Directed Acyclic Graphs: Marginal Independence

Cornell STSCI / INFO / ILRST 3900
Fall 2023
causal3900.github.io

14 Sep 2023

Learning goals for today

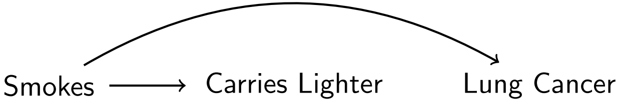
At the end of class, you will be able to

- draw a causal Directed Acyclic Graph
- enumerate edges in the graph
- read statistical dependence of nodes in the graph
- determine marginal exchangeability in the graph

After class:

- Hernán and Robins 2020 Chapter 6.1 and 6.2

Directed Acyclic Graphs (DAGs) formalize **causal beliefs**



Causal beliefs:

1) Smoking may cause you to carry a lighter
2) Smoking may cause lung cancer
3) Carrying a lighter does not cause lung cancer

Directed Acyclic Graphs (DAGs) formalize **causal beliefs**

$$\begin{tikzcd}
\text{Smokes} \arrow[r] \arrow[rr, bend left] & \text{Carries Lighter} & \text{Lung Cancer}
\end{tikzcd}$$

Directed Acyclic Graphs (DAGs) formalize **causal beliefs**

$$
\begin{tikzcd}
\textbf{Smokes} \arrow[r] \arrow[rr, bend left] & \textbf{Carries Lighter} & \textbf{Lung Cancer}
\end{tikzcd}
$$

Nodes represent random variables

Directed Acyclic Graphs (DAGs) formalize **causal beliefs**

$$
\begin{tikzcd}
\textbf{Smokes} \arrow[r] \arrow[rr, bend left] & \textbf{Carries Lighter} & \textbf{Lung Cancer}
\end{tikzcd}
$$

Nodes represent random variables

Edges represent direct causal effects

Directed Acyclic Graphs (DAGs) formalize **causal beliefs**

Nodes represent random variables

Edges represent direct causal effects

Requirement	Any node with edges to at least two other nodes must be included

Directed Acyclic Graphs (DAGs) formalize **statistical dependence**

$$
\begin{tikzcd}
\text{Smokes} \arrow[r] \arrow[rr, bend left] & \text{Carries Lighter} & \text{Lung Cancer}
\end{tikzcd}
$$

Directed Acyclic Graphs (DAGs) formalize **statistical dependence**

$$\begin{tikzcd} \text{Smokes} \arrow[r] \arrow[rr, bend left] & \text{Carries Lighter} & \text{Lung Cancer} \end{tikzcd}$$

(Smokes, Carries Lighter) are statistically dependent

— because (Smokes) causes (Carries Lighter)

Directed Acyclic Graphs (DAGs) formalize **statistical dependence**

$$\begin{tikzcd}
\text{Smokes} \arrow[r] \arrow[rr, bend left] & \text{Carries Lighter} & \text{Lung Cancer}
\end{tikzcd}$$

(Smokes, Carries Lighter) are statistically dependent

— because (Smokes) causes (Carries Lighter)

(Smokes, Lung Cancer) are statistically dependent

— because (Smokes) causes (Lung Cancer)

Directed Acyclic Graphs (DAGs) formalize **statistical dependence**

$$\begin{tikzcd} \text{Smokes} \arrow[r] \arrow[rr, bend left] & \text{Carries Lighter} & \text{Lung Cancer} \end{tikzcd}$$

(Smokes, Carries Lighter) are statistically dependent
— because (Smokes) causes (Carries Lighter)

(Smokes, Lung Cancer) are statistically dependent
— because (Smokes) causes (Lung Cancer)

(Carries Lighter, Lung Cancer) are statistically dependent
— because (Smokes) causes both

Task. Propose a rule for when two nodes are dependent

Two nodes are dependent if and only if __________

$$\begin{tikzcd}
\text{Smokes} \arrow[r] \arrow[rr, bend left] & \text{Carries Lighter} & \text{Lung Cancer}
\end{tikzcd}$$

(Smokes, Carries Lighter) are statistically dependent

— because (Smokes) causes (Carries Lighter)

(Smokes, Lung Cancer) are statistically dependent

— because (Smokes) causes (Lung Cancer)

(Carries Lighter, Lung Cancer) are statistically dependent

— because (Smokes) causes both

Task. Propose a rule for when two nodes are dependent

Two nodes are dependent if and only if __________

$$\begin{tikzcd}
\text{Smokes} \arrow[r] \arrow[rr, bend left] & \text{Carries Lighter} & \text{Lung Cancer}
\end{tikzcd}$$

Possible rule (not yet correct)	Two nodes are dependent if and only if they are connected by a path

Task. Propose a rule for when two nodes are dependent

Two nodes are dependent if and only if __________

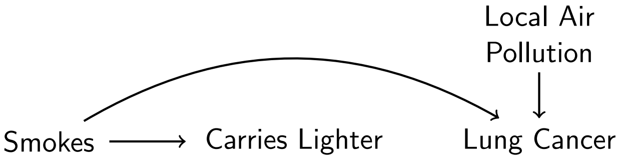
$$\begin{tikzcd} \text{Smokes} \arrow[r] \arrow[rr, bend left] & \text{Carries Lighter} & \text{Lung Cancer} \end{tikzcd}$$

Possible rule (not yet correct) — Two nodes are dependent if and only if they are connected by a path

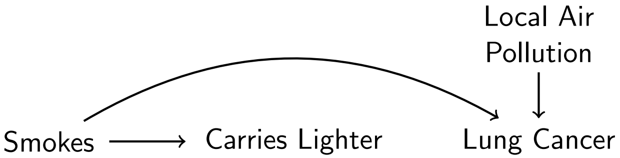
Path — A sequence of edges connecting two nodes

Smokes $\rightarrow$ Carries Lighter
Smokes $\rightarrow$ Lung Cancer
Carries Lighter $\leftarrow$ Smokes $\rightarrow$ Lung Cancer

Task. Propose a rule for when two nodes are dependent

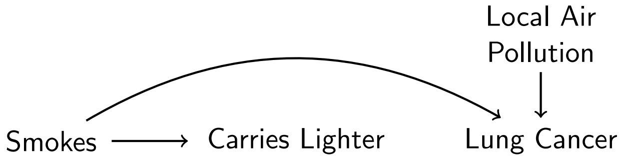


Task. Propose a rule for when two nodes are dependent



(Local Air Pollution) causes (Lung Cancer)

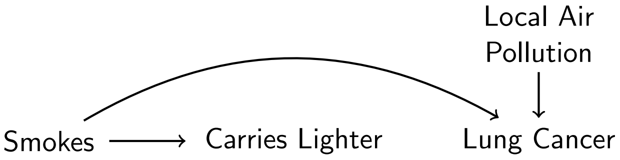
Task. Propose a rule for when two nodes are dependent



(Local Air Pollution) causes (Lung Cancer)

There are no common causes of (Smokes, Local Air Pollution)

Task. Propose a rule for when two nodes are dependent



(Local Air Pollution) causes (Lung Cancer)

There are no common causes of (Smokes, Local Air Pollution)

There is a path: (Smokes) $\rightarrow$ (Lung Cancer) $\leftarrow$ (Local Air Pollution)

Task. Propose a rule for when two nodes are dependent

$$
\begin{tikzcd}
 & & \text{Local Air Pollution} \arrow[d] \\
\text{Smokes} \arrow[r] \arrow[rr, bend left] & \text{Carries Lighter} & \text{Lung Cancer}
\end{tikzcd}
$$

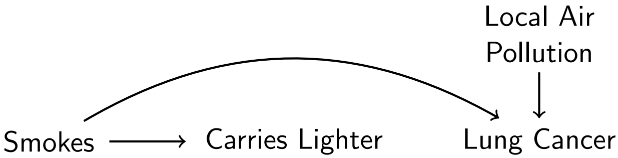
(Local Air Pollution) causes (Lung Cancer)

There are no common causes of (Smokes, Local Air Pollution)

There is a path: (Smokes) $\rightarrow$ (Lung Cancer) $\leftarrow$ (Local Air Pollution)

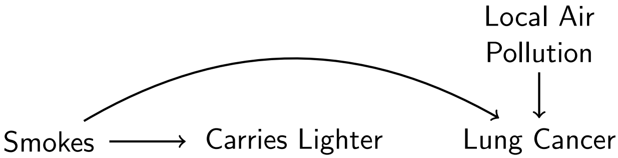
Is (Smokes) statistically related to (Local Air Pollution)?

Task. Propose a rule for when two nodes are dependent



Lung cancer is a **collider** on the path
(Smokes) $\rightarrow$ (Lung Cancer) $\leftarrow$ (Local Air Pollution)

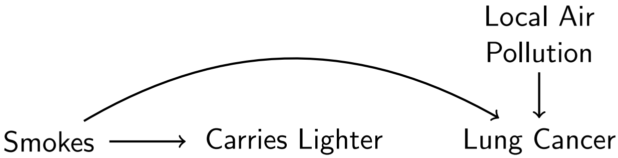
Task. Propose a rule for when two nodes are dependent



Lung cancer is a **collider** on the path
(Smokes) $\rightarrow$ (Lung Cancer) $\leftarrow$ (Local Air Pollution)

Collider A node on a path where two edges collide $\rightarrow \bullet \leftarrow$

Task. Propose a rule for when two nodes are dependent

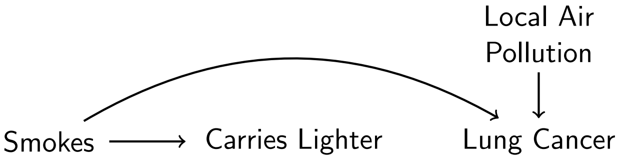


Lung cancer is a **collider** on the path
(Smokes) $\rightarrow$ (Lung Cancer) $\leftarrow$ (Local Air Pollution)

Collider A node on a path where two edges collide $\rightarrow \bullet \leftarrow$

A collider **blocks the path**.
A blocked path does not create statistical dependence.

Task. Propose a rule for when two nodes are dependent



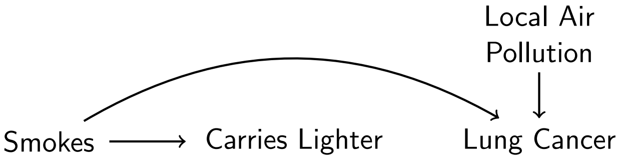
Lung cancer is a **collider** on the path
(Smokes) $\rightarrow$ (Lung Cancer) $\leftarrow$ (Local Air Pollution)

Collider A node on a path where two edges collide $\rightarrow \bullet \leftarrow$

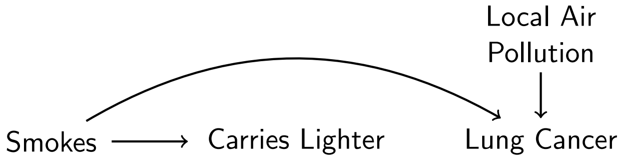
A collider **blocks the path**.
A blocked path does not create statistical dependence.

Intuition: If two variables affect one outcome,
that does not make those two variables related

Task. Propose a rule for when two nodes are dependent



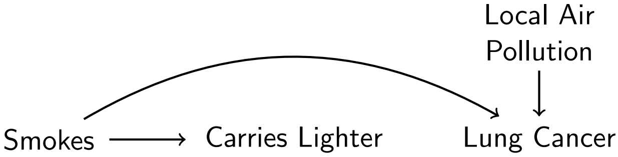
Task. Propose a rule for when two nodes are dependent



Possible rule (not yet correct)

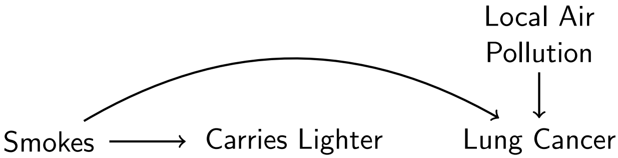
Two nodes are dependent if and only if they are connected by a path

Task. Propose a rule for when two nodes are dependent



Possible rule (not yet correct) Two nodes are dependent if and only if they are connected by an unblocked path

Task. Propose a rule for when two nodes are dependent



Rule Two nodes are dependent if and only if they are connected by an unblocked path

DAGs help us reason about exchangeability

DAGs help us reason about exchangeability

DAGs tell us why two variables are statistically dependent

- A set of unblocked paths

DAGs help us reason about exchangeability

DAGs tell us why two variables are statistically dependent

- A set of unblocked paths

Exchangeability requires statistical independence: $A \perp\!\!\!\perp Y^a$

- Exchangeability holds if the only reason A and Y are related is the causal effect of A on Y

DAGs help us reason about exchangeability

DAGs tell us why two variables are statistically dependent

- A set of unblocked paths

Exchangeability requires statistical independence: $A \perp\!\!\!\perp Y^a$

- Exchangeability holds if the only reason A and Y are related is the causal effect of A on Y

Exchangeability holds if all unblocked paths between A and Y are causal paths that point from A to Y

DAGs help us reason about exchangeability

DAGs help us reason about exchangeability

Procedure

1) List all paths between A to Y
2) Cross out the blocked paths
3) Exchangeability holds if all remaining paths are causal

DAGs help us reason about exchangeability

Procedure

1) List all paths between A to Y
2) Cross out the blocked paths
3) Exchangeability holds if all remaining paths are causal

DAG 1

$$\begin{tikzcd} & U \arrow[d] \\ A \arrow[r] & Y \end{tikzcd}$$

DAG 2

$$\begin{tikzcd} U \arrow[d] \arrow[dr] & \\ A \arrow[r] & Y \end{tikzcd}$$

DAG 3

$$\begin{tikzcd} & U \arrow[dr] & \\ A \arrow[ur] \arrow[rr] & & Y \end{tikzcd}$$

DAG 4

$$\begin{tikzcd} & U \\ A \arrow[ur] \arrow[r] & Y \arrow[u] \end{tikzcd}$$

DAG 5

$$\begin{tikzcd} U_1 \arrow[dd] \arrow[dr] & & U_2 \arrow[dl] \arrow[dd] \\ & U_3 & \\ A \arrow[rr] & & Y \end{tikzcd}$$

Learning goals for today

At the end of class, you will be able to

- draw a causal Directed Acyclic Graph
- enumerate edges in the graph
- read statistical dependence of nodes in the graph
- determine marginal exchangeability in the graph

After class:

- Hernán and Robins 2020 Chapter 6.1 and 6.2