for joint distribution:

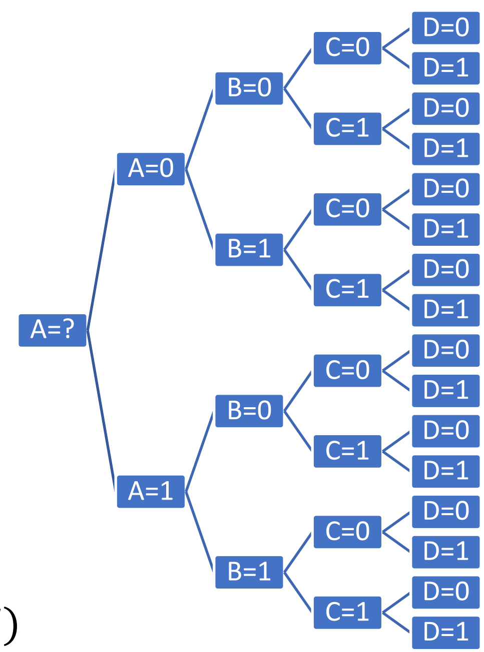
$$P(X_1, \dots, X_n) = P(X_1)P(X_2|X_1) \dots P(X_n|X_{n-1}, \dots, X_1)$$

In binary cases, how many possible combinations of values for n variables?

2^n combinations



$$P(A, B, C, D)$$
$$= P(A)P(B|A)P(C|A, B)P(D|A, B, C)$$



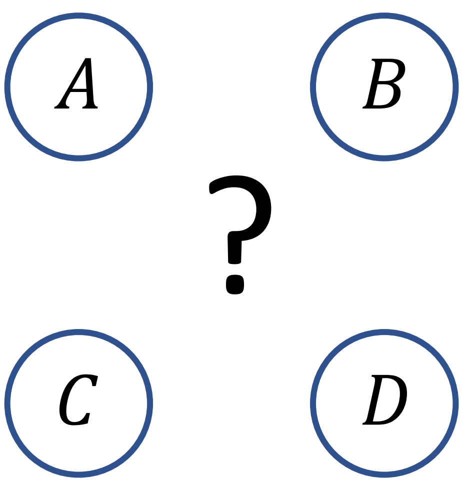
Joint Distribution

- Naïve modeling for joint distribution:

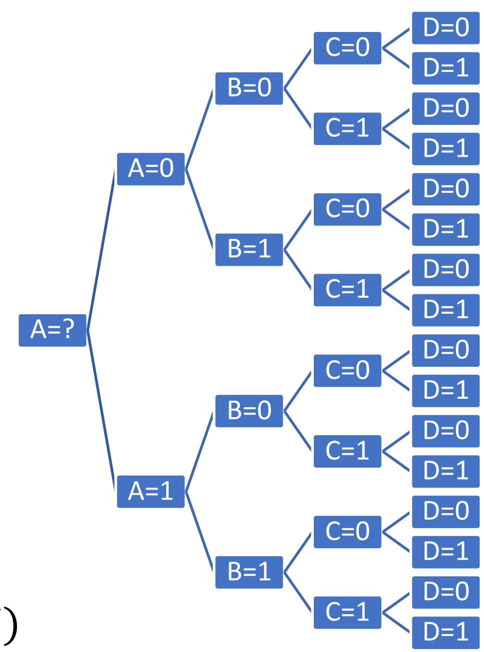
$$P(X_1, \ldots, X_n) = P(X_1)P(X_2|X_1) \ldots P(X_n|X_{n-1}, \ldots, X_1)$$

How many parameters are needed to describe the joint distribution?

$2^n - 1$ parameters

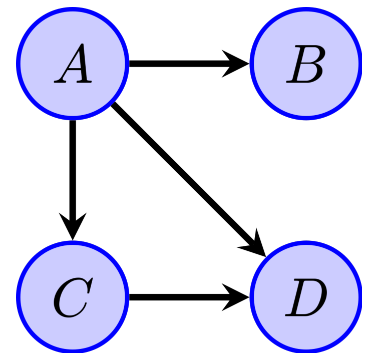


$$P(A, B, C, D)$$
$$= P(A)P(B|A)P(C|A,B)P(D|A,B,C)$$

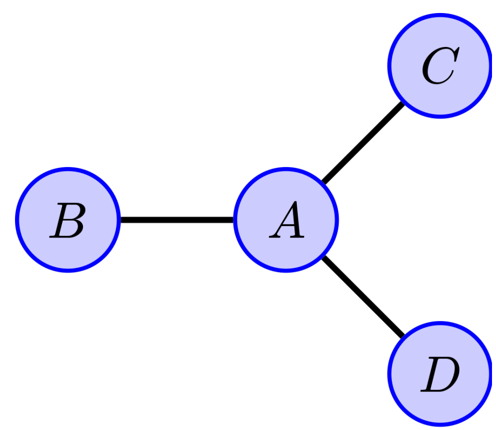


Graph Directionality

- Directed graphical models
  - Direction in edges
  - Bayesian networks
  - More popular in AI and statistics



- Undirected graphical models
  - Edges without direction
  - Markov random fields (MRFs)
    - Better suited to express soft constraints between variables
  - More popular in Vision and Physics

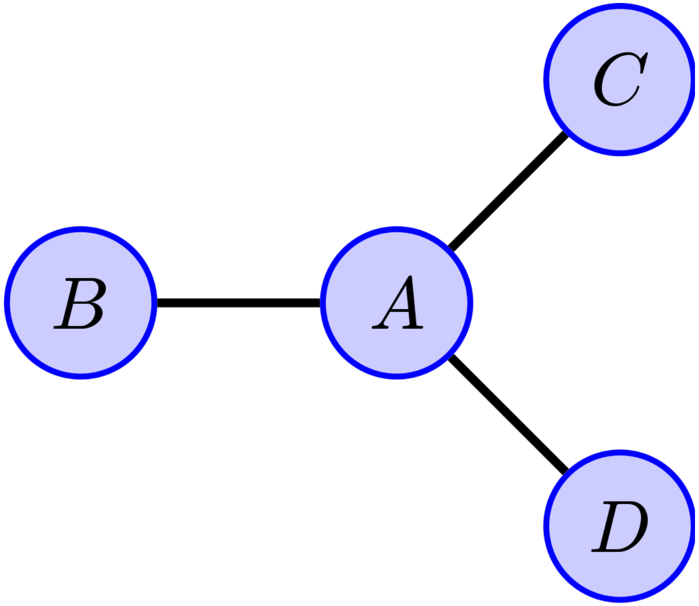


Outline

- Introduction to Graphical Models
  - Undirected graphical models
  - Directed graphical models
- Structural Causal Model
  - Causal graph
  - Structural equations
  - Intervention
  - Backdoor adjustment

Undirected Graphical Model

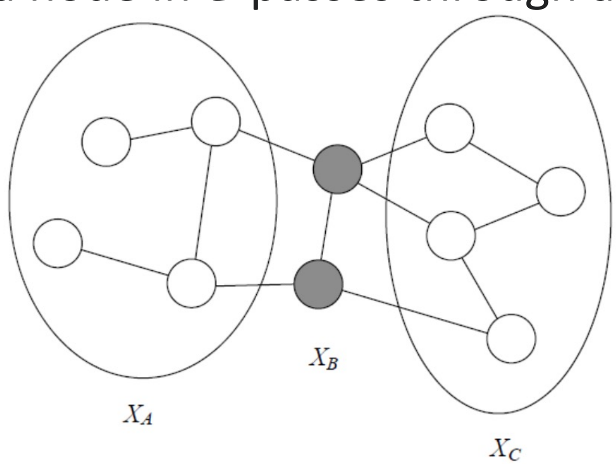
- An edge implies dependence between the corresponding random variables.



$P(A, B, C, D) = ?$

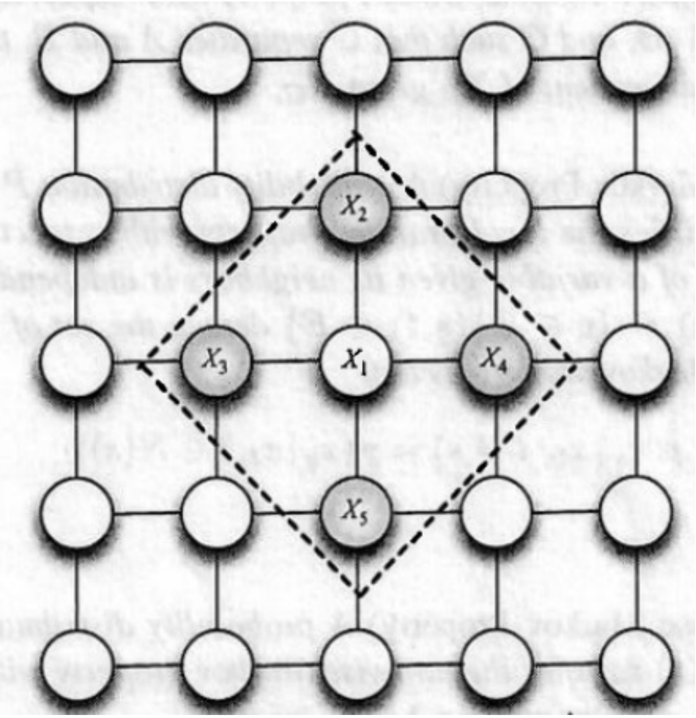
Markov Properties on Undirected Graphs

- **Local Markov Property**: For each variable, given its neighbors, this variable is conditionally independent of other variables.
- **Global Markov Property**: For any disjoint node subsets A, B, and C, such that B separates A and C, the random variables X_A are conditionally independent of X_C given X_B.
  - Here, we say B separates A and C if every path from a node in A to a node in C passes through a node in B.



B separates A and $C \Rightarrow X_A \perp\!\!\!\perp X_C \mid X_B$.

Separation



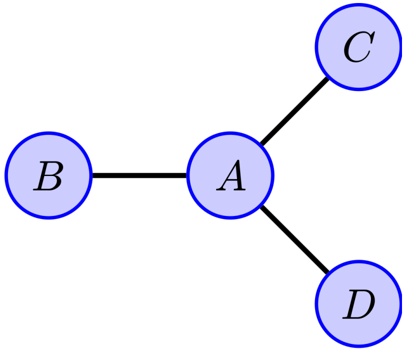
Markov Random Field (MRF)

- MRF, Markov network or undirected graphical model is a set of random variables having a Markov property described by an undirected graph G.

$$P(X = x) = \prod_{C \in cl(G)} \phi_c(x_C)$$

$cl(G)$: the set of cliques of G

- From this graph, B,C,D are all mutually independent, once A is known.



$$P(A, B, C, D) = f_{AB}(A, B) \cdot f_{AC}(A, C) \cdot f_{AD}(A, D)$$

non-negative functions

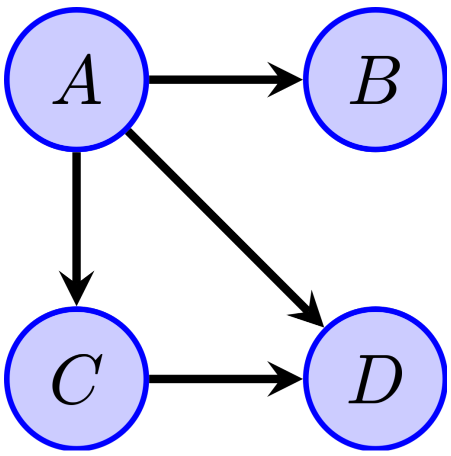
Outline

- Introduction to Graphical Models
  - Undirected graphical models
  - Directed graphical models
- Structural Causal Model
  - Causal graph
  - Structural equations
  - Intervention
  - Backdoor adjustment

Bayesian Network

- If the network is a directed acyclic graph (DAG), the model represents a factorization of the joint probability of all random variables.
- For $X_1, \ldots, X_n$, the joint probability satisfies

$$P(X_1, \ldots, X_n) = ?$$



Example of a directed acyclic graph on four vertices.

Markov Properties on Directed Graphs

- **Local Markov Property**: Each variable is conditionally independent of its non-descendants given its parent variables.

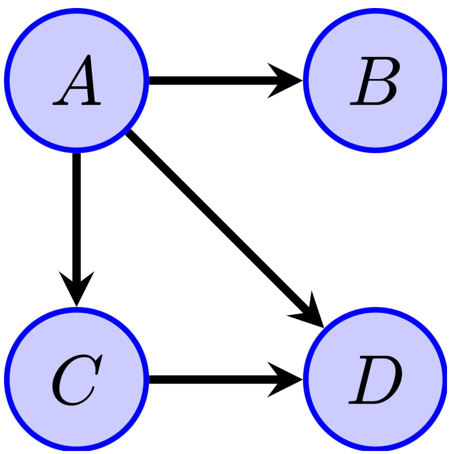
Bayesian Network Factorization

- For $X_1, \dots, X_n$, the joint probability satisfies

$$P(X_1, \dots, X_n) = \prod_{i=1}^{n} P(X_i | pa(X_i))$$

Parents of node X_i

local Markov assumption => Bayesian network factorization
Bayesian network factorization => local Markov assumption

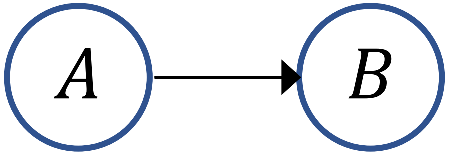


Example of a directed acyclic graph on four vertices.

$$P(A, B, C, D) = \mathrm{P(A)} \cdot P(B|A) \cdot P(C|A) \cdot P(D|A, C)$$

Minimality Assumption

- Another important assumption we use in this course
- Two parts:
  - (Local Markov assumption): Given its parents in the DAG, a node is independent of all its non-descendants.
  - Adjacent nodes in the DAG are dependent.



$$P(A, B) = P(A) \cdot P(B|A)$$

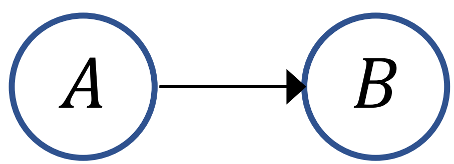
$$\cancel{P(A, B) = P(A) \cdot P(B)}$$

Minimality Assumption

- Another important assumption we use in this course
- Two parts:

Statistical Independencies

  - (Local Markov assumption): Given its parents in the DAG, a node is independent of all its non-descendants.
  - Adjacent nodes in the DAG are dependent.

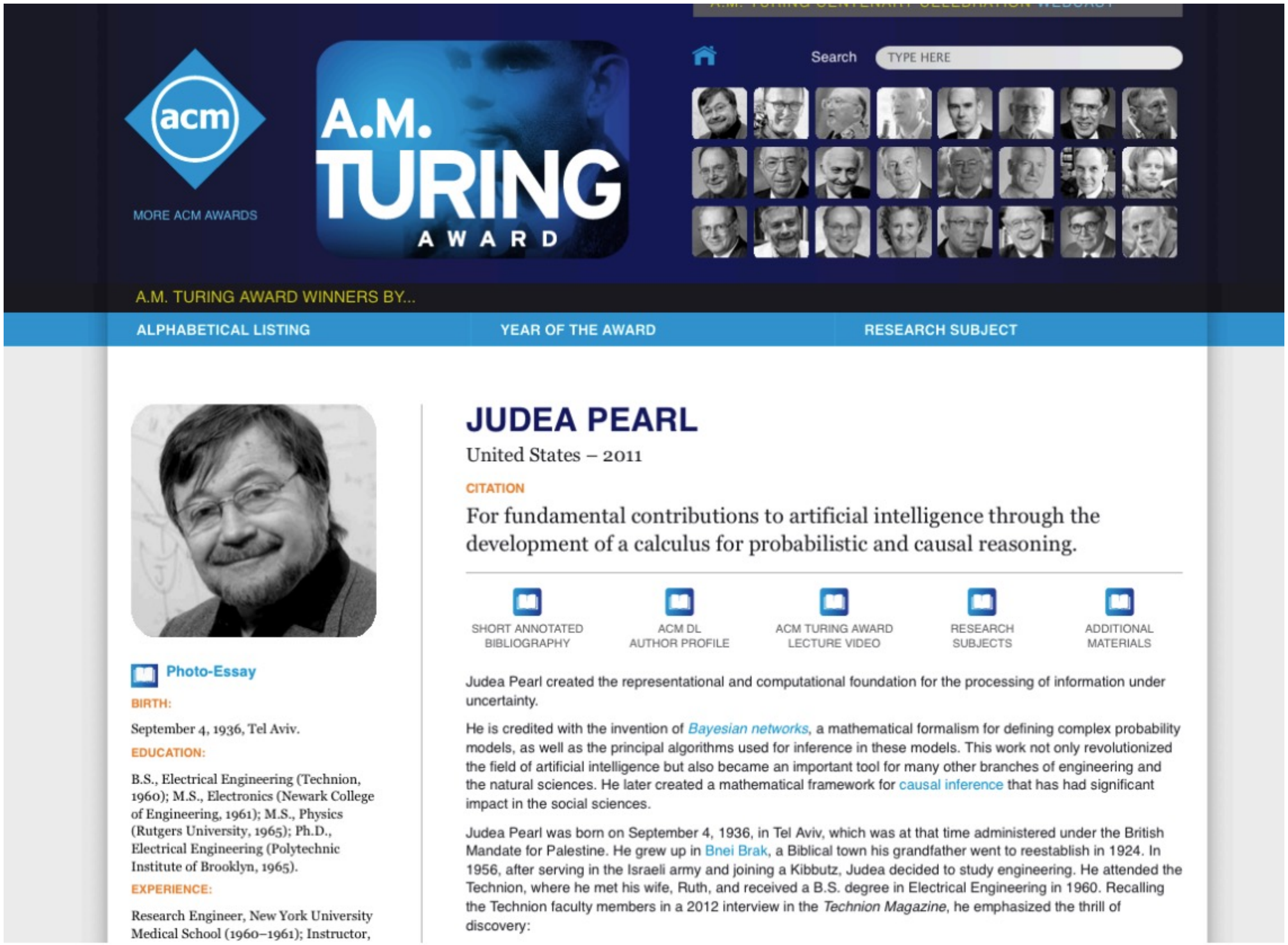


Statistical Dependencies

$$P(A,B) = P(A) \cdot P(B|A)$$

$$\cancel{P(A,B) = P(A) \cdot P(B)}$$

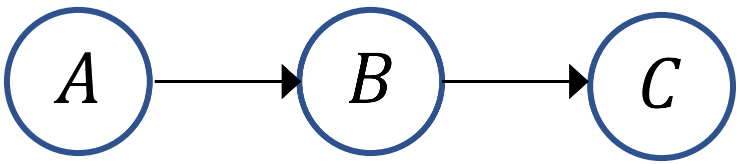
2011 Turing award was for Bayesian networks



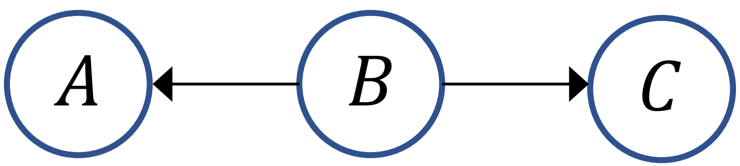
D-separation

- D stands for "directional"
- For three disjoint subsets A, B, C of nodes in graphical model, we say A and C are d-separated by B if all of the paths between (any node in) A and (any node in) C are blocked by B.

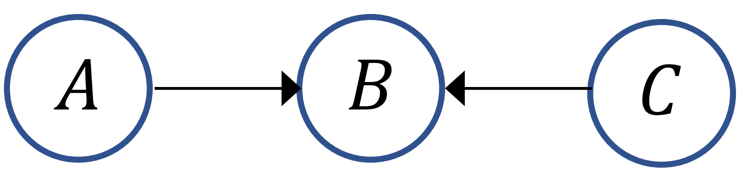
Junction Patterns



Chain



Fork



Collider

Junction Patterns

$A \rightarrow B \rightarrow C$ **Chain**

$A \not\perp\!\!\!\perp C$

$A \perp\!\!\!\perp C \mid B$

$A \leftarrow B \rightarrow C$ **Fork**

$A \not\perp\!\!\!\perp C$

$A \perp\!\!\!\perp C \mid B$

$A \rightarrow B \leftarrow C$ **Collider**

$A \perp\!\!\!\perp C$

$A \not\perp\!\!\!\perp C \mid B$

Blocked Paths

- A path between nodes X and Y is blocked by a (potentially empty) conditioning set Z if either of the following is true:
  - Along the path, there is a chain $\ldots \rightarrow W \rightarrow \cdots$ or a fork $\ldots \leftarrow W \rightarrow \cdots$ where W is conditioned on ($W \in Z$).
  - There is a collider W on the path that is not conditioned on ($W \notin Z$) and none of its descendants are conditioned on ($\mathrm{des}(W) \nsubseteq Z$).

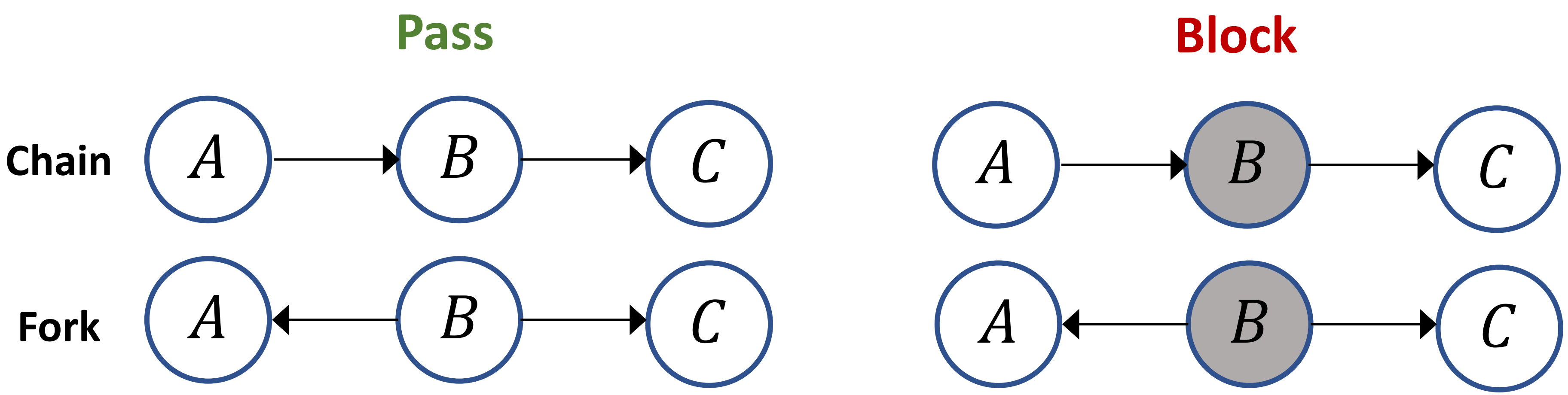
Blocking Conditions

Pass **Block**

Chain

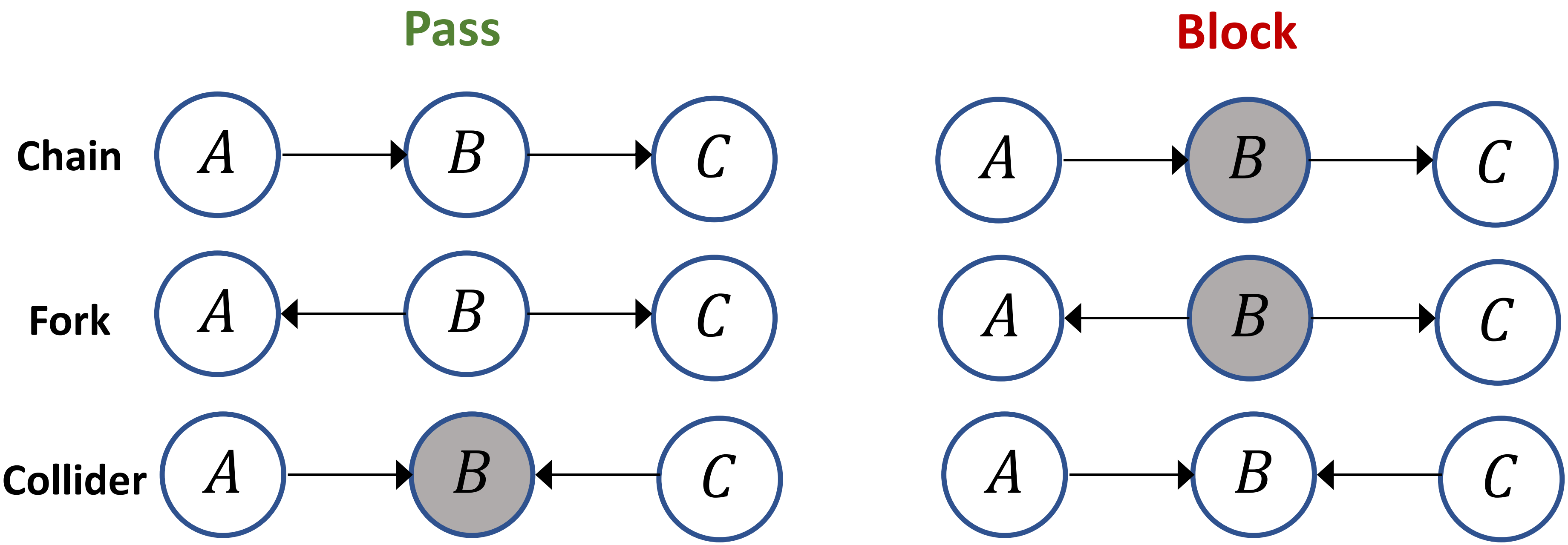
Grey color means "given"

Blocking Conditions



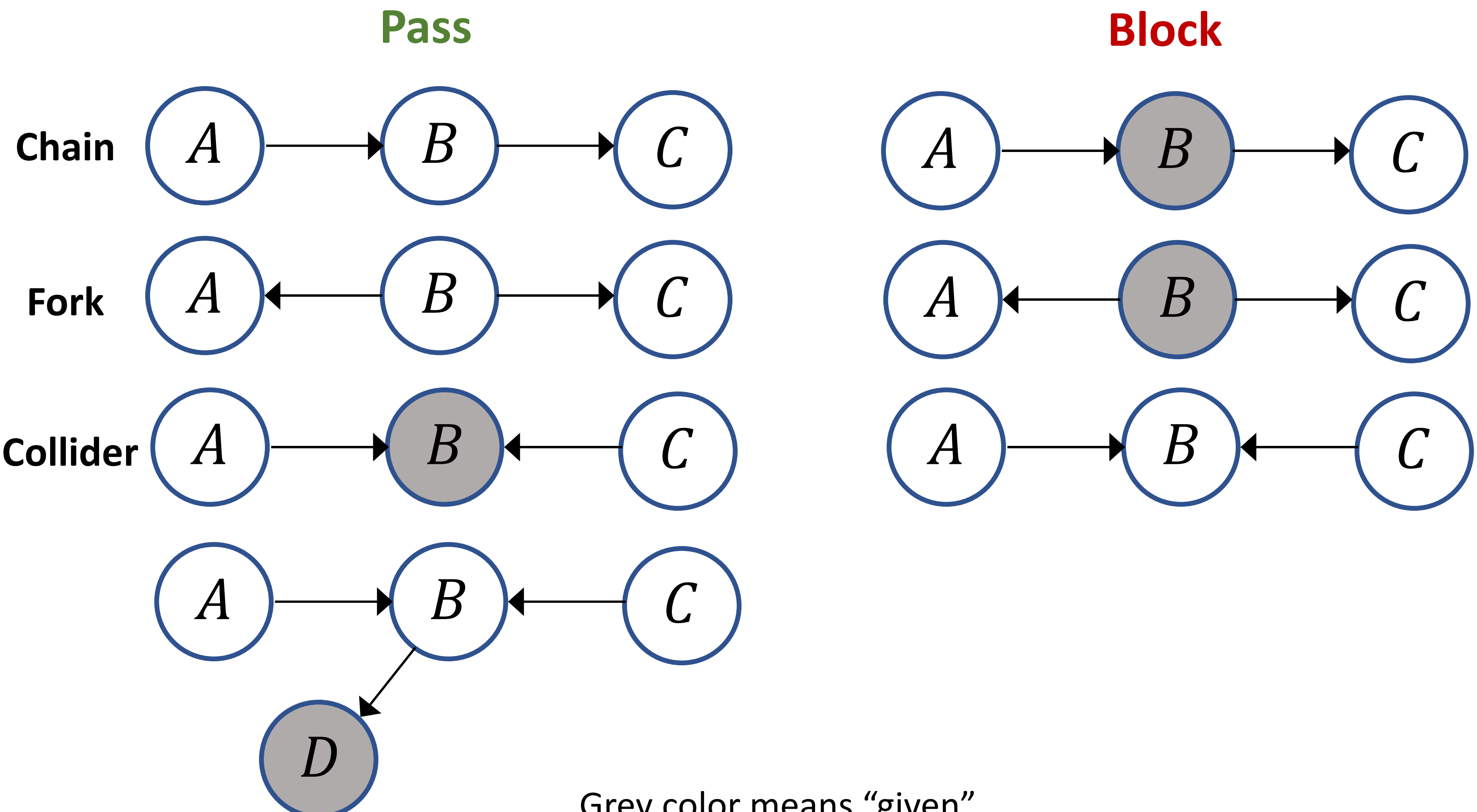
Grey color means “given”

Blocking Conditions



Grey color means “given”

Blocking Conditions



Grey color means "given"

Outline

- Introduction to Graphical Models
  - Undirected graphical models
  - Directed graphical models
- Structural Causal Model
  - Causal graph
  - Structural equations
  - Intervention
  - Backdoor adjustment

Causation in Graphical Model

- A variable X is said to be a cause of a variable Y if Y can change in response to changes in X.
- **Causal edges assumption**: In a directed graph, every parent is a direct cause of all its children

Causal Dependencies

Causation in Graphical Model

- A variable X is said to be a cause of a variable Y if Y can change in response to changes in X.
- **Causal edges assumption**: In a directed graph, every parent is a direct cause of all its children

Causal Dependencies

- A **causal graph** is a Bayesian network with the requirement that the relationships be causal.

DAG + Markov assumption + Causal edges assumption => Causal graph

Markov Assumption

- D-separation: For three disjoint subsets A, B, C of nodes, we say A and C are d-separated by B if all of the paths between (any node in) A and (any node in) C are blocked by B.
- **Global Markov assumption**: Given causal graph G and distribution P (w.r.t., G),

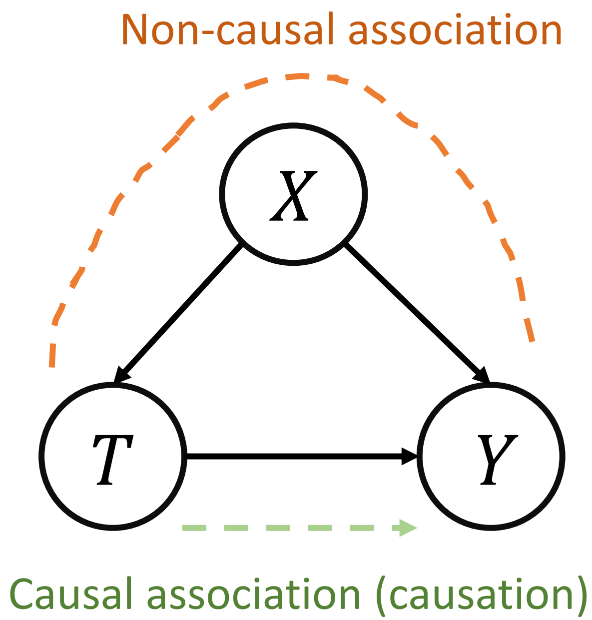
$$X_A \perp\!\!\!\perp_G X_C \mid X_B \Rightarrow X_A \perp\!\!\!\perp_P X_C \mid X_B.$$

D-separation in G

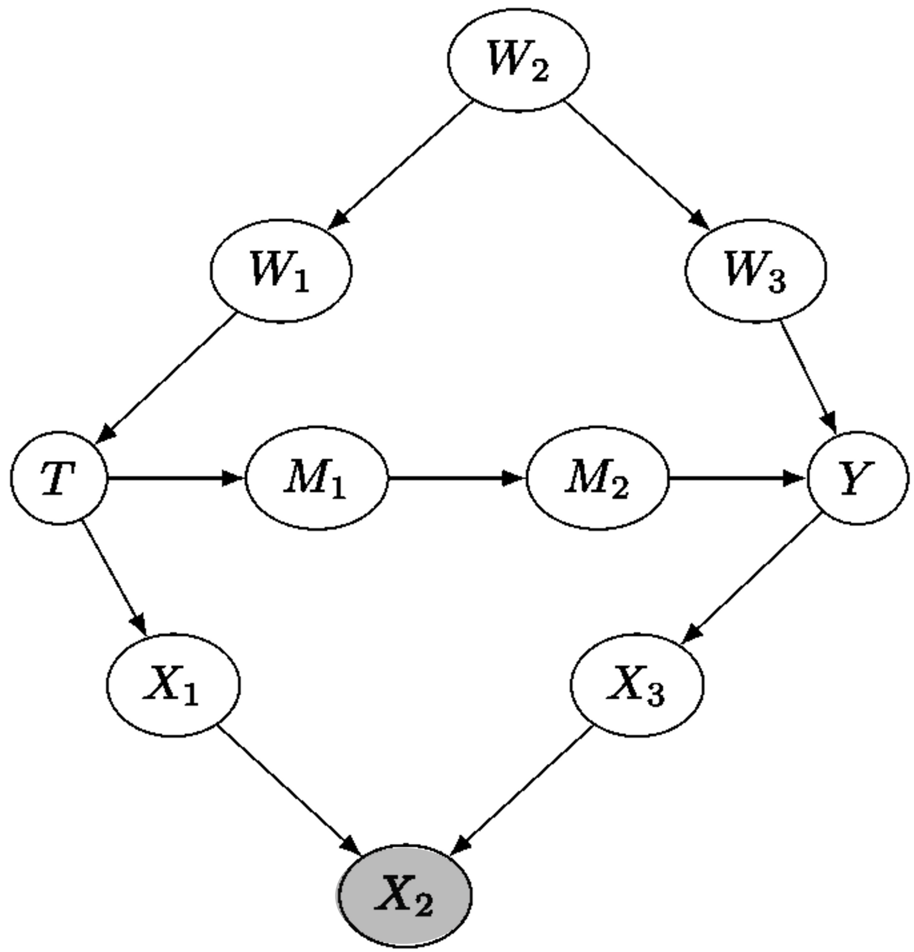
Conditional independence in P

local Markov assumption ⇔ global Markov assumption

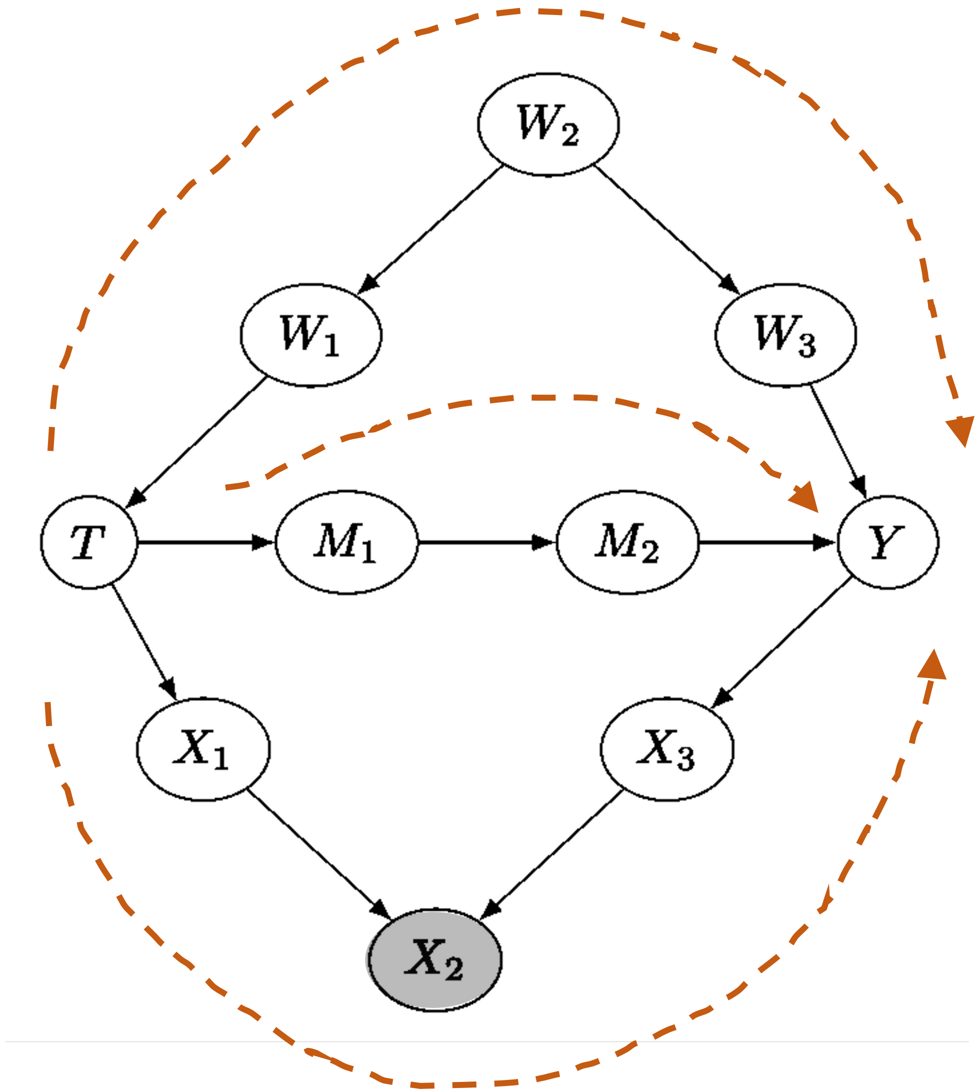
Association and Causation



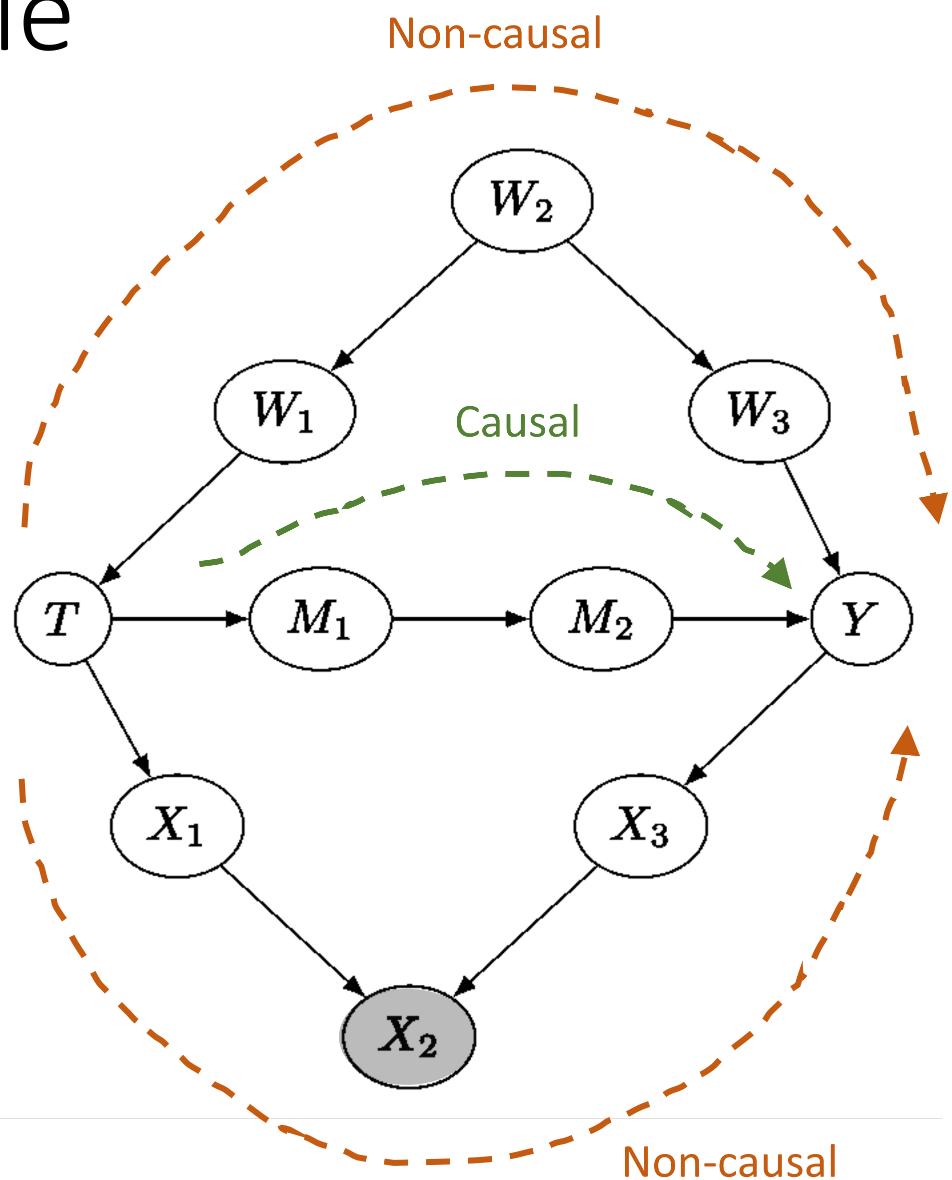
Example



Example



Example



Outline

- Introduction to Graphical Models
  - Undirected graphical models
  - Directed graphical models
- Structural Causal Model
  - Causal graph
  - Structural equations
  - Intervention
  - Backdoor adjustment

Identification and Estimation

Underlying Causality

Identifying assumptions ↓ ↑ Identification

Causal Model

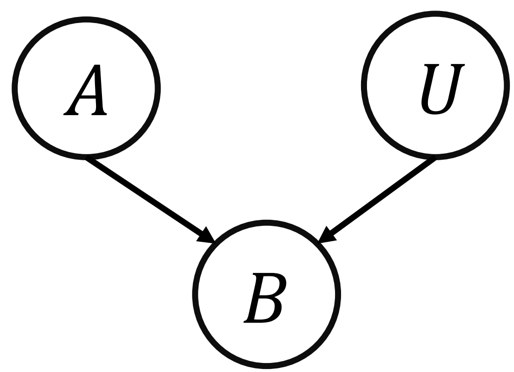
Population distribution

Sampling ↓ ↑ Estimation, inference

Observational data

Structural Equation

- The “equals sign” does not convey any causal information.
  - $B = A \Leftrightarrow A = B$ (symmetric)
- Structural equation for A as a cause of B:
  - $B := f(A)$
  - $B := f(A, U)$



Structural Causal Model (SCM)

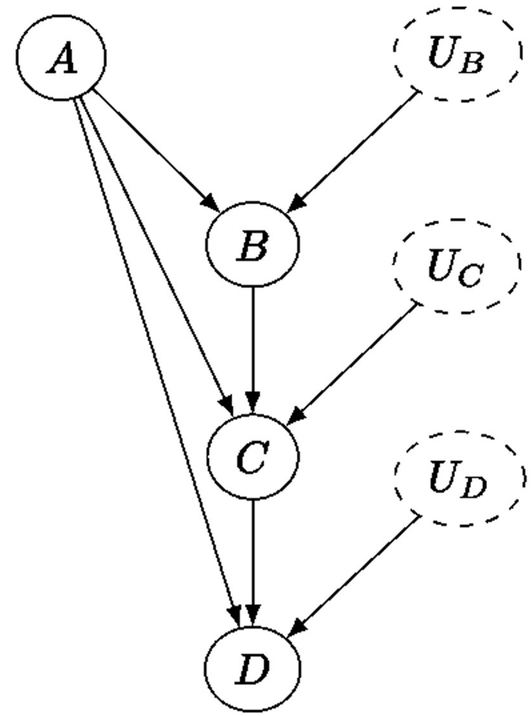
- A triple (U, V, F):
  - A set of exogenous variables (U)
  - A set of endogenous variables (V), determined by variables in $U \cup V$
  - A set of functions $F = \{f_1(\cdot), f_2(\cdot), \ldots, f_{|V|}(\cdot)\}$ (a.k.a. structural equations), each function generate an endogenous variable as a function:

$$V_i = f_i(\mathrm{pa}_i, U_{\mathrm{pa}_i})$$

$\mathrm{pa}_i \subseteq V \backslash \{V_i\}$ $\qquad$ $U_{\mathrm{pa}_i} \subseteq U$

Structural Causal Model (SCM)

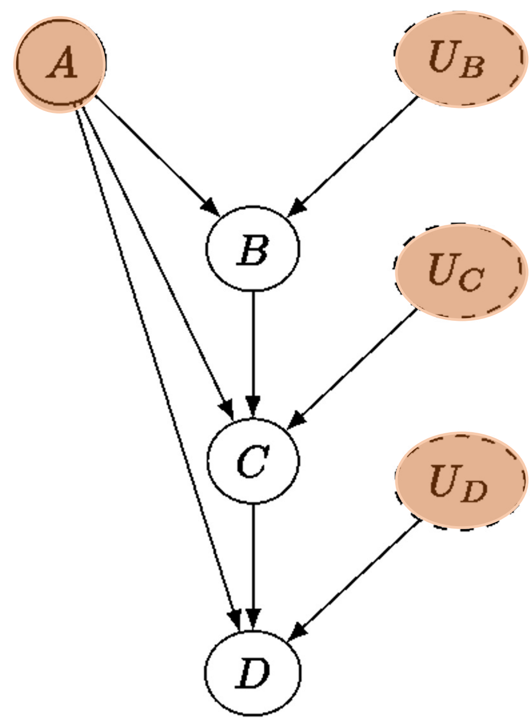
$$\boldsymbol{M}: \quad \begin{aligned} B &:= f_B\,(A,\ U_B) \\ C &:= f_C\,(A,\ B,\ U_C) \\ D &:= f_D\,(A,\ C,\ U_D) \end{aligned}$$



Structural Causal Model (SCM)

$$B := f_B(A, U_B)$$
$$C := f_C(A, B, U_C)$$
$$D := f_D(A, C, U_D)$$

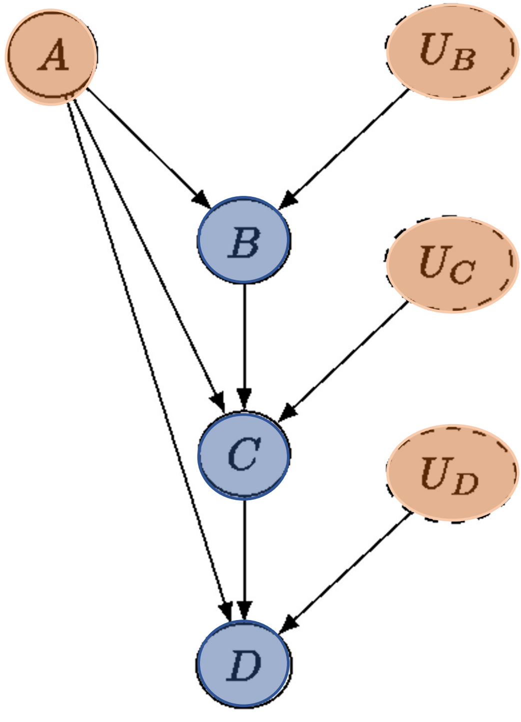
- Exogenous variables



Structural Causal Model (SCM)

$$B := f_B(A, U_B)$$
$$C := f_C(A, B, U_C)$$
$$D := f_D(A, C, U_D)$$

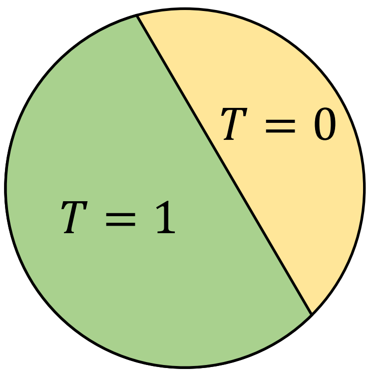
- Exogenous variables
- Endogenous variables



Outline

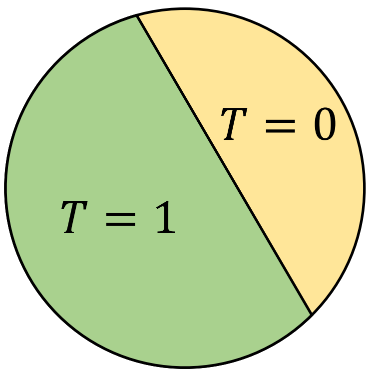
- Introduction to Graphical Models
  - Undirected graphical models
  - Directed graphical models
- Structural Causal Model
  - Causal graph
  - Structural equations
  - Intervention
  - Backdoor adjustment

Conditioning vs. intervening

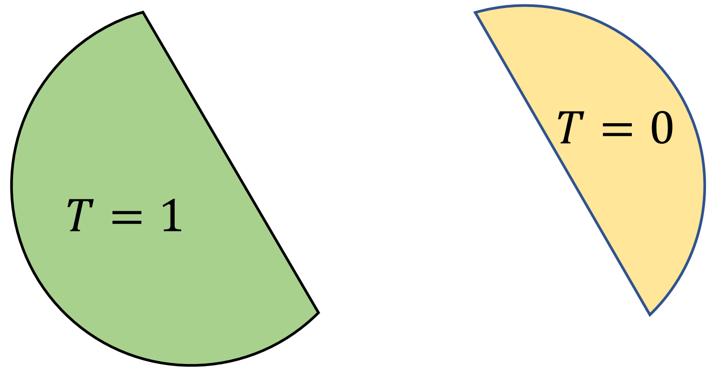


Population

Conditioning vs. intervening

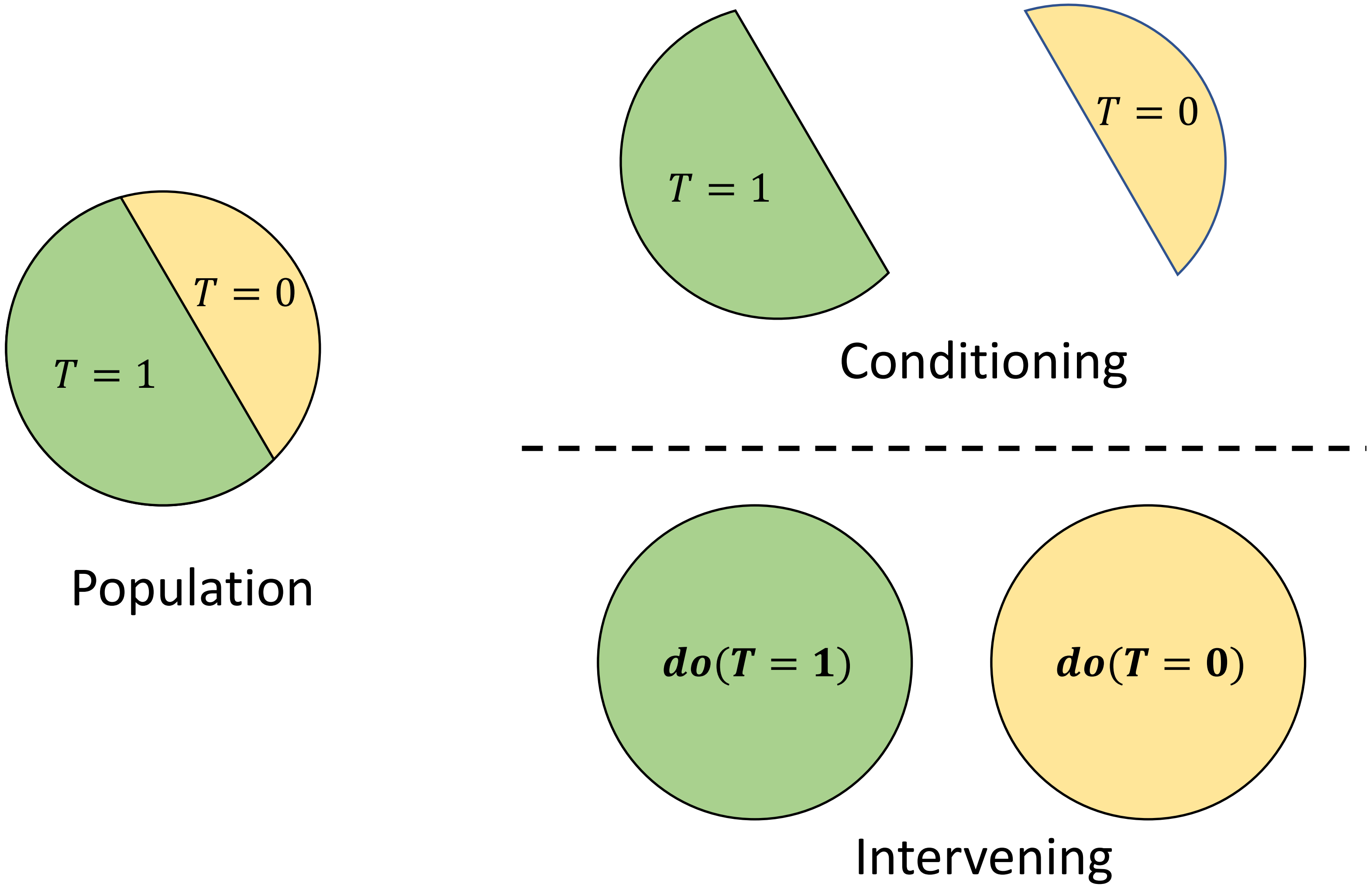


Population



Conditioning

Conditioning vs. intervening



Intervention

- Interventional distributions

$$P(Y(t) = y) \triangleq P(Y = y|do(T = t)) \triangleq P(y|do(t))$$

$$P(Y|T = t) \qquad\qquad P(Y|do(T = t))$$

Observational Interventional

Intervention

- Interventional distributions

$$P(Y(t) = y) \triangleq P(Y = y|do(T = t)) \triangleq P(y|do(t))$$

$$P(Y|T = t) \qquad\qquad P(Y|do(T = t))$$

Observational Interventional

- Average treatment effect (ATE):

$$E[Y|do(T = 1)] - E[Y|do(T = 0)]$$

Identification and Estimation

$P(Y|do(t))$ Underlying Causality

Identifying assumptions ↓ ↑ Identification

Causal Model

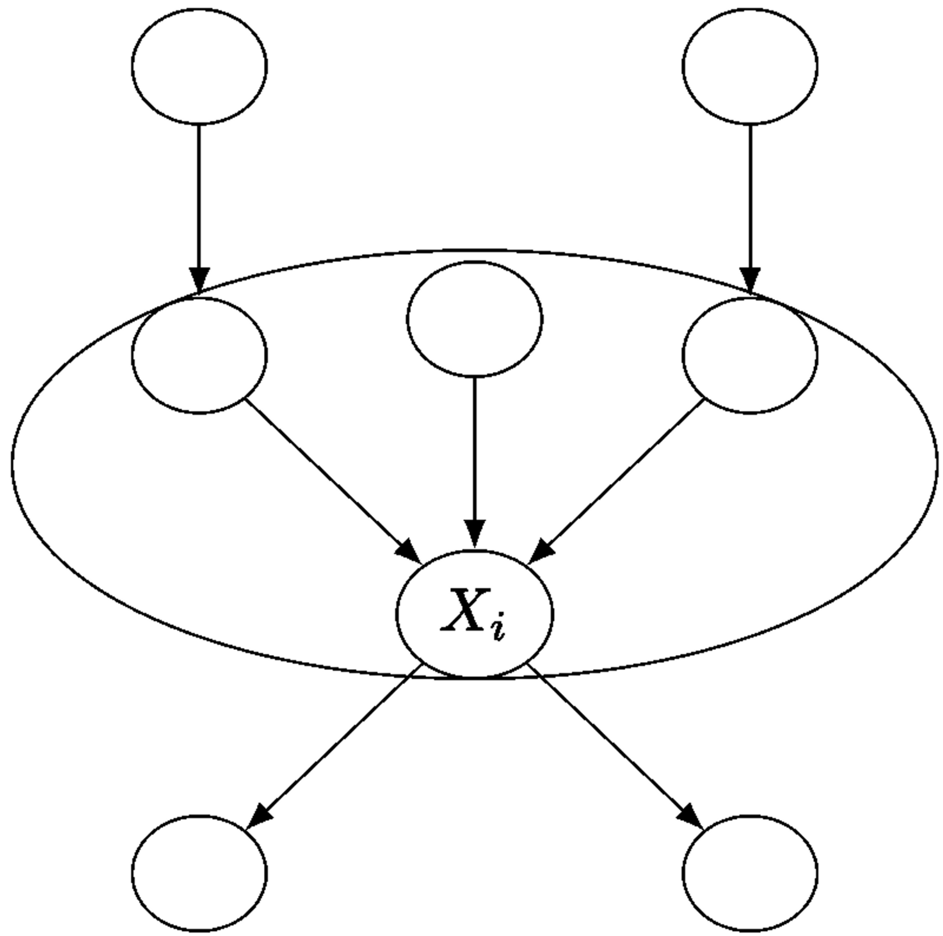
$P(Y|t)$ Population distribution

Sampling ↓ ↑ Estimation, inference

Observational data

Causal Mechanism

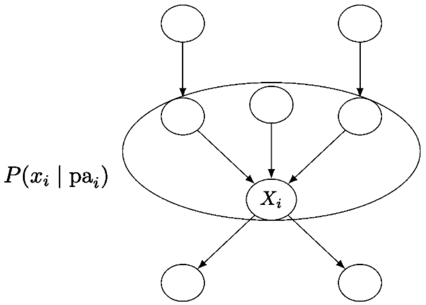
$P(x_i \mid \mathrm{pa}_i)$



Modularity Assumption

- If we intervene on a node X_i, then only the mechanism $P(x_i|\mathrm{pa}_i)$ changes. All other mechanisms remain unchanged.
  - In other words, the causal mechanisms are modular.
  - Other names: independent mechanisms, autonomy, invariance, etc.

Modularity Assumption

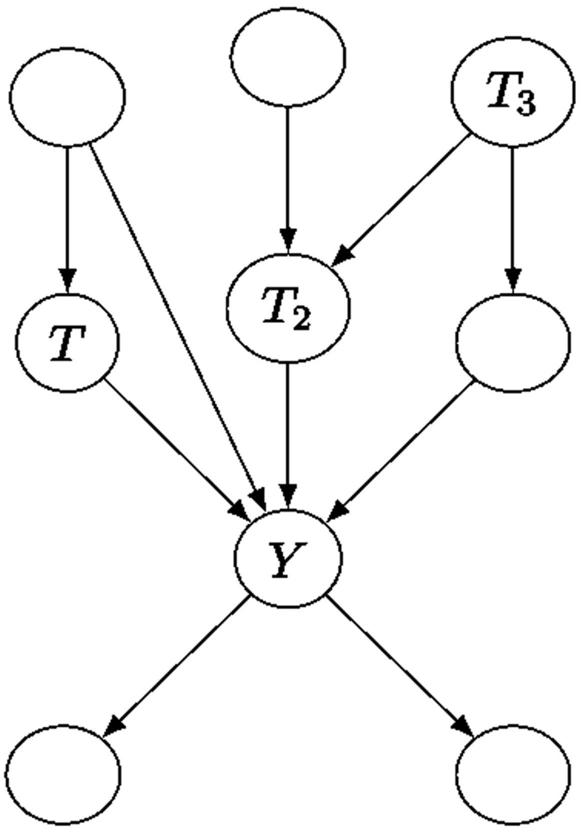


- If we intervene on a node X_i, then only the mechanism $P(x_i|\mathrm{pa}_i)$ changes. All other mechanisms remain unchanged.
  - In other words, the causal mechanisms are modular.
  - Other names: independent mechanisms, autonomy, invariance, etc.

- More formally: If we intervene on a set of nodes $S \subseteq [n]$, setting them to constants, then for all i, we have the following:
  - If $i \notin S$, then $P(x_i|\mathrm{pa}_i)$ remains unchanged.
  - If $i \in S$, then $P(x_i|\mathrm{pa}_i) = 1$ if x_i is the value that X_i was set to by the intervention; otherwise, $P(x_i|\mathrm{pa}_i) = 0$.

Observation v.s. Intervention

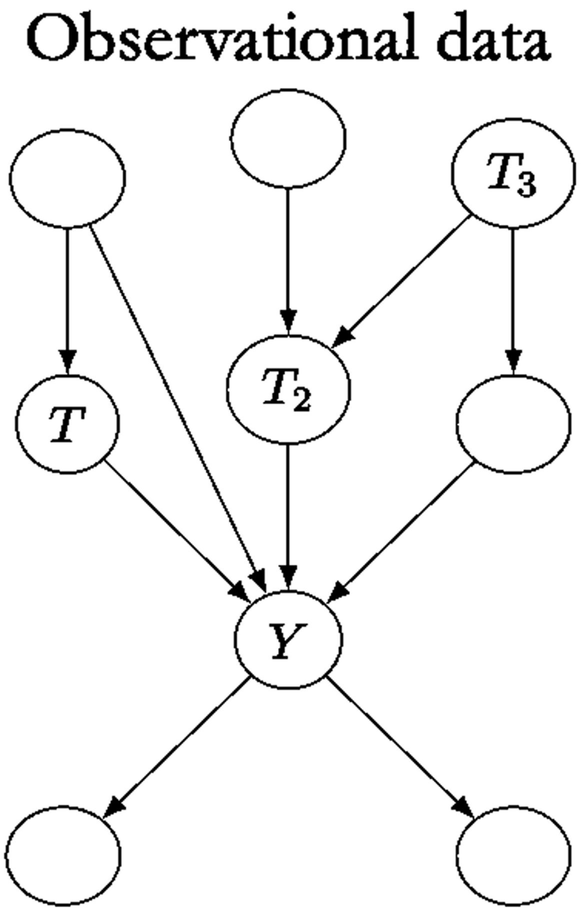
Observational data

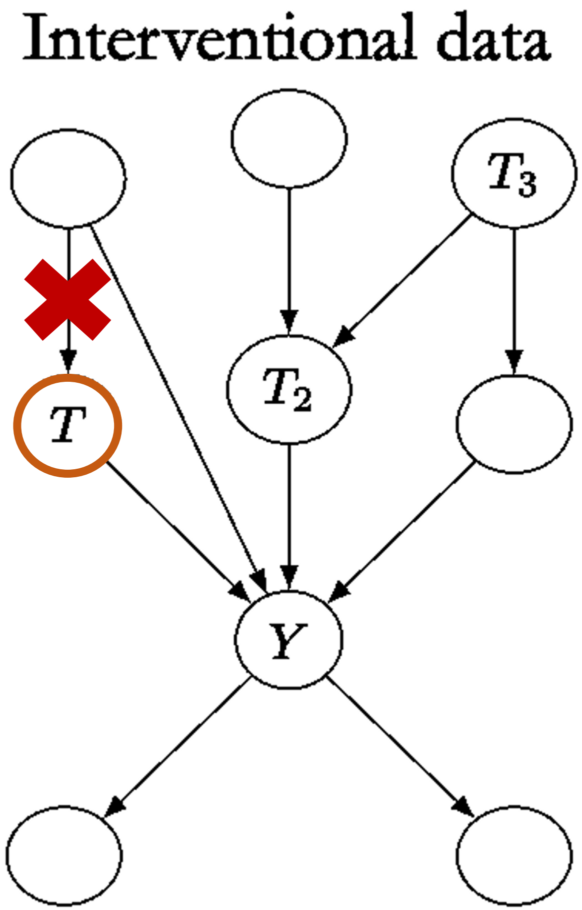


$\boldsymbol{M}$: $T \coloneqq f_T(X, U_T)$

$Y \coloneqq f_Y(X, T, U_Y)$

Observation v.s. Intervention





$\boldsymbol{M}$:
$$T := f_T(X, U_T)$$
$$Y := f_Y(X, T, U_Y)$$

$\boldsymbol{M_t}$:
$$T := t$$
$$Y := f_Y(X, T, U_Y)$$

Modularity Assumption for SCMs

- Consider an SCM M and an interventional SCM M_t got with intervention $do(T = t)$.
- The **modularity assumption** states that M and M_t share all of their structural equations except the structural equation for T, which is $T := t$ in M_t.

$\boldsymbol{M}$: $$T := f_T(X, U_T)$$ $$Y := f_Y(X, T, U_Y)$$

$\boldsymbol{M_t}$: $$T := t$$ $$Y := f_Y(X, T, U_Y)$$

Outline

- Introduction to Graphical Models
  - Undirected graphical models
  - Directed graphical models
- Structural Causal Model
  - Causal graph
  - Structural equations
  - Intervention
  - Backdoor adjustment

Truncated factorization

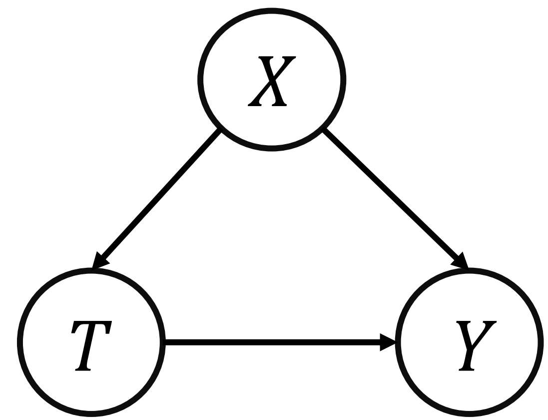
$$P(x_1, \ldots, x_n | do(S = s)) = \prod_{i \notin S} P(x_i | \mathrm{pa}_i)$$

if x is consistent with the intervention.

Otherwise,

$$P(x_1, \ldots, x_n | do(S = s)) = 0$$

Simple identification via truncated factorization



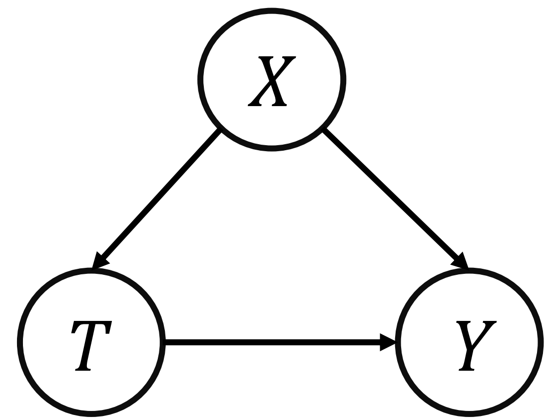
- Goal: identify $P(y|do(t))$
- Bayesian network factorization:

$$P(y, t, x) = P(x)P(t|x)P(y|t, x)$$

- Truncated factorization:

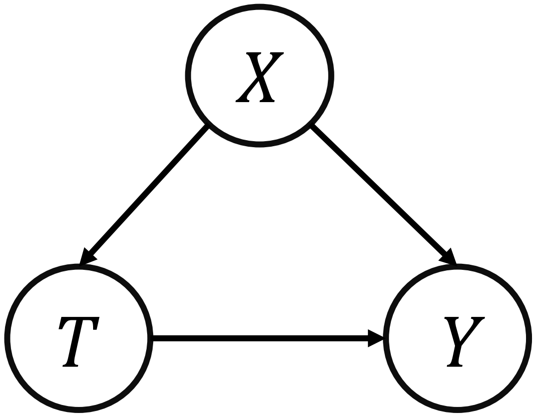
$$P(y, x|do(t)) = P(x)P(y|t, x)$$

Simple identification via truncated factorization



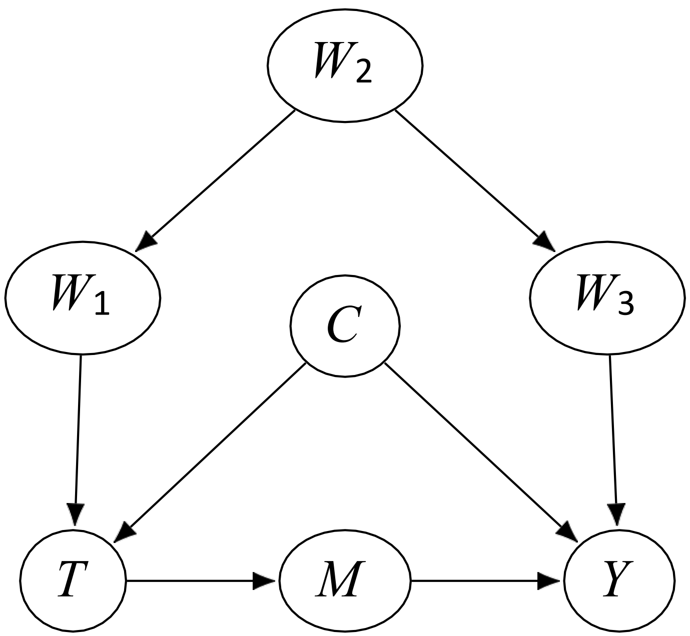
- Goal: identify $P(y|do(t))$
- Bayesian network factorization:
$$P(y, t, x) \; = \; P(x)P(t|x)P(y|t, x)$$
- Truncated factorization:
$$P(y, x|do(t)) = P(x)P(y|t, x)$$
- Marginalize:
$$P(y|do(t)) = \sum_{x} P(x)P(y|t, x)$$

Association vs. causation revisited

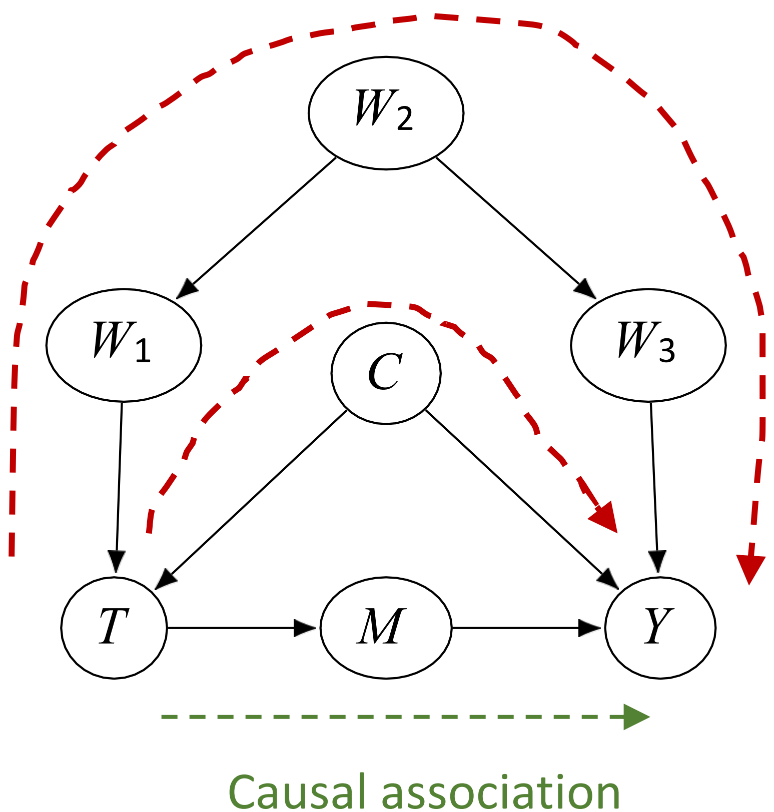


- $P(y|do(t)) = \sum_x P(x) P(y|t,x)$
- $P(y|do(t)) \neq P(y|t) = \sum_x P(x|t) P(y|t,x)$

Backdoor Paths

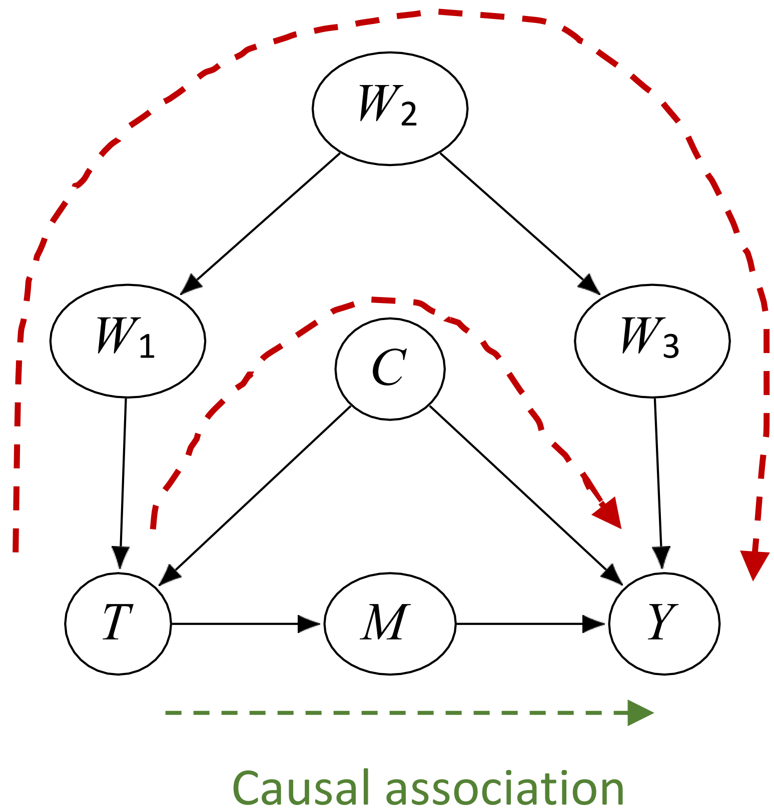


Backdoor Paths

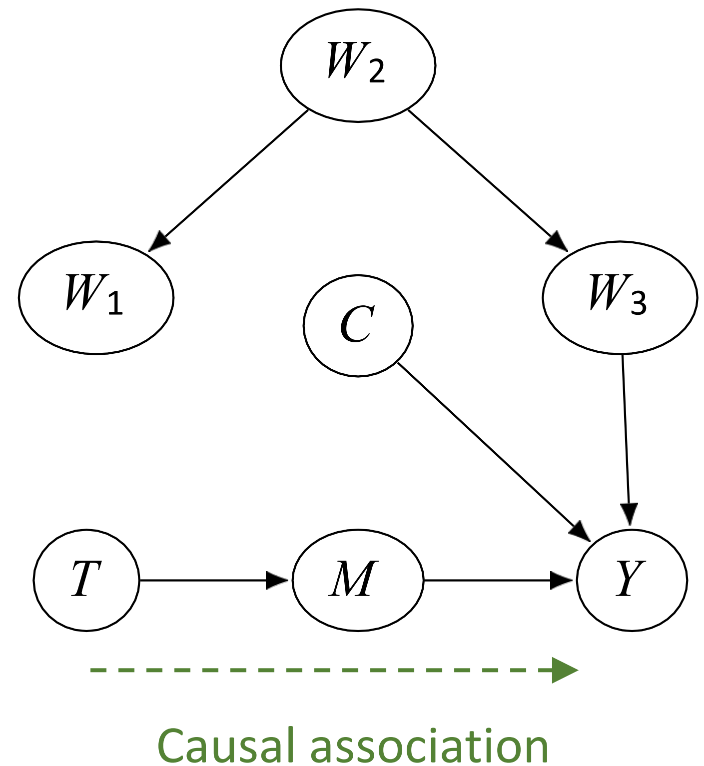


Backdoor Paths

$P(Y|t)$



$P(Y|do(t))$



Backdoor criterion and backdoor adjustment

- A set of variables W satisfies the **backdoor criterion** relative to T and Y if the following are true:
  - W blocks all backdoor paths from T to Y
  - W does not contain any descendants of T

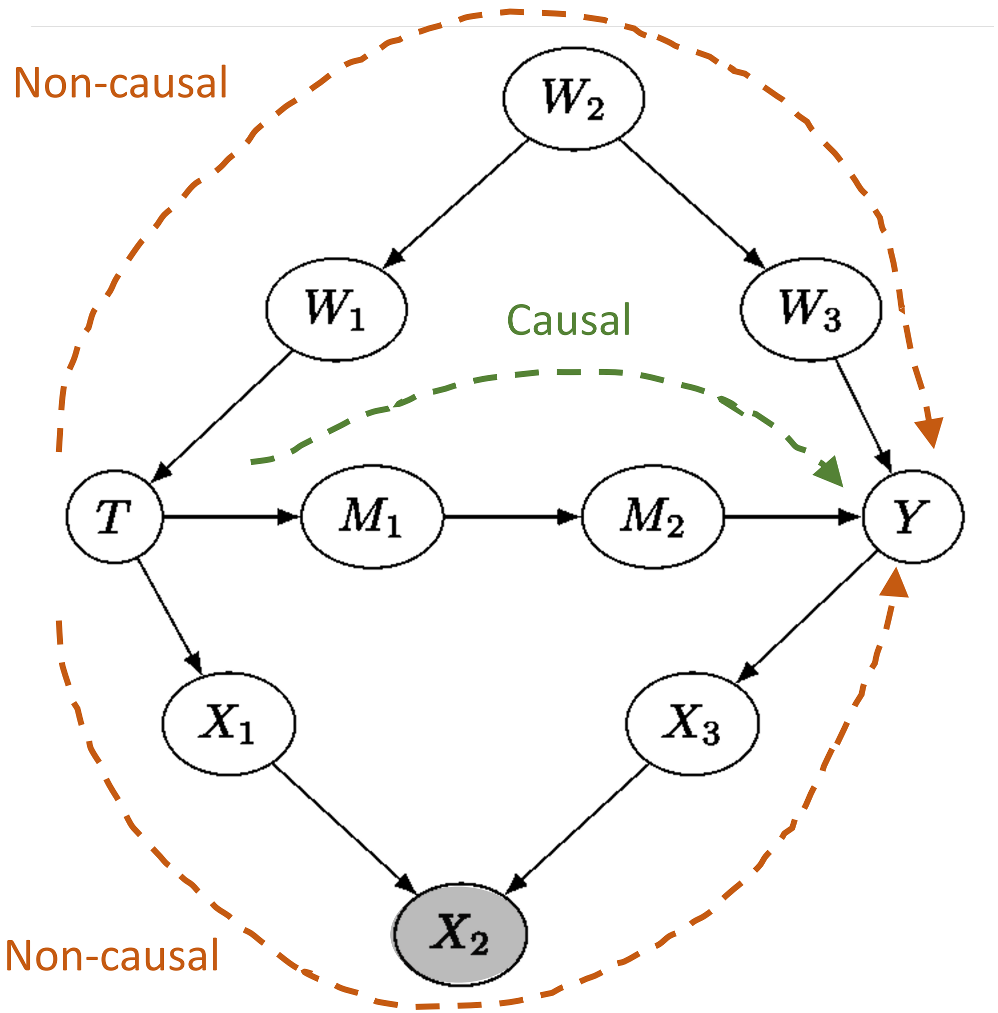
Backdoor criterion and backdoor adjustment

- A set of variables W satisfies the **backdoor criterion** relative to T and Y if the following are true:
  - W blocks all backdoor paths from T to Y
  - W does not contain any descendants of T

- Given the modularity assumption and that W satisfies the backdoor criterion, we can identify the causal effect of T on Y:

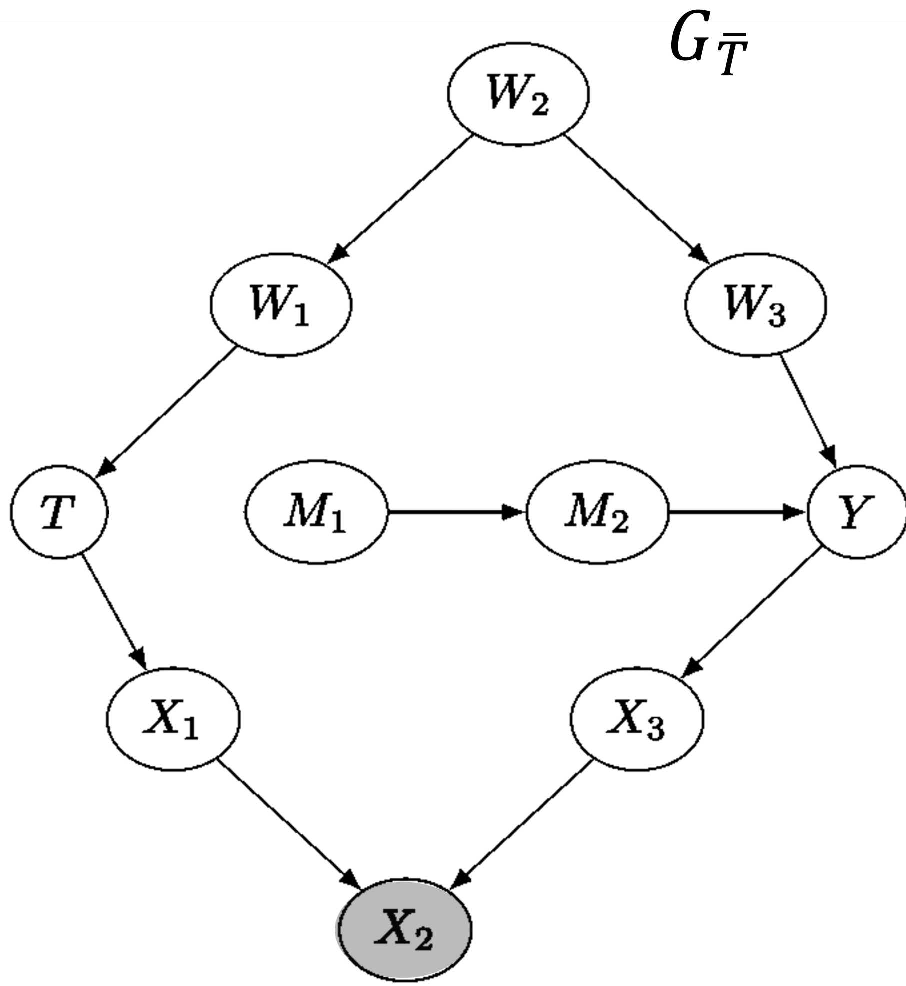
$$P(y|do(t)) = \sum_{w} P(w)P(y|t,w)$$

Backdoor criterion as d-separation



- W blocks all backdoor paths from T to Y
- W does not contain any descendants of T

Backdoor criterion as d-separation



- W blocks all backdoor paths from T to Y
- W does not contain any descendants of T

$$Y \perp\!\!\!\perp_{G_{\overline{T}}} T \mid W$$

Backdoor Adjustment and Adjustment in Potential Outcome

- Backdoor adjustment:

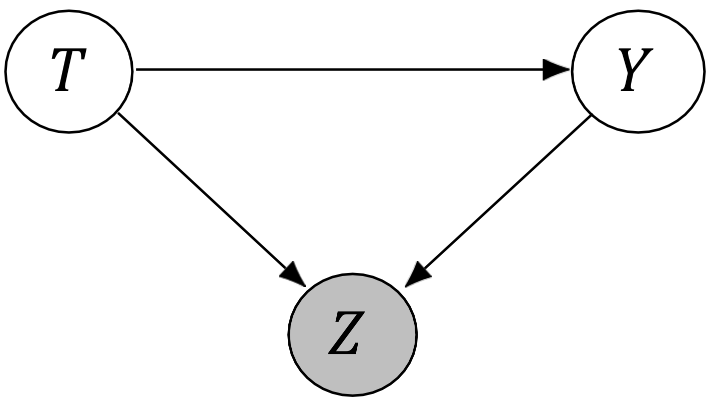
$$P(y|do(t)) = \sum_{w} P(w)P(y|t,w)$$

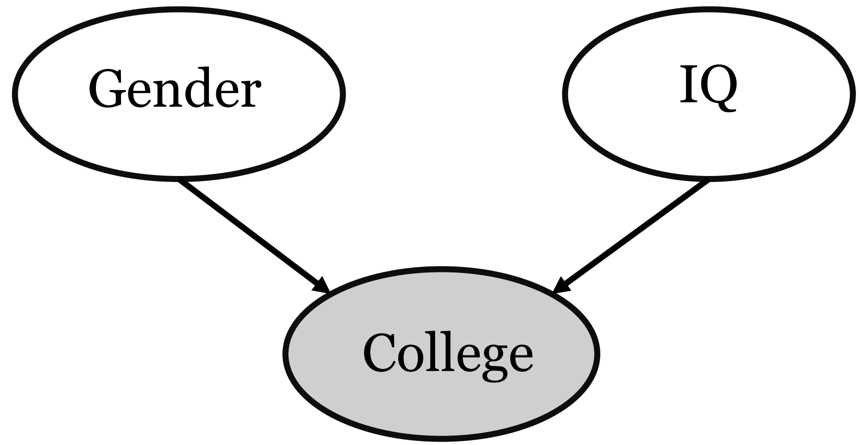
- Adjustment formula in Potential Outcome:

$$E[Y(1) - Y(0)] = E_W[E[Y|T = 1, W] - E[Y|T = 0, W]]$$

Why not condition on descendants of treatment

- Collider bias

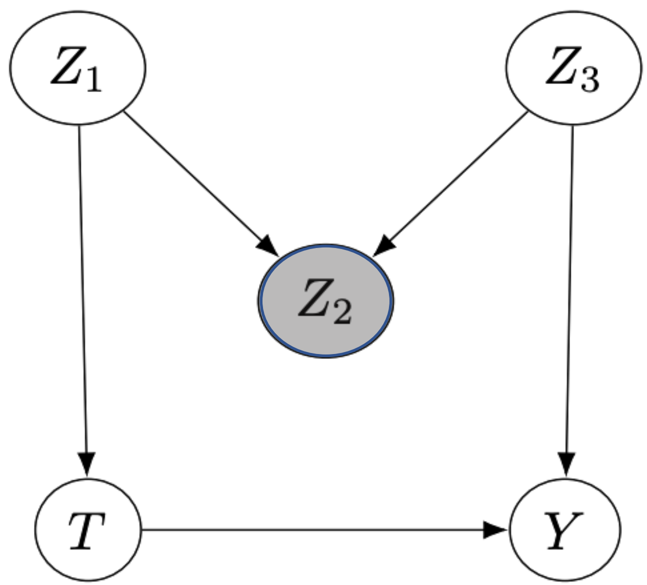




Rule: don’t condition on post-treatment covariates

M-bias

- Z_2 is a pre-treatment covariate, but adjusting for it can still lead to bias

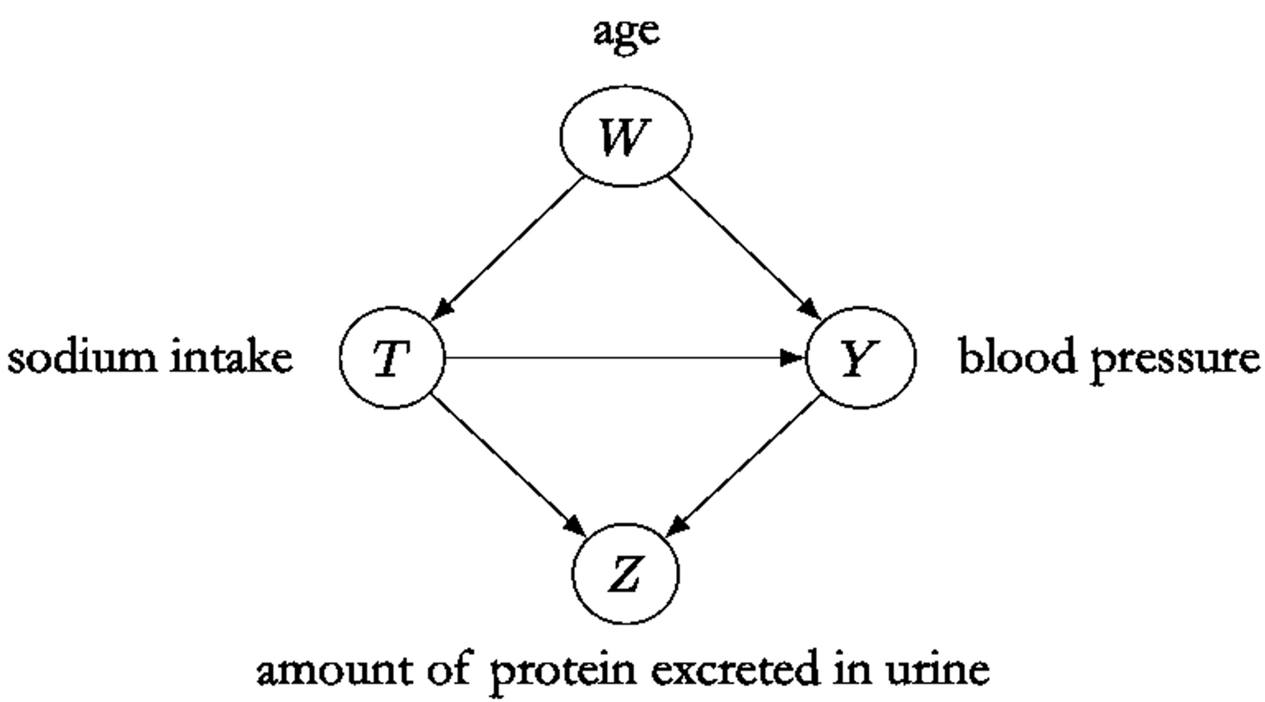


Example problem: effect of sodium intake on blood pressure

- Motivation: 46% of Americans have high blood pressure and high blood pressure is associated with increased risk of mortality [1]
- Data:
  - Outcome Y: (systolic) blood pressure (continuous)
  - Treatment T: sodium intake (1 if above 3.5 mg and 0 if below)
  - Covariates X: age and amount of protein excreted in urine
  - Simulation: we know the true ATE is 1.05

Luque-Fernandez et al. (2018), 'Educational Note: Paradoxical collider effect in the analysis of non-communicable disease epidemiological data: a reproducible illustration and web application'

Causal graph for this problem



Recall: Estimator of ATE

- True ATE: $E[Y(1) - Y(0)] = 1.05$
- Identification:
  $E[Y(1) - Y(0)] = E_X[E[Y|T = 1, X] - E[Y|T = 0, X]]$
- Estimation:

  $\frac{1}{n} \sum_x [E[Y|T = 1, x] - E[Y|T = 0, x]]$

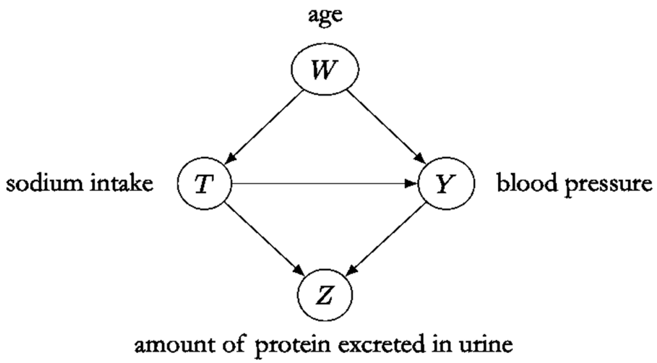
Recall: Estimator of ATE

- True ATE: $E[Y(1) - Y(0)] = 1.05$
- Identification:

$$E[Y(1) - Y(0)] = E_X[E[Y|T = 1, X] - E[Y|T = 0, X]]$$

- Estimation:

$$\frac{1}{n}\sum_x [E[Y|T = 1, x] - E[Y|T = 0, x]]$$



Estimates:
X = {} (naive): 5.33 Bias: $\frac{|5.33-1.05|}{1.05} \times 100\% = 407\%$
X = {W, Z} (last week): 0.85 Bias: $\frac{|0.85-1.05|}{1.05} \times 100\% = 19\%$
X = {W} (unbiased): 1.0502

Potential Outcome Framework v.s. SCM

- The two frameworks are logically equivalent, which means an assumption in one can always be translated to its counterpart in the other [1].
- **Potential outcome framework**: can model the causal effects of interest without knowing the complete causal graph, more straightforward.
- **SCM**: can study the causal effect of any variable. Therefore, SCMs are often preferred when learning causal relations among a set of variables.

[1] Judea Pearl. 2009. Causal inference in statistics: An overview. Statistics Surveys 3 (2009), 96–146

Reading Materials

- Judea Pearl. Causality. Cambridge University Press, 2009 --- Chapter 1: Introduction to Probabilities, Graphs, and Causal Models

Thank you!
Questions?