DUKE COMPUTER SCIENCE

CompSci 590.01

Causal Inference in Data Analysis with Applications to Fairness and Explanations

Lecture 2: Simpson's Paradox, d-Separation, Causal models

Sudeepa Roy

Resources

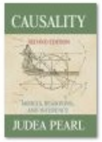
- Causality books and several articles by Judea Pearl are available at Duke library online – can be “loaned” up to 365 days – search for “Judea Pearl”

4. Causality : models, reasoning, and inference

Judea Pearl.

Pearl, Judea

Cambridge, U.K. ; New York : Cambridge University Press, 2000. / Book / Online



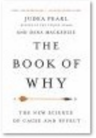
View Online

1. The book of why : The New Science of Cause and Effect

Judea Pearl.

Pearl, Judea, author

2018. / Book / Online



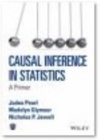
View Online

10. Causal inference in statistics : a primer

Judea Pearl, Madelyn Glymour, Nicholas Jewell.

Pearl, Judea, author

Chichester, West Sussex, UK : John Wiley & Sons Ltd, 2016. / Book / Online

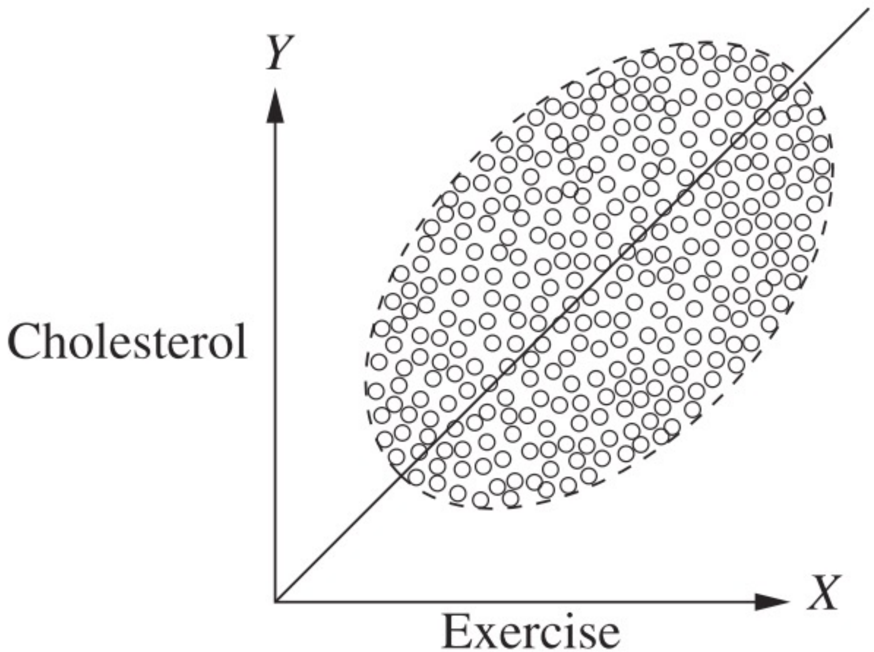


View Online

Review: Probability

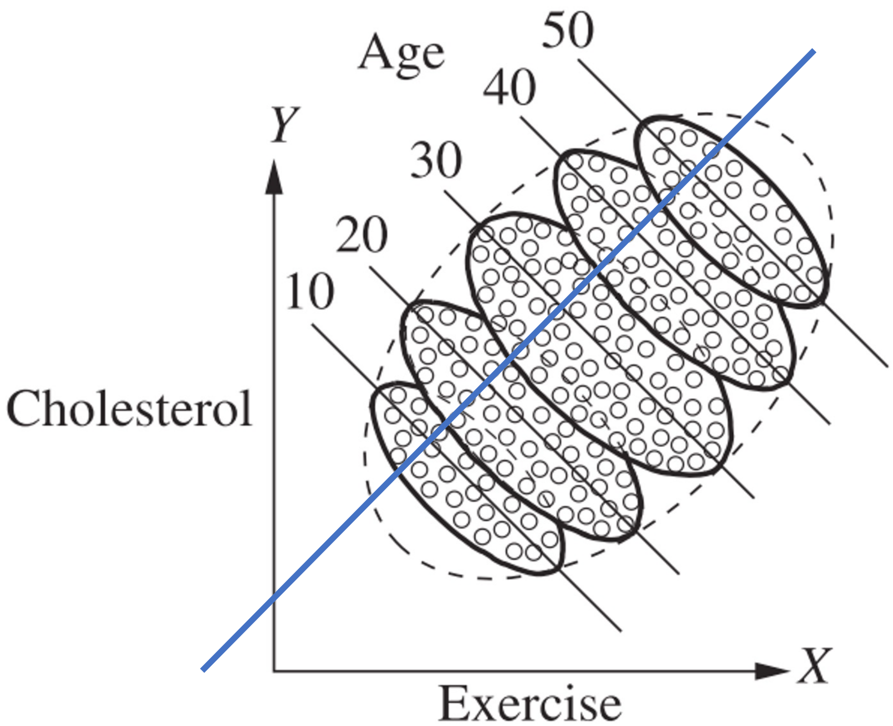
- $\Pr(X = x)$
- $\Pr(AB) = \Pr(A \wedge B)$
- $\Pr(A|B) = \Pr(A \wedge B) / \Pr(B)$
- $\Pr(A \mid B) = \Pr(B \mid A) \Pr(A) / \Pr(B)$ --- Bayes' Rule
- If A and B are independent
  - $\Pr(A \wedge B) = \Pr(A)\Pr(B)$
  - $\Pr(A|B) = \Pr(A)$

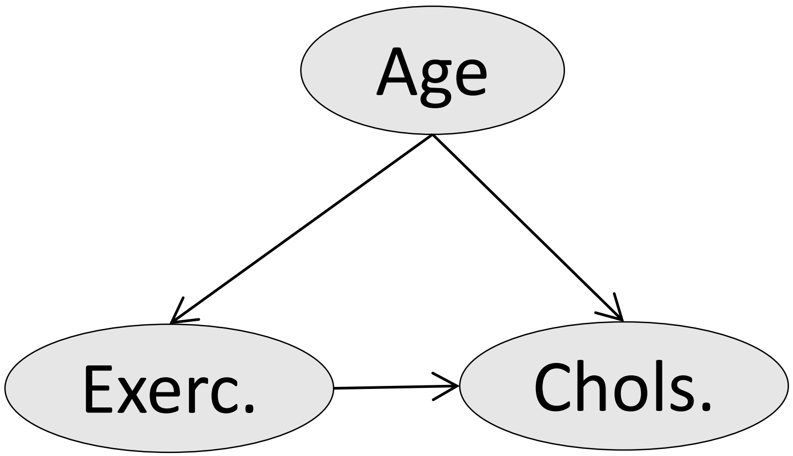
Simpson’s Paradox Revisited - 1



Graph from the Primer book

Simpson's Paradox Revisited





Need to understand the story behind the data—the causal mechanism generated, the results we see.
Need to look at age-segregated data and compare people with same age
People who exercise more have high cholesterol due to their age not due to exercising

Simpson's Paradox Revisited - 2

Table 1.1 Results of a study into a new drug, with gender being taken into account

	Drug	No drug
Men	81 out of 87 recovered (93%)	234 out of 270 recovered (87%)
Women	192 out of 263 recovered (73%)	55 out of 80 recovered (69%)
Combined data	273 out of 350 recovered (78%)	289 out of 350 recovered (83%)

Table from the Primer book

Simpson's Paradox Revisited - 2

Table 1.1 Results of a study into a new drug, with gender being taken into account

	Drug	No drug
Men	81 out of 87 recovered (93%)	234 out of 270 recovered (87%)
Women	192 out of 263 recovered (73%)	55 out of 80 recovered (69%)
Combined data	273 out of 350 recovered (78%)	289 out of 350 recovered (83%)

Estrogen has a negative effect on recovery – women are less likely to recover & women are likely to take the drug more
Why does the drug look harmful overall? If we select a person at random, more likely to be a woman, hence less likely to recover

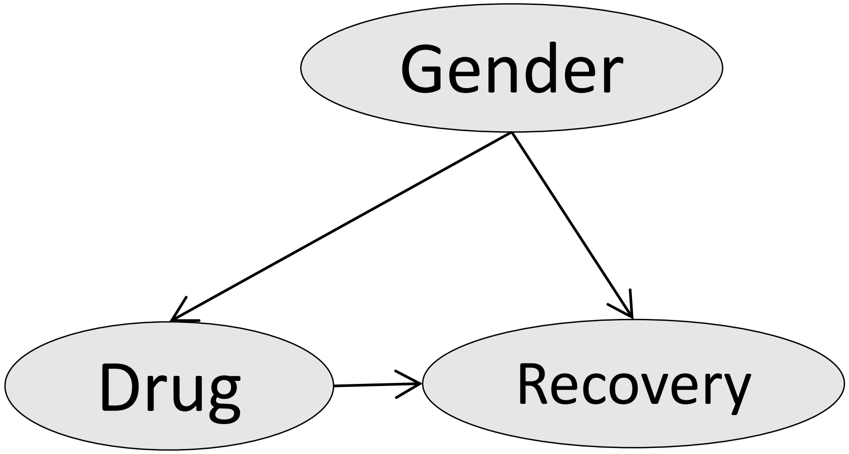
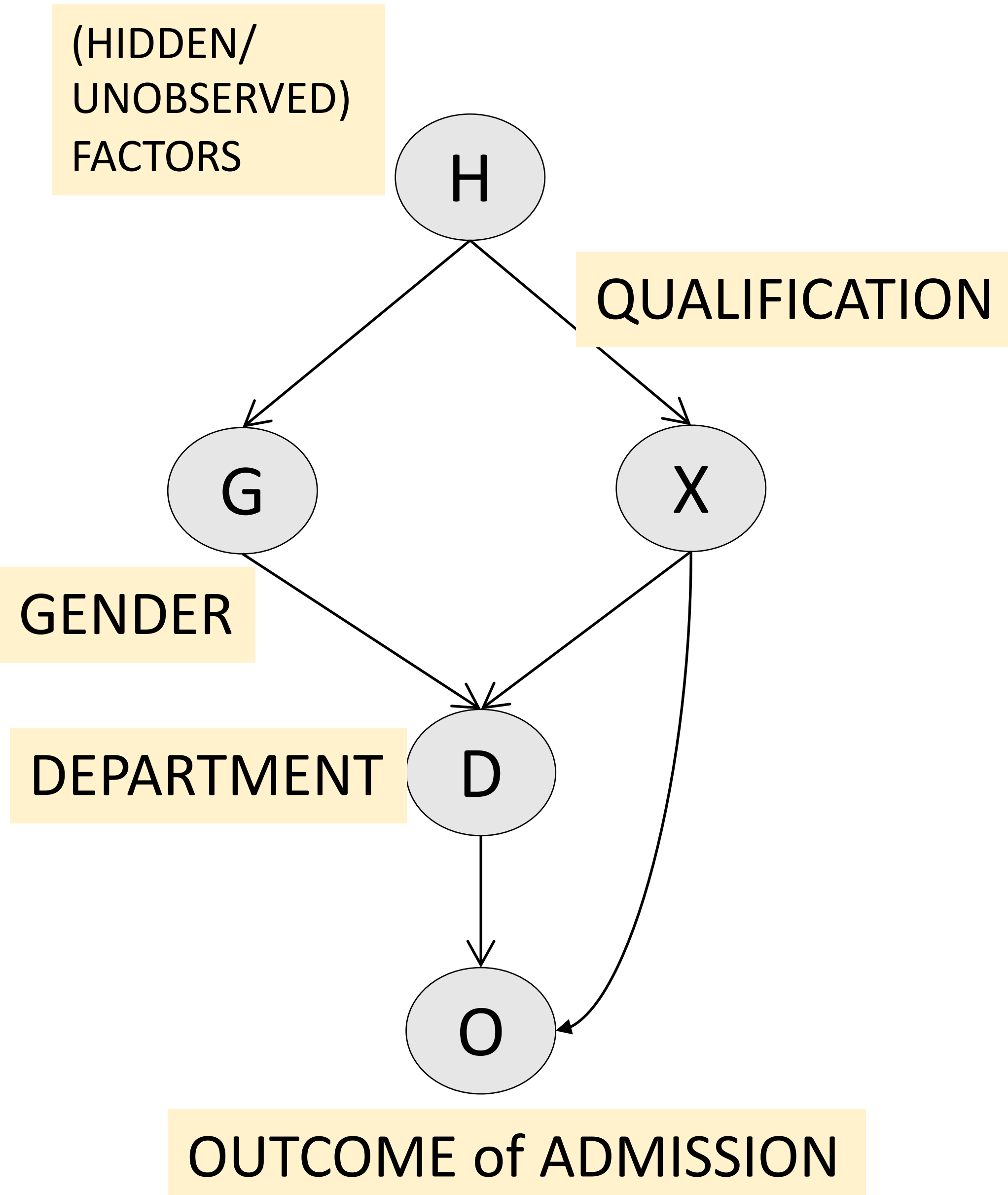


Table from the Primer book

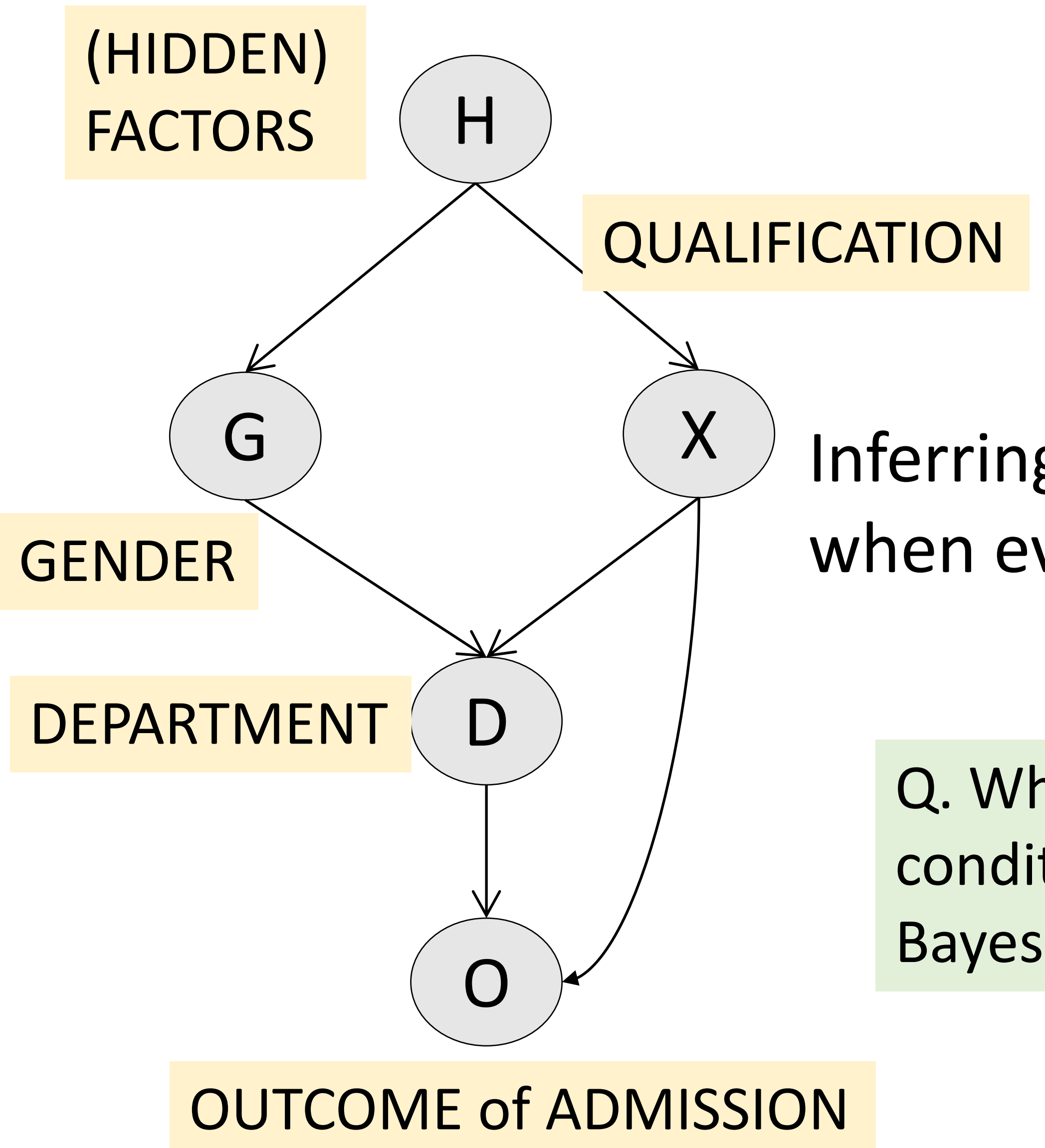
Default assumption

Review: Directed Acyclic Graphs



- Parent
  - H is a parent of X
- Child
  - X is a child of H
- Ancestor
  - H is an ancestor of D
- Descendant
  - D is a descendant of H
- Path (directed & undirected)
  - Directed: H → X → D → O
  - Undirected: X – D – G – H

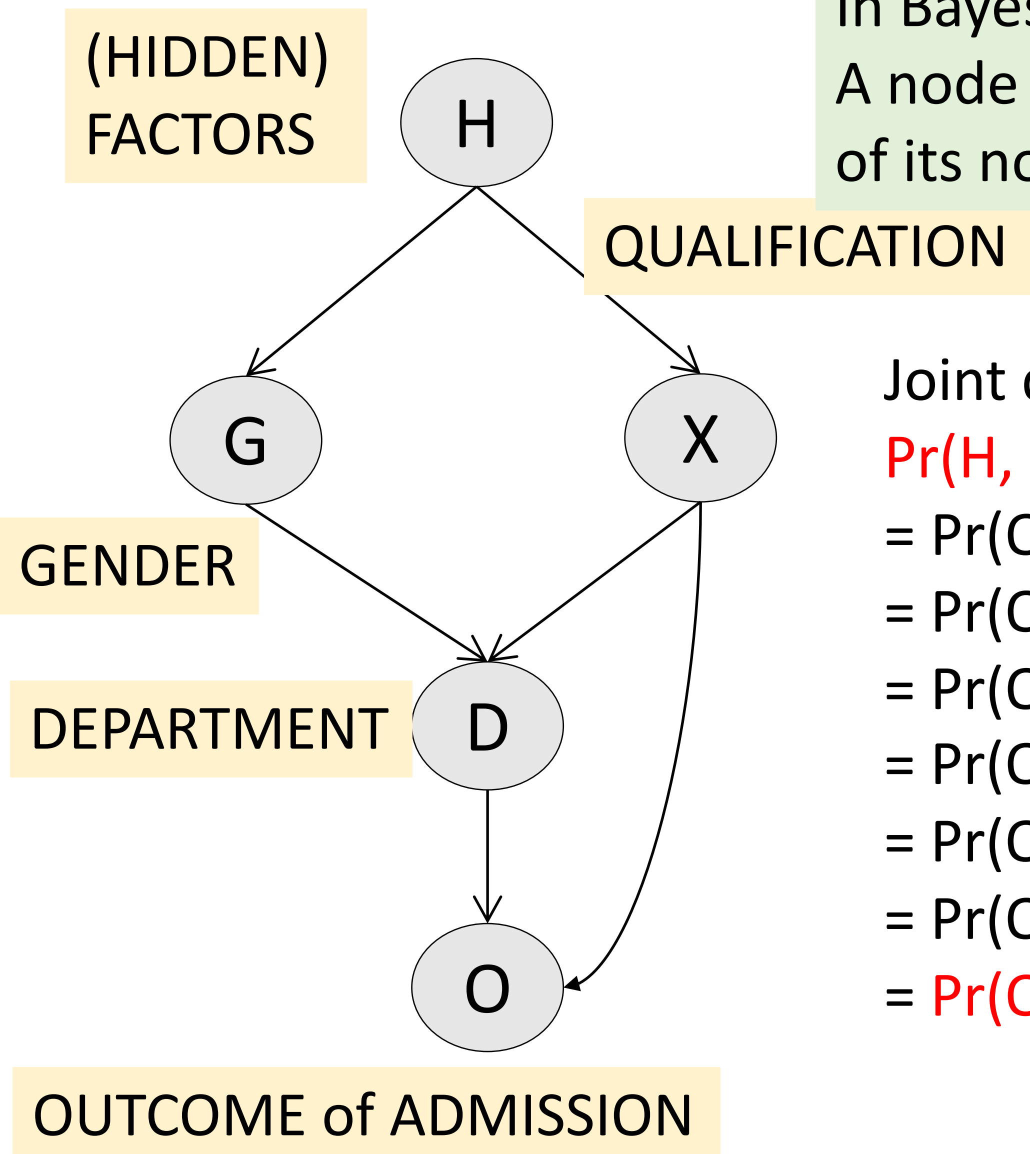
Next: Concepts from (Directed) Graphical Models



Inferring conditional independence when events are nodes

Q. What do you remember about conditional independence in Bayesian Network?

Review: Bayesian Network

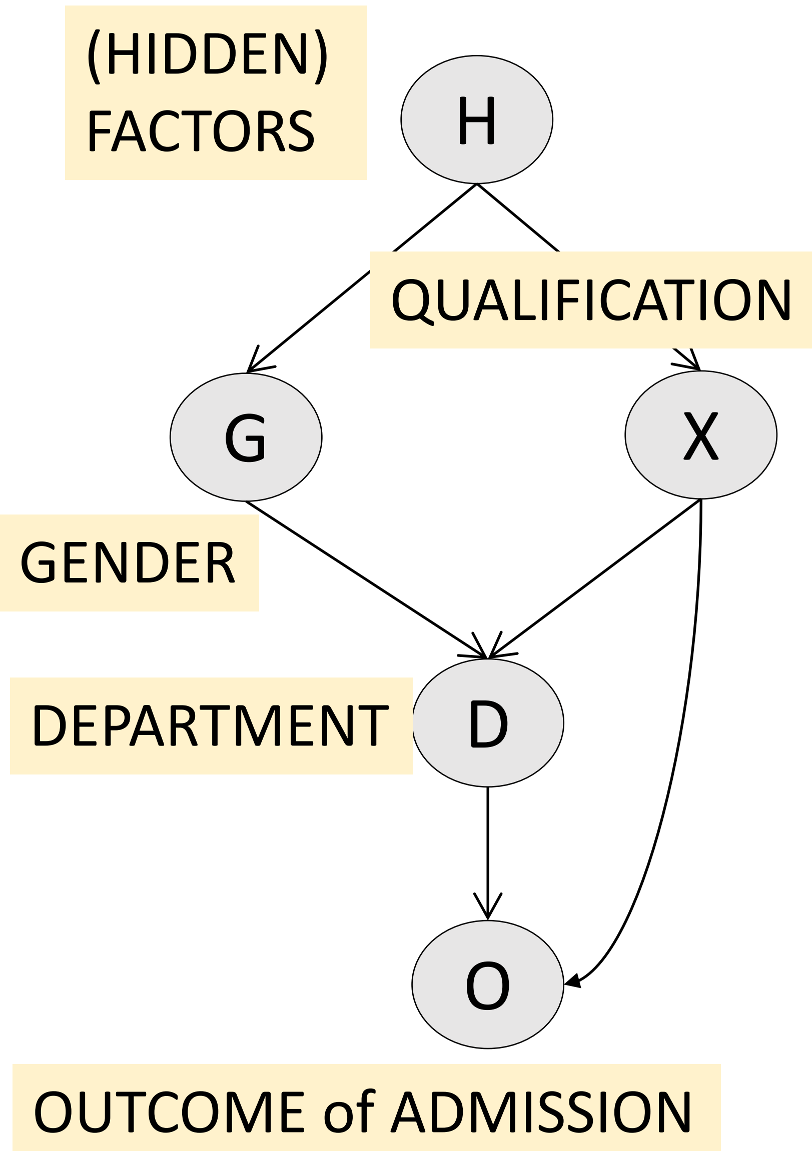


In Bayesian Networks:
A node is conditionally independent of its non-descendants given its parents

Joint distributions can be factorized

$Pr(H, G, X, D, O)$
$= Pr(O \mid DXGH)\, Pr(DXGH)$
$= Pr(O \mid DX)\, Pr(DXGH)$
$= Pr(O|DX)\, Pr(D|GXH)\, Pr(GXH)$
$= Pr(O|DX)\, Pr(D|GX)\, Pr(GXH)$
$= Pr(O|DX)\, Pr(D| GX)\, Pr(G|XH) Pr(XH)$
$= Pr(O|DX)\, Pr(D \mid GX)\, Pr(G|H)\, Pr(XH)$
$= Pr(O|DX)\, Pr(D \mid GX)\, Pr(G|H)\, Pr(X|H)\, Pr(H)$

Review: Bayesian Network



In Bayesian Networks (DAGs):
Joint distribution can be expressed as Products of $\Pr(x_i \mid pa_i)$, where pa_i denotes the parents of x_i

Joint distributions can be factorized

$$\Pr(H, G, X, D, O) = \Pr(O|DX)\Pr(D|GX)\Pr(G|H)\Pr(X|H)\Pr(H)$$

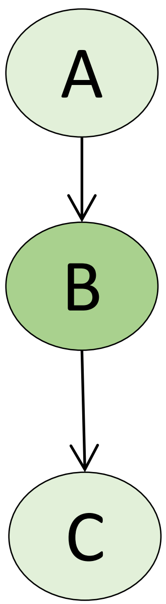
Product decomposition

Directed Graphical Models generalize Bayesian Networks

Next – Chain, Fork, Collider

We won’t cover conditional independence in undirected graphical models in this course

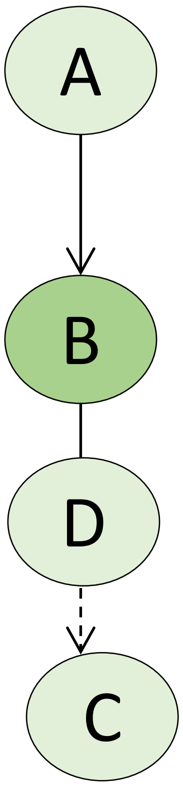
Chains



Chain

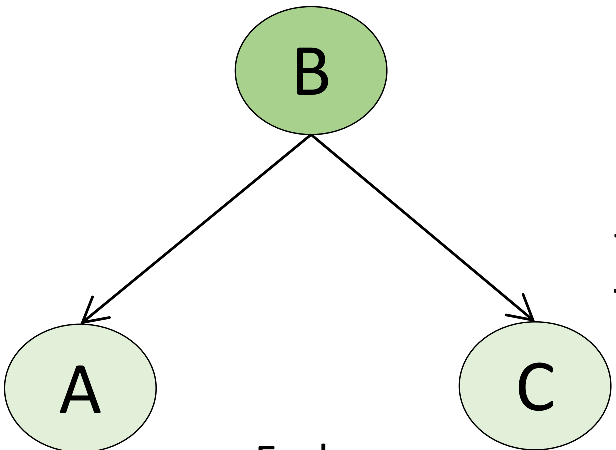
Conditional independence in chains:
Two variables A & C are conditionally independent given B
If there is only one unidirectional path between A & C
And B is any set of variables that intercepts that path

e.g.,
Switch -> circuit state -> light bulb on



A & C are (likely) correlated
A & C are independent conditioned on B

Forks



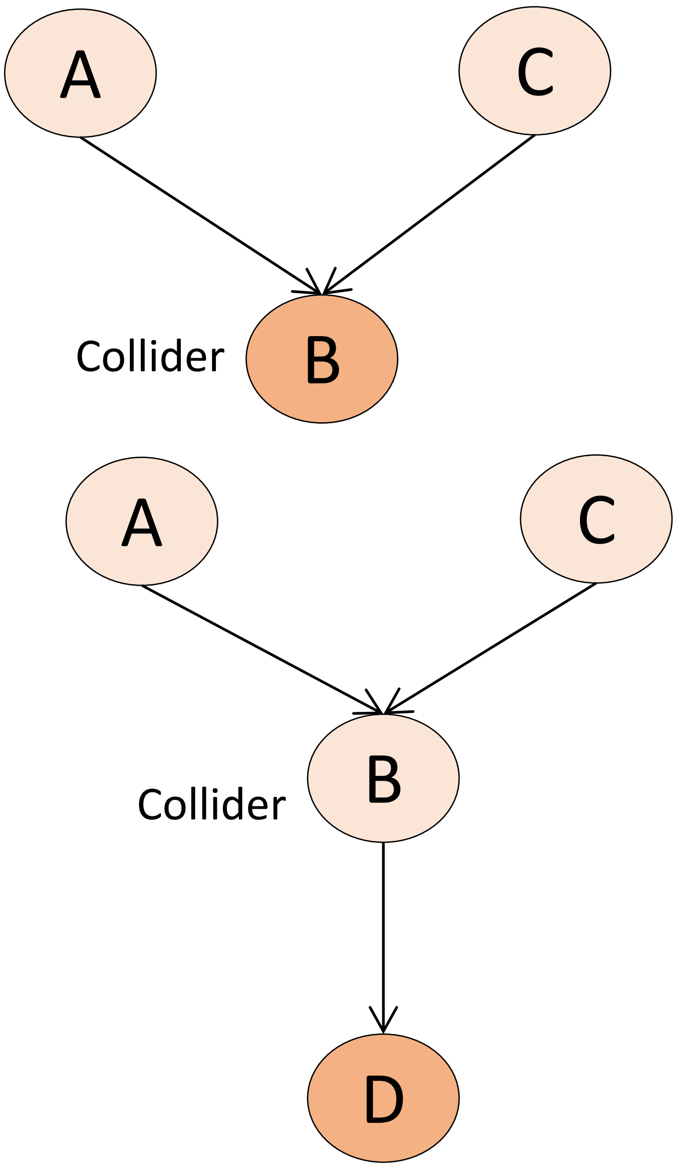
Fork

Conditional independence in forks:

If a variable B is a common cause of variables A & C, and there is only one path between A & C,
Then A & B are conditionally independent given B

A & C are correlated
A & C are independent conditioned on B

Colliders

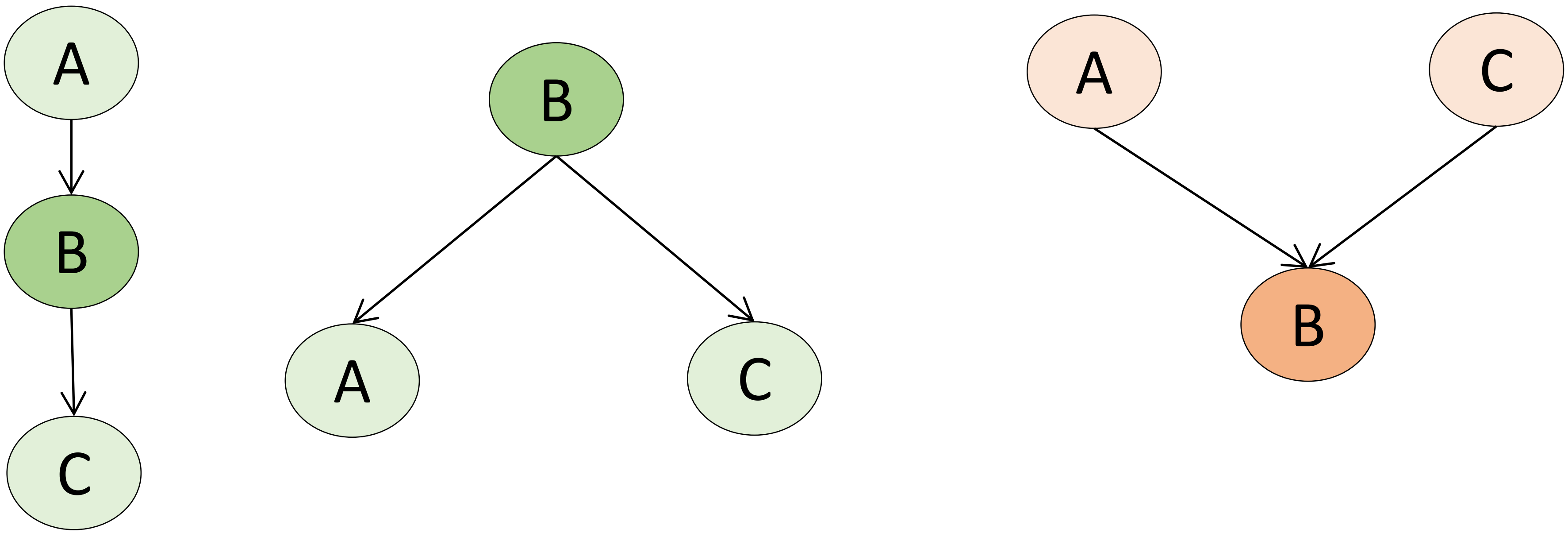


Example:
A, C: Random unbiased coin tosses independent of each other
B: Ring a bell if A = C = Head or A = C = Tail

Pr(C = Head) = ½ = Pr(C = Head | B = rings)
Pr(C = Head | A = Head, B = rings) = 1
Pr(C = Head | A = Tail, B = rings) = 0

A & C are independent
A & C are correlated conditioned on B or any descendant of B
(B “explains away” A & C)

Chain, Fork, Collider: Summary

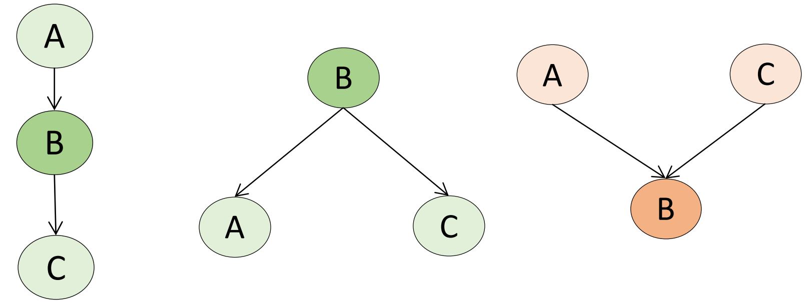


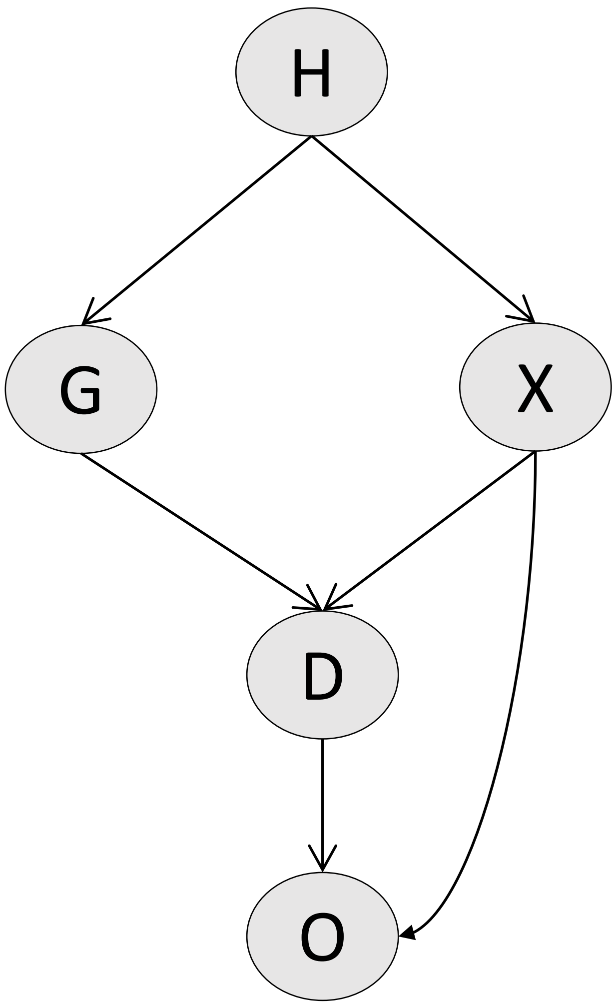
Chain Fork Collider

A & C are correlated
A & C are independent conditioned on B

A & C are independent
A & C are correlated conditioned on B or any descendant of B
(B “explains away” A & C)

Blocking a path



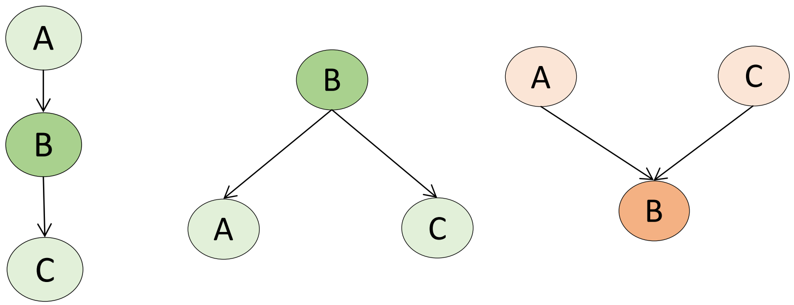


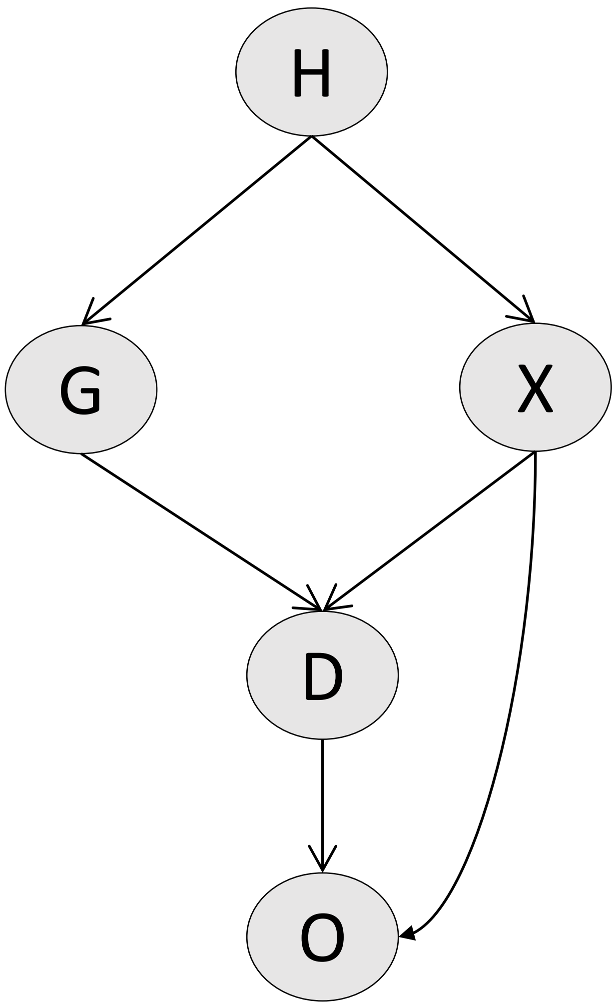
A (undirected) path p is blocked by a set of nodes Z if

- P contains a chain of the form A→B→C, or a fork of the form A←B→C such that B∈Z,

  or
- p contains a collider node B of the form A → B ← C such that neither B nor any descendants of B is in Z.

Blocking a path

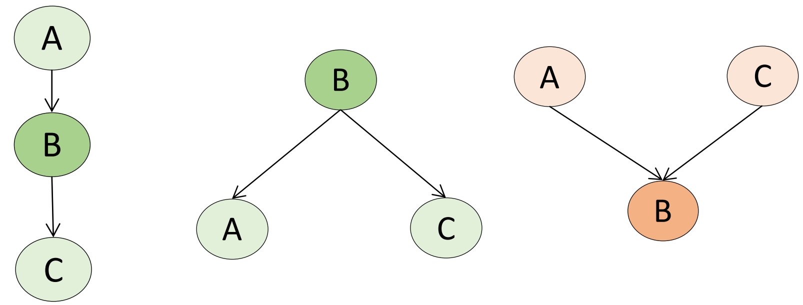


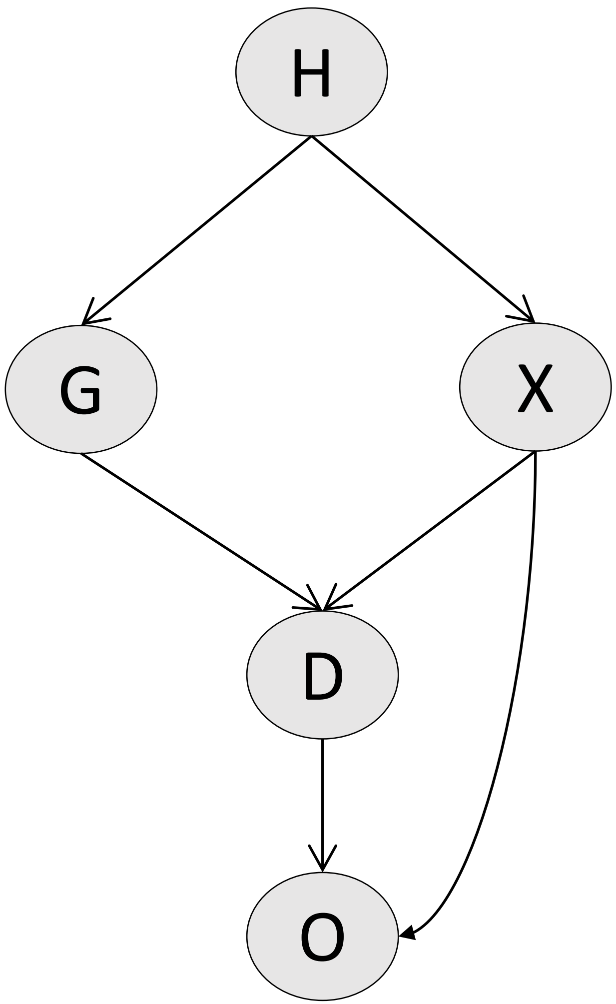


A path p is blocked by a set of nodes Z if

- P contains a chain of the form A→B→C, or a fork of the form A←B→C such that B∈Z,
  or
- p contains a collider node B of the form A → B ← C such that neither B nor any descendants of B is in Z.

Blocking a path



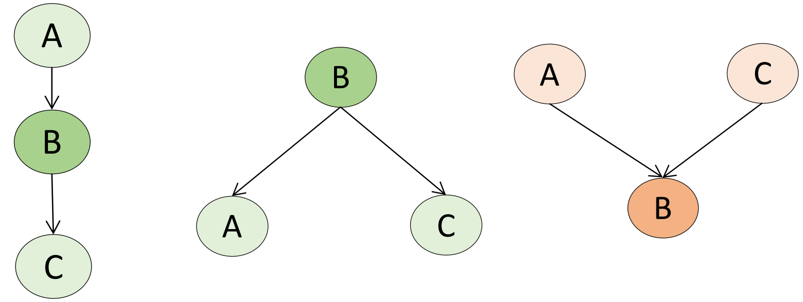


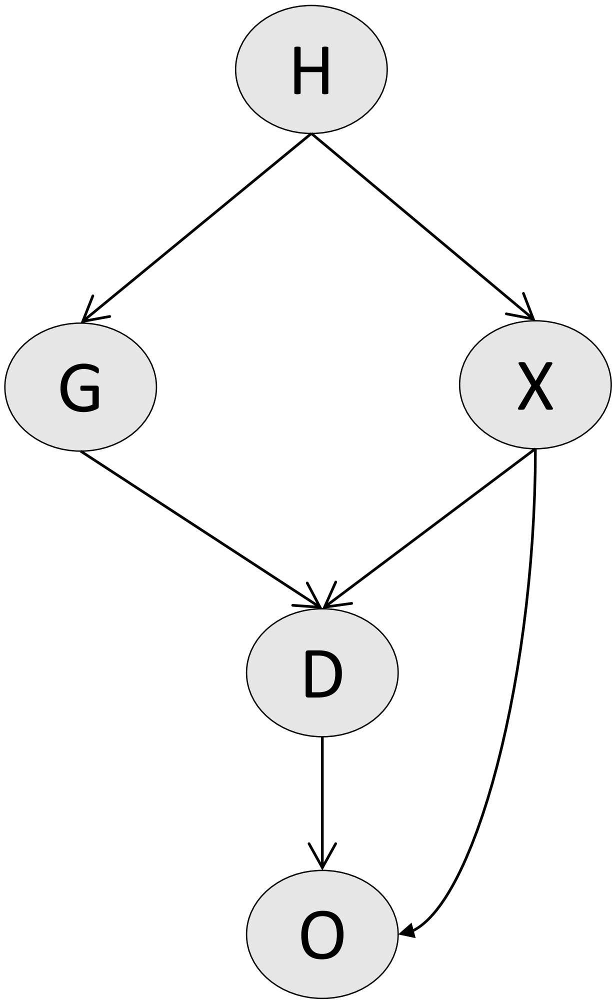
A path p is blocked by a set of nodes Z if

- P contains a chain of the form A→B→C, or a fork of the form A←B→C such that B∈Z,
  or
- p contains a collider node B of the form A → B ← C such that neither B nor any descendants of B is in Z.

X blocks the path H – X – D – O
G blocks the path D – G – H – X
D unblocks the path H – G – D – X
{DG} blocks the path H – G – D – X

d-Separation

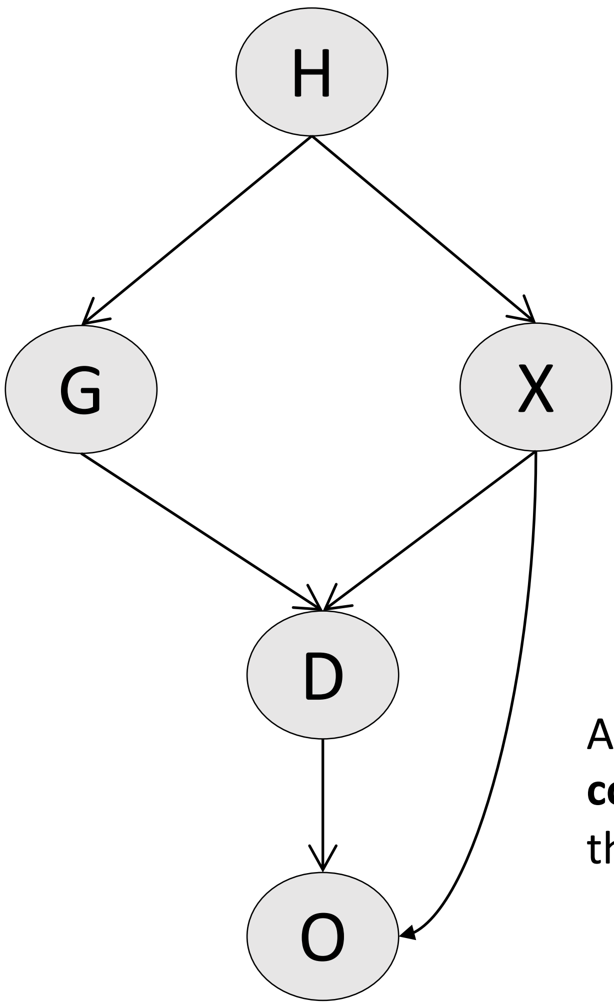




If a set of nodes Z blocks every path between two nodes X and Y, then X and Y are d-separated conditioned on Z

H and D are d-separated by {XG}
G and X are d-separated by {H}
G and X are NOT d-separated by {HD}

d-Separation and Conditional Independence



If a set of nodes Z blocks every path between two nodes X and Y, then X and Y are d-separated conditioned on Z

H and D are d-separated by {XG}
G and X are d-separated by {H}
G and X are NOT d-separated by {HD}

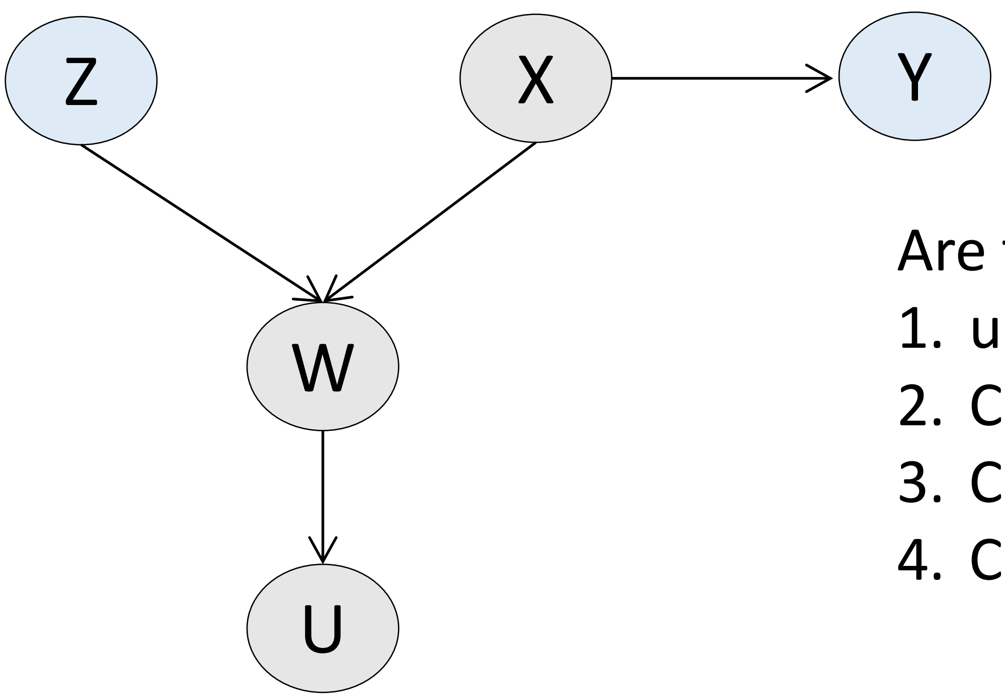
A probability distribution Pr and DAG G are **Markov compatible**: if X and Y are d-separated conditioned on Z, then X and Y are also conditionally independent given Z in Pr

H and D are conditionally independent given {XG}

Special case: Independence in Bayesian Network:
A node is conditionally independent of its non-descendants given its parents

Example-2 of d-separation

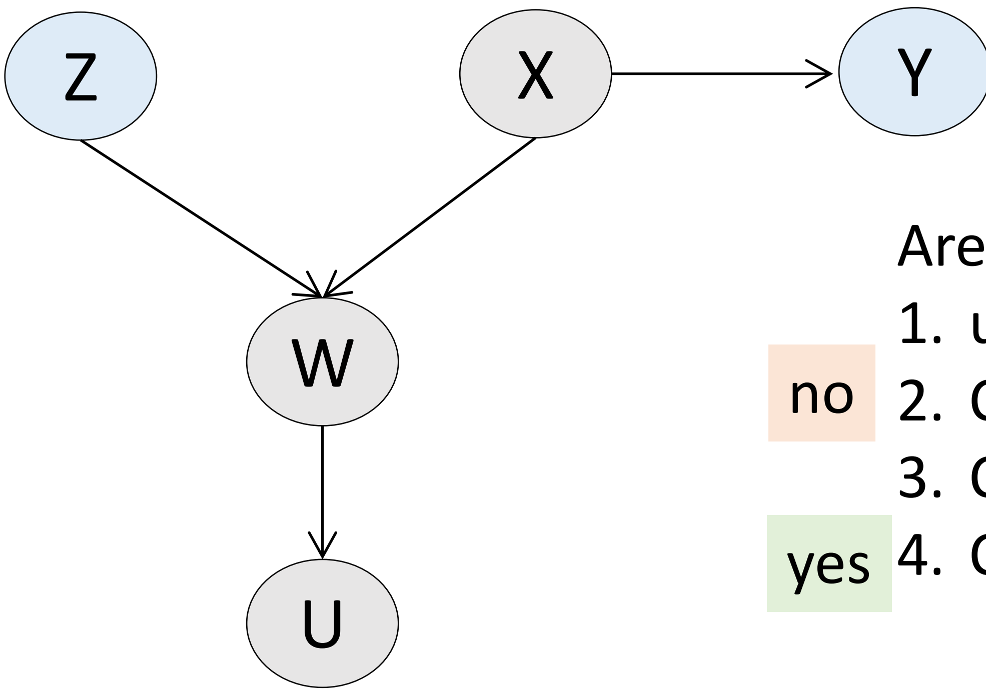
Consider nodes Z and Y



Are they d-separated
1. unconditionally?
2. Conditioned on {W}?
3. Conditioned on {U}?
4. Conditioned on {W, X}?

Example-2 of d-separation

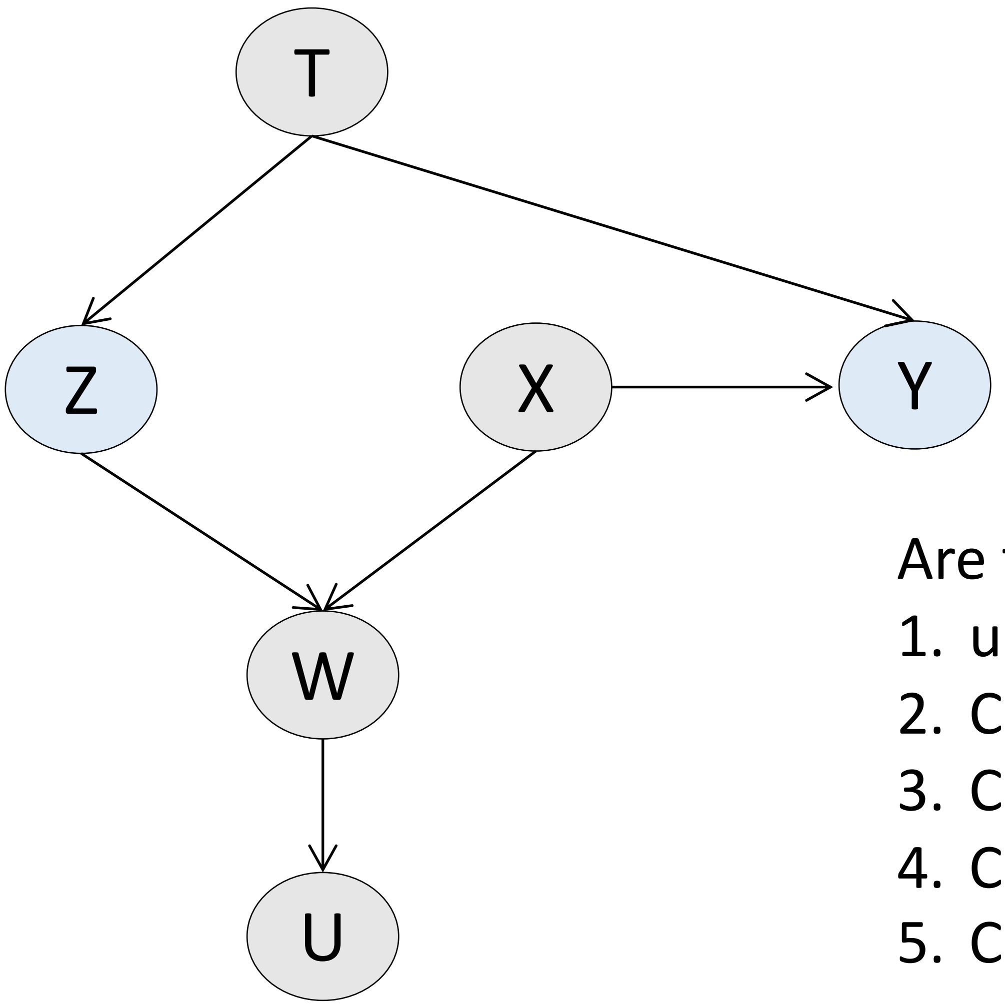
Consider nodes Z and Y



Are they d-separated

1. unconditionally? yes
2. Conditioned on {W}? no
3. Conditioned on {U}? no
4. Conditioned on {W, X}? yes

Example-3 of d-separation

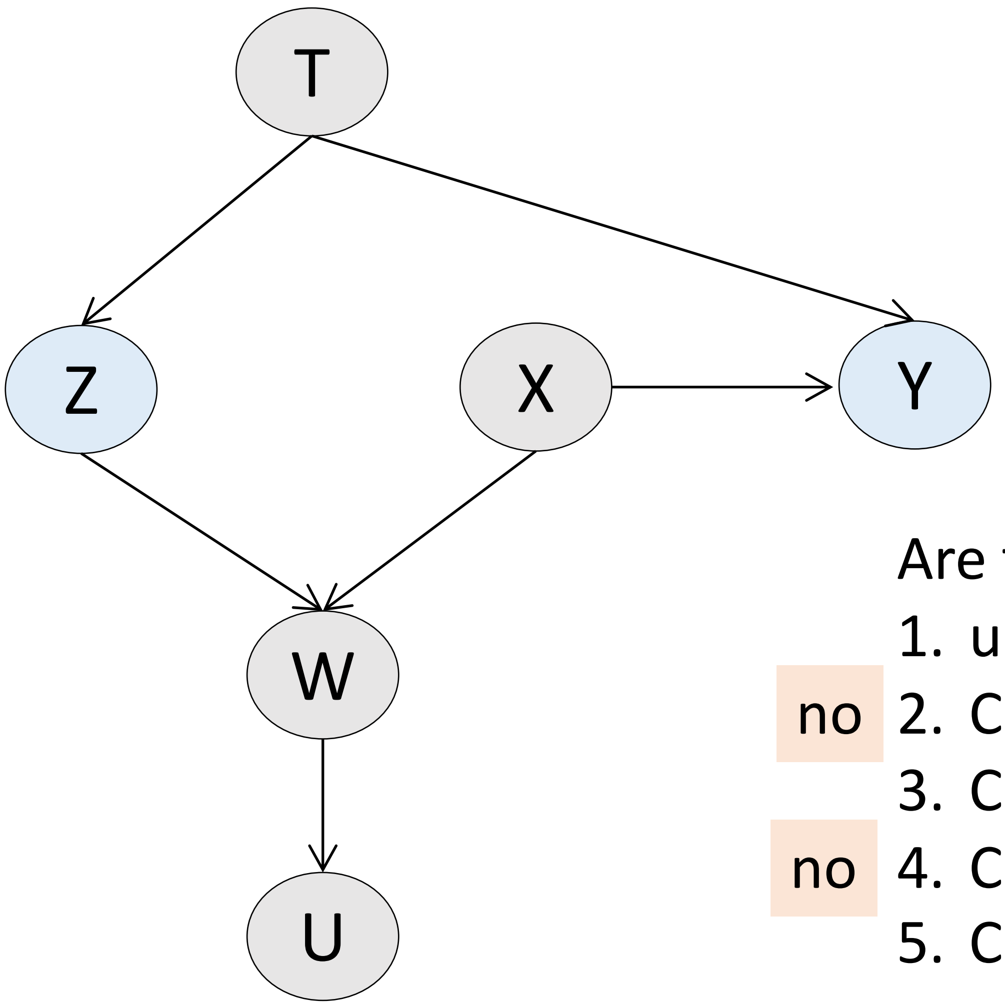


Consider nodes Z and Y

Are they d-separated

1. unconditionally?
2. Conditioned on {W, X}?
3. Conditioned on {T}?
4. Conditioned on {T, W}?
5. Conditioned on {T, W, X}?

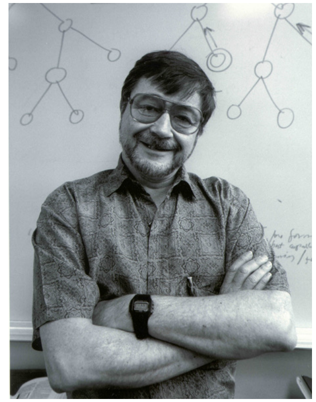
Example-3 of d-separation

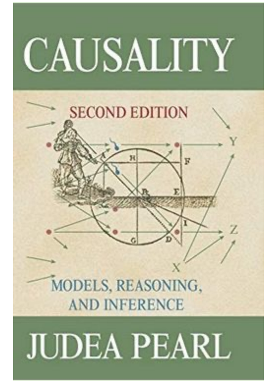


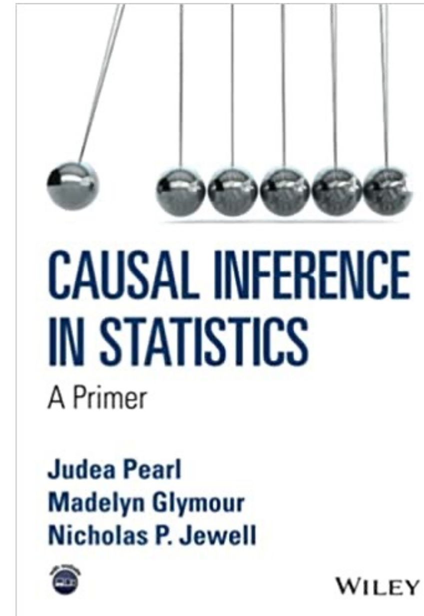
Consider nodes Z and Y

Are they d-separated

1. unconditionally? no
2. Conditioned on {W, X}? no
3. Conditioned on {T}? yes
4. Conditioned on {T, W}? no
5. Conditioned on {T, W, X}? yes







Main reference used here

Structural, Graphical, and Probabilistic Causal Models

Structural Causal Model

- $M = \langle U, V, F \rangle$
  - a set of observable or endogenous variables V that are inside the model,
  - a set of noise or exogenous variables U that are outside of the model, and
  - a set of structural equations F, one F_X for each endogenous variable $X \in V$ The structural equations assign every endogenous variable a value based on other endogenous and exogenous variables.
  - $F_X : Dom(Pa_V(X)) \times Dom(Pa_U(X)) \rightarrow Dom(X)$

Endogenous parents of X

Exogenous parents of X

Domain

Structural Causal Model as a Graphical Causal Model

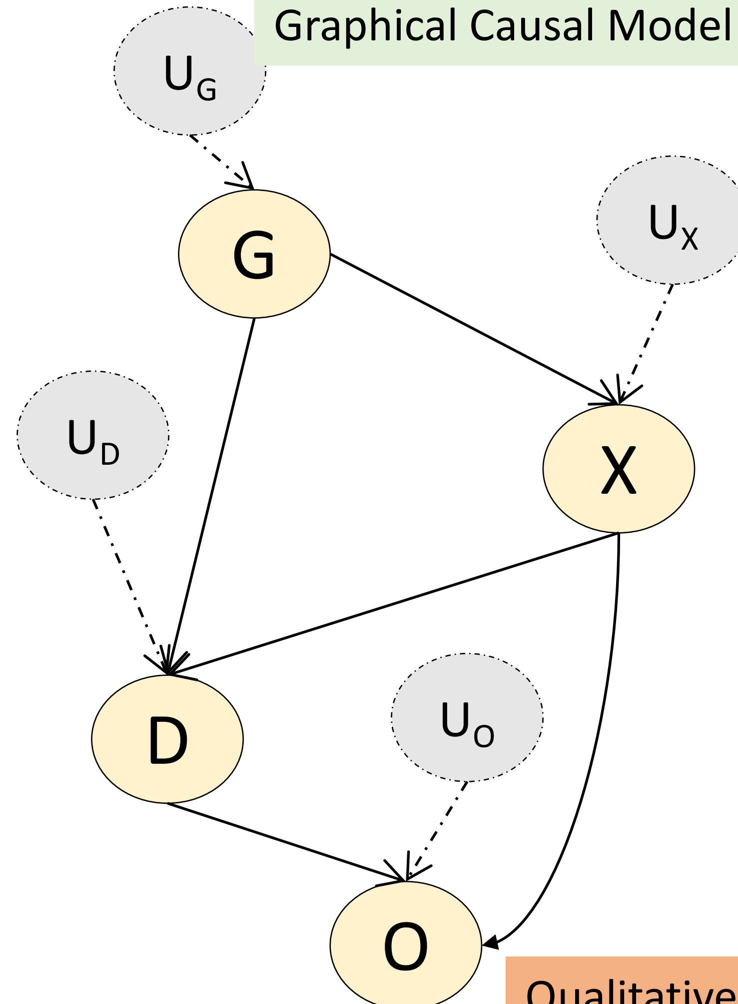
- $M = \langle U, V, F \rangle$
- Endogenous (observable) variables $V = \{G, X, D, O\}$
- Exogenous (noise) variables $U = \{U_G, U_X, U_D, U_o\}$
- Structural equations F:

$$\{G = F_G(U_G),$$
$$X = F_X(U_X, G),$$
$$D = F_D(U_D, G, X),$$
$$O = F_O(U_O, X, D)\}$$

Can be linear, exp, …

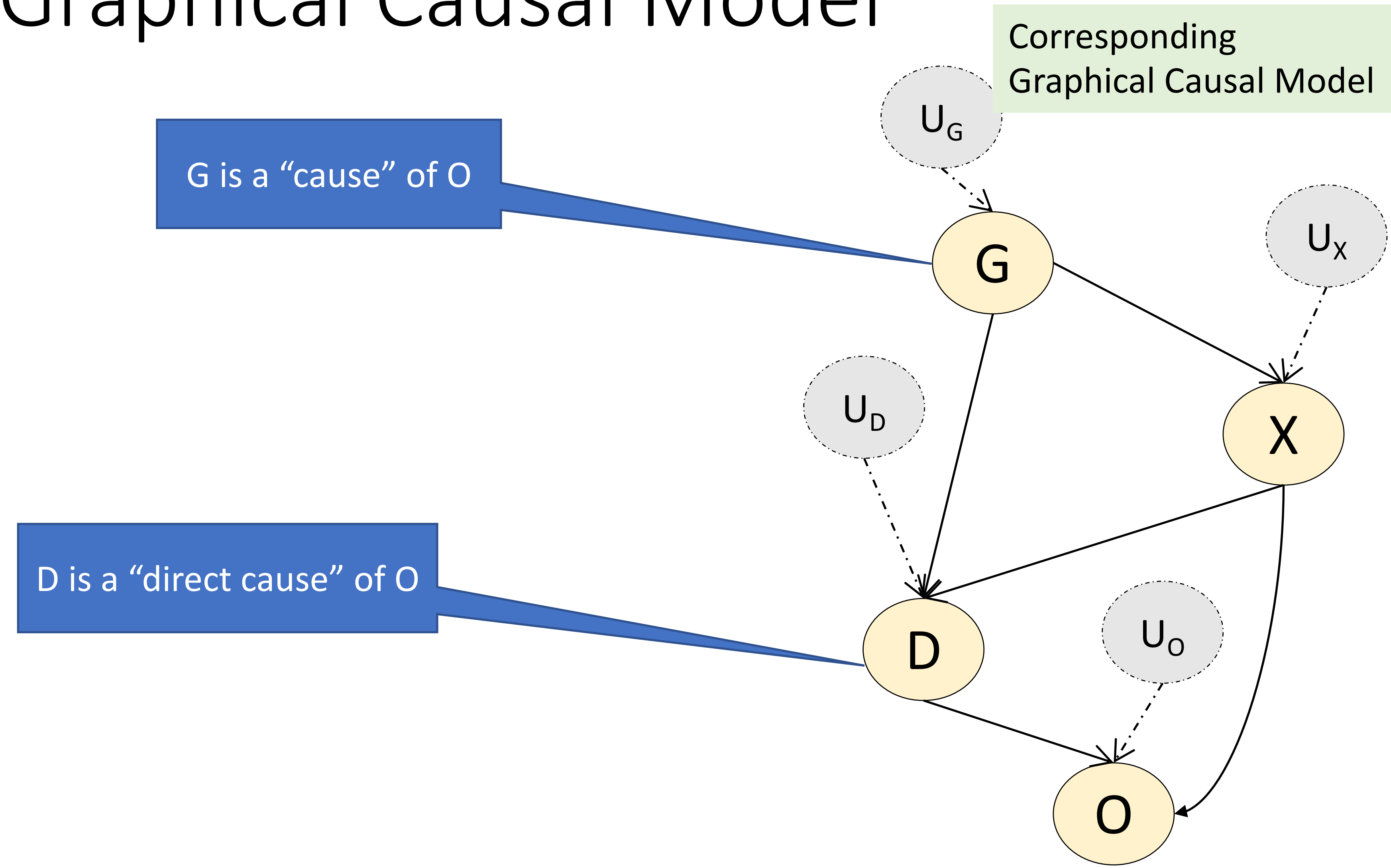
Quantitative

Corresponding Graphical Causal Model



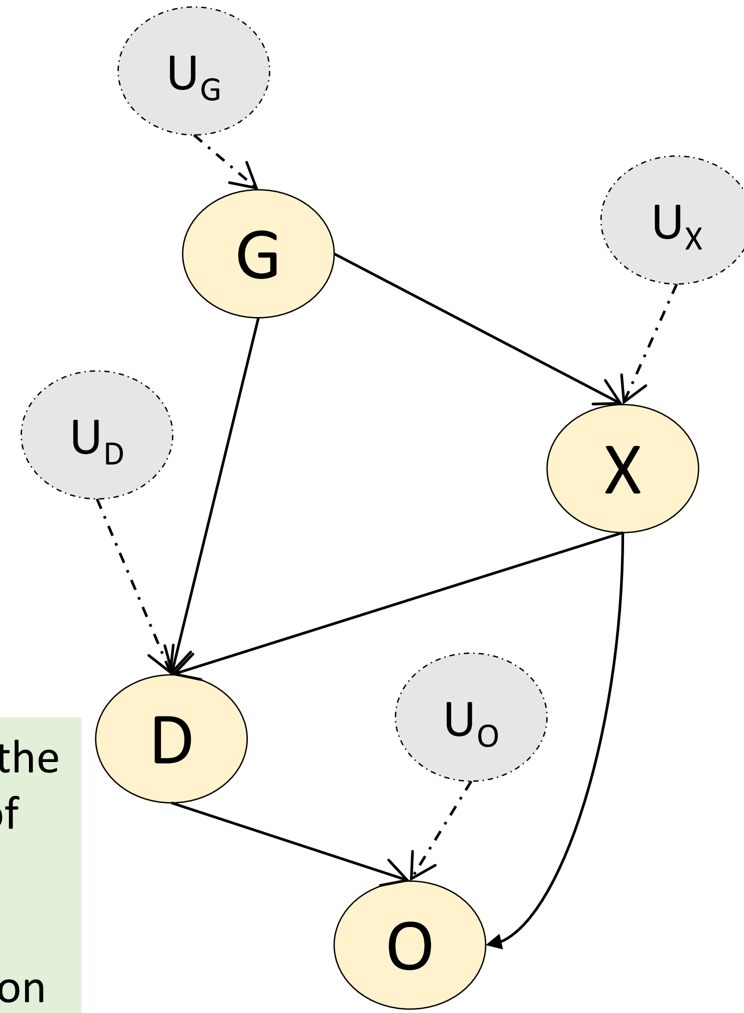
Qualitative

Structural Causal Model as a Graphical Causal Model



Structural/Graphical Causal Model to Probabilistic Model

- <M, Pr>
- M is a Structural Causal Model
- Pr is the Probability distribution
  - Satisfies Causal Markov Condition
  - Conditional independence in directed graphical models
- $Pr(X_1, X_2, ...) = \prod_i Pr(X_i \mid Pa(X_i))$



If we knew the values of the exogenous variables and the structural equations in F, we exactly know the values of endogenous V

But not in practice – so assume a probability distribution Pr(U = u), which gives a Pr distribution on V

Model for “Intervention” and “Counterfactuals”

Intervention (do-operators) and Counterfactuals

Intervention:

Change the reality by setting X to x: or $X \leftarrow x$

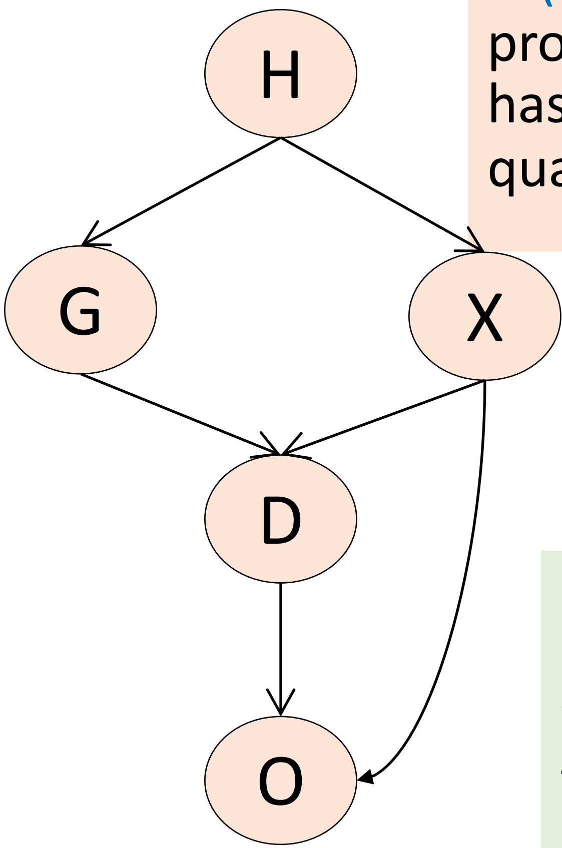
- Modeled by do-operator
- $Pr(Y = y \mid do(X = x))$

Counterfactuals:

- “If X was set to x, what would have been the value of Y”
- $Y_{X=x}$ (or Y_x) $= y$

do-operators vs. conditional probabilities

Conditioning

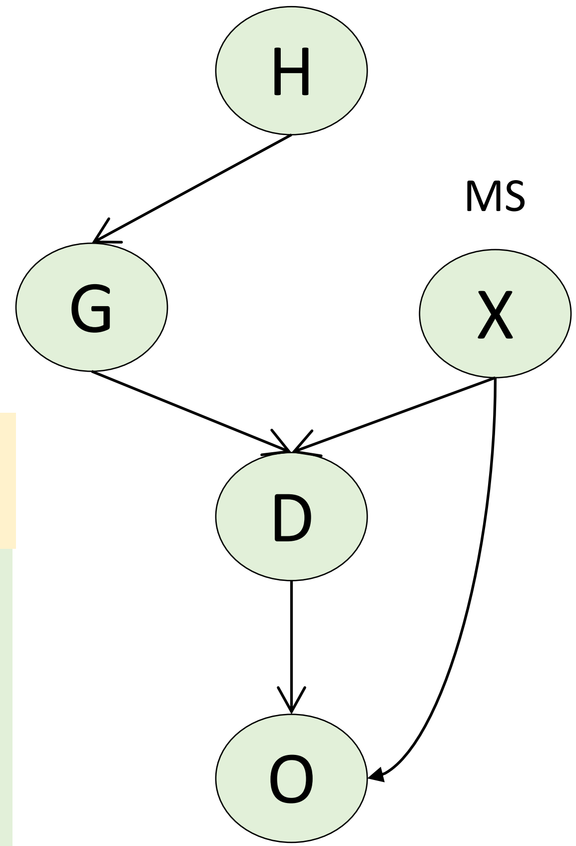


Pr(O = yes | X = MS):
probability of admission if it has been "observed" that the qualification is MS degree

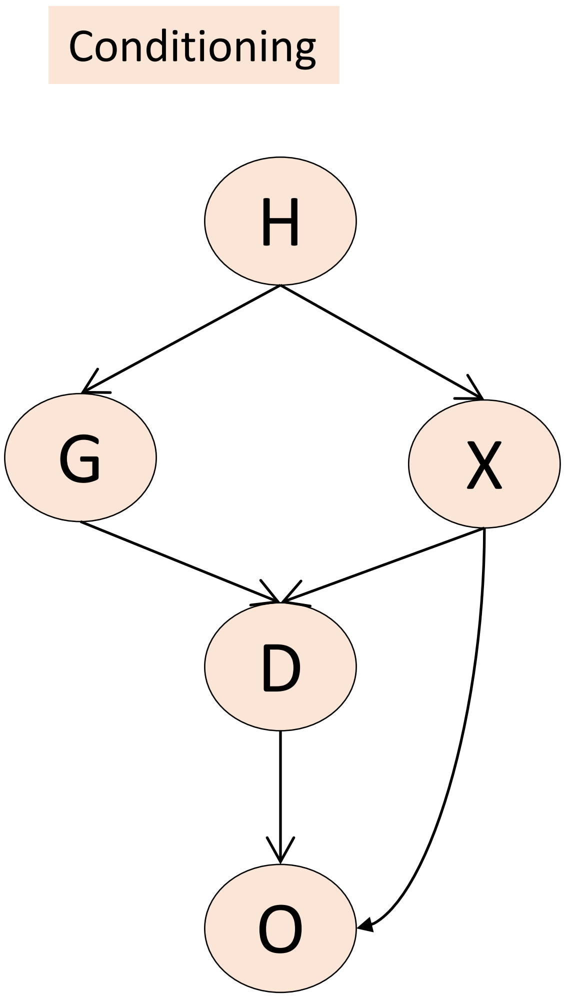
X = f(H) changes to
X = MS

Pr(O = yes | do(X = MS)):
probability of admission if we have an intervention on the world and "force" the degree to be MS

Intervention



Observational Study by Pearl's Model



Goal:

Express causal relationship as do-operators

to

conditional probabilities

- Need

1. A valid causal DAG
2. Graph Surgery
3. Observed Data