Graphical Models for Causal Inference

Karthika Mohan and Judea Pearl

University of California, Los Angeles

August 14, 2012
Introduction

Why do we need graphs?

Figure: Motivating Example
Introduction

Variables in the study:
- Season
- Sprinkler
- Rain
- Wetness of pavement (Wet)
- Slipperiness of pavement (Slippery)

Figure: Motivating Example
## Introduction

**Figure: Motivating Example**

<table>
<thead>
<tr>
<th># Variables</th>
<th>Table size</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>32</td>
</tr>
<tr>
<td>6</td>
<td>64</td>
</tr>
<tr>
<td>7</td>
<td>128</td>
</tr>
<tr>
<td>8</td>
<td>256</td>
</tr>
<tr>
<td>9</td>
<td>512</td>
</tr>
<tr>
<td>10</td>
<td>1,024</td>
</tr>
<tr>
<td>20</td>
<td>1,048,576</td>
</tr>
<tr>
<td>30</td>
<td>1,073,741,824</td>
</tr>
</tbody>
</table>
Introduction

**Figure: DAG Representation**

- **Conditional Probability Distributions**
  - $P(X_1) : 2$
  - $P(X_3|X_1) : 4$
  - $P(X_2|X_1) : 4$
  - $P(X_4|X_2, X_3) : 8$
  - $P(X_5|X_4) : 4$

Total # of Table Entries = 22
Graphs: Notations

Figure: Bayesian Network representing dependencies

- Adjacent Nodes
- Root and Leaf Nodes
- Skeleton
- Path
- Kinship Terminology
Graphs: Notations

Figure: Bayesian Network representing dependencies
Background Factors & Bi-directed Edges

Figure: (a) Causal Model with background factors (b) & (c) Causal Model with correlated background factors
Decomposing joint distribution-$P(V)$

How would you decompose joint distribution $P(V)$ into smaller distributions?

By applying Chain rule

Let $X_1, X_2, ..., X_n$ be any arbitrary ordering of nodes in a DAG.

$P(x_1, x_2, ..., x_n) = \prod_j P(x_j | x_1, ..., x_{j-1})$

Is it possible that conditional probability of some variable $X_j$ is not sensitive to all its predecessors? Yes!
Decomposing joint distribution-$P(V)$

How would you decompose joint distribution $P(V)$ into smaller distributions?

By applying Chain rule
Decomposing joint distribution-$P(V)$

How would you decompose joint distribution $P(V)$ into smaller distributions?

By applying Chain rule

Let $X_1, X_2, ..., X_n$ be any arbitrary ordering of nodes in a DAG.

$$P(x_1, x_2, ..., x_n) = \prod_j P(x_j | x_1, ..., x_{j-1})$$

Is it possible that conditional probability of some variable $X_j$ is not sensitive to all its predecessors?

Yes!
Decomposing joint distribution-$P(V)$

How would you decompose joint distribution $P(V)$ into smaller distributions?

By applying Chain rule

Let $X_1, X_2, ..., X_n$ be any arbitrary ordering of nodes in a DAG. $P(x_1, x_2, ..., x_n) = \prod_j P(x_j|x_1, ..., x_{j-1})$

Is it possible that conditional probability of some variable $X_j$ is not sensitive to all its predecessors?
Decomposing joint distribution-$P(V)$

How would you decompose joint distribution $P(V)$ into smaller distributions?

By applying Chain rule

Let $X_1, X_2, ..., X_n$ be any arbitrary ordering of nodes in a DAG.

\[
P(x_1, x_2, ..., x_n) = \prod_j P(x_j | x_1, ..., x_{j-1})
\]

Is it possible that conditional probability of some variable $X_j$ is not sensitive to all its predecessors?

Yes!
Markovian Parents

Markovian Parents

$X_1 : \phi$
$X_2 : \{X_1\}$
$X_3 : \{X_1\}$
$X_4 : \{X_2, X_3\}$
$X_5 : \{X_4\}$

Figure: Bayesian Network representing dependencies
Markovian Parents

\[
P(x_1, x_2, x_3, x_4, x_5) = P(x_1)P(x_2|x_1)P(x_3|x_1)P(x_4|x_2, x_3)P(x_5|x_4)
\]
Markov Compatibility

Let $V = \{x_1, x_2, ..., x_n\}$ be the set of observed nodes and $pa_i$ be the Markovian parents of $x_i$. Then,

$$P(v) = P(x_1, x_2, .., x_n) = \prod_i P(x_i|pa_i).$$
**Markov Compatibility**

Let \( V = \{x_1, x_2, \ldots, x_n\} \) be the set of observed nodes and \( pa_i \) be the Markovian parents of \( x_i \). Then, 
\[
P(v) = P(x_1, x_2, \ldots, x_n) = \prod_i P(x_i | pa_i).
\]

**Definition (Markov Compatibility)**

If a probability distribution \( P \) admits Markovian factorization of observed nodes relative to DAG \( G \), we say that \( G \) and \( P \) are Markov compatible.
Markov Compatibility

Let $V = \{x_1, x_2, \ldots, x_n\}$ be the set of observed nodes and $pa_i$ be the Markovian parents of $x_i$. Then,

$$P(v) = P(x_1, x_2, \ldots, x_n) = \prod_i P(x_i | pa_i).$$

**Definition (Markov Compatibility)**

If a probability distribution $P$ admits Markovian factorization of observed nodes relative to DAG $G$, we say that $G$ and $P$ are Markov compatible.

**Example**

<table>
<thead>
<tr>
<th>$X$</th>
<th>$Y$</th>
<th>$P(X, Y)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>0.225</td>
</tr>
<tr>
<td>1</td>
<td>0</td>
<td>0.375</td>
</tr>
<tr>
<td>0</td>
<td>1</td>
<td>0.125</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>0.275</td>
</tr>
</tbody>
</table>

Markov Compatible DAGs:

- $X \rightarrow Y$
- $X \leftarrow Y$
Testing Markov Compatibility

Given a DAG $G$ and distribution $P$, how can you conclude that $P$ and $G$ are compatible?
Testing Markov Compatibility

Given a DAG $G$ and distribution $P$, how can you conclude that $P$ and $G$ are compatible?

- Parents shielding tests
  - non-descendants
  - predecessors
- d-separation
d-Separation

Definition
Let X, Y and Z be disjoint sets in DAG G. X and Y are d-separated by Z (written \((X \perp \!\!\!\perp Y | Z)_G\)) if and only if Z blocks every path from a node in X to a node in Y.
d-Separation

**Definition**
Let $X$, $Y$ and $Z$ be disjoint sets in DAG $G$. $X$ and $Y$ are d-separated by $Z$ (written $(X \perp Y|Z)_G$) if and only if $Z$ blocks every path from a node in $X$ to a node in $Y$.

A path $p$ is said to be d-separated (or blocked) by a set of nodes $Z$ if and only if:

1. $p$ contains a chain $i \rightarrow m \rightarrow j$ or a fork $i \leftarrow m \rightarrow j$ such that the middle node $m$ is in $Z$, or

2. ...
Example: d-Separation

\[ X_1 \perp X_4 | X_2, X_3 \]
\[ X_3 \perp X_5 | X_4 \]
\[ X_1 \perp X_5 | X_4 \]
d-Separation

Definition
Let $X$, $Y$ and $Z$ be disjoint sets in DAG $G$. $X$ and $Y$ are d-separated by $Z$ (written $(X \perp Y | Z)_G$) if and only if $Z$ blocks every path from a node in $X$ to a node in $Y$.

A path $p$ is said to be d-separated (or blocked) by a set of nodes $Z$ if and only if:

(1) $p$ contains a chain $i \rightarrow m \rightarrow j$ or a fork $i \leftarrow m \rightarrow j$ such that the middle node $m$ is in $Z$, or

(2) $p$ contains an inverted fork (or collider) $i \rightarrow m \leftarrow$ such that the middle node $m$ is not in $Z$ and such that no descendant of $m$ is in $Z$. 
Example: d-Separation

\[ X_3 \perp\!\!\!\!\!\!\perp X_2 \mid X_1 \]

Graphical Models for Causal Inference

Karthika Mohan and Judea Pearl
Example: d-Separation

(a) X \downarrow Y \mid φ

(b) X \downarrow Y

Karthika Mohan and Judea Pearl
Graphical Models for Causal Inference
Example: d-Separation

Case (a): $X \perp Y \mid \phi$

Case (b): $X \perp Y$
Example: d-Separation

Case (a): $X \perp \!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
d-Separation

When is it impossible to d-separate 2 non-adjacent nodes $X$ and $Y$?
Do we need to test all sets for possible separation?
**Inducing path**

**Definition**
Path between 2 nodes $X$ and $Y$ is termed inducing if every non-terminal node on the path:
(i) is a collider and
(ii) an ancestor of either $X$ or $Y$ (or both)

![Diagram](image)

Note: There are no separators for $X$ and $Y$. 
The Five Necessary Steps of Causal Analysis

Define Express the target quantity $Q$ as property of the model $M$.

Assume Express causal assumptions in structural or graphical form.

Identify Determine if $Q$ is identifiable.

Estimate Estimate $Q$ if it is identifiable; approximate it, if it is not.

Test If $M$ has testable implications
A “Mini” Turing Test in Causal Conversation

Figure: Turing Test
A “Mini” Turing Test in Causal Conversation

Input: Story

Question: What if? What is? Why?

Answers: I believe that...

Figure: Turing Test
A “Mini” Turing Test in Causal Conversation

The Story

**Figure:** Turing Test
A “Mini” Turing Test in Causal Conversation

Figure: Turing Test

Q1: If the season is dry and the pavement is slippery, did it rain?
A “Mini” Turing Test in Causal Conversation

Figure: Turing Test

Q1: If the season is dry and the pavement is slippery, did it rain?
A1: Unlikely, it is more likely that the sprinkler was ON with a very slight possibility that it is not even wet.

Karthika Mohan and Judea Pearl
Graphical Models for Causal Inference
A “Mini” Turing Test in Causal Conversation

The Story

Figure: Turing Test

Q2: But what if we see that the sprinkler is OFF?
A “Mini” Turing Test in Causal Conversation

Figure: Turing Test

Q2: But what if we see that the sprinkler is OFF?
A2: Then it is more likely that it rained.
Without graphs, # of Table Entries = 32
A “Mini” Turing Test in Causal Conversation

Figure: Turing Test

Q3: Do you mean that if we actually turn the sprinkler ON, the rain will be less likely?
Figure: Turing Test

Q3: Do you mean that if we actually turn the sprinkler ON, the rain will be less likely?
A3: No, the likelihood of rain would remain the same but the pavement would surely get wet.
Without graphs, # of Table Entries = 32 * 32
A “Mini” Turing Test in Causal Conversation

Figure: Turing Test

Q4: Suppose we see that the sprinkler is ON and pavement is wet. What if the sprinkler were OFF?
A “Mini” Turing Test in Causal Conversation

Figure: Turing Test

Q4: Suppose we see that the sprinkler is ON and pavement is wet. What if the sprinkler were OFF?
A4: The pavement would be dry because the season is likely to be dry.
Without graphs, what would be the # of table entries?
Interventions

**Query:** Would the pavement be slippery if we *make sure that* the sprinkler is on?

**Compute:** \( P(x_5 \mid do(x_3)) \)

May be equivalently represented as:

(a) \( P(x_5 \mid \hat{x}_3) \)

(b) \( P_{x_3}(x_5) \)

---

**Figure:** DAG before intervention
Interventions

Compute: $P(x_5 | do(x_3))$

Figure: DAG before intervention

$$P(v) = P(x_1)P(x_2 | x_1)P(x_3 | x_1)P(x_4 | x_2, x_3)P(x_5 | x_4)$$
Interventions

**Compute:** \( P(x_5|do(x_3)) \)

\[
P(v) = P(x_1)P(x_2|x_1)P(x_3|x_1)P(x_4|x_2, x_3)P(x_5|x_4)\]
\[
P(x_1, x_2, x_4, x_5|do(x_3)) = P(x_1)P(x_2|x_1)P(x_4|x_2, x_3)P(x_5|x_4)\]

Figure: DAG before intervention

Figure: DAG after intervention

Karthika Mohan and Judea Pearl
Graphical Models for Causal Inference
Interventions

Compute: \( P(x_5|do(x_3)) \)

\[
P(v) = P(x_1)P(x_2|x_1) \cdot P(x_3|x_1) \cdot P(x_4|x_2, x_3) \cdot P(x_5|x_4)
\]

\[
P(x_5|do(x_3)) = \sum_{x_1, x_2, x_4} P(x_1)P(x_2|x_1)P(x_4|x_2, x_3)P(x_5|x_4)
\]

Note: \( P(x_5|do(x_3)) \neq P(x_5|x_3) \) i.e. Doing \neq\ Seeing
Question: Can you estimate $P(y|do(x))$, given $P(x,y)$?
Examples

Question: Can you estimate $P(y|\text{do}(x))$, given $P(x,y)$?

NO!
Question: Can you estimate \( P(y|do(x)) \), given \( P(x, y) \)?

**NO!**

\[
P(x, y) = \sum_u P(x, y, u) = \sum_u P(y|x, u)P(x|u)P(u)
\]

\[
P(y|do(x)) = \sum_u P(y|x, u)P(u)
\]
Identifiability

Definition

Let $Q(M)$ be any computable quantity of a model $M$. 
Identifiability

Definition
Let $Q(M)$ be any computable quantity of a model $M$. We say that $Q$ is identifiable in a class $M$ of models if, for any pairs of models $M_1$ and $M_2$ from $M$, $Q(M_1) = Q(M_2)$ whenever $P_{M_1}(v) = P_{M_2}(v)$. 
Estimating causal effect

Adjustment for direct causes

Compute: \( P(y|\hat{x}) \)

\[
P(x, y, z, w) = P(y|x, w)P(x|z)P(w|z)P(z)
\]

\[
P(y, z, w|do(x)) = P(y|x, w)P(w|z)P(z) \frac{P(x|z)}{P(x|z)}
\]

\[
P(y, z, w|do(x)) = \frac{P(x,y,z,w)}{P(x|z)}
\]

\[
P(y|do(x)) = \sum_{z,w} P(yw|x, z)P(z) = \sum_z P(y|x, z)P(z)
\]

Figure: DAGs before and after intervention
Theorem (Adjustment for direct causes)

Let $PA_i$ denote the set of direct causes of $X_i$ and let $Y$ be any set of variables disjoint of $\{X_i \cup Pa_i\}$. The causal effect of $X_i$ on $Y$ is given by:

$$P(y|\hat{x}_i) = \sum_{pa_i} P(y|x_i, pa_i)P(pa_i)$$

where $P(y|x_i, pa_i)$ and $P(pa_i)$ represent pre-interventional probabilities.
Example: Adjustment for direct causes

**Query:** Would the pavement be slippery if we *make sure that* the sprinkler is on?

$$P(x_5|x_3) = \sum_{x_1} P(x_5|x_3, x_1)P(x_1)$$

---

**Figure:** DAG before intervention

**Figure:** DAG after intervention
Estimating Causal Effect

Compute: \( P(X_j|do(X_i)) \)

How can we find a set \( Z \) of concomitants that are sufficient for identifying causal effect?

![Graphical Model](image-url)
Back-door Criterion for Identifiability

Definition (Pearl-1993)

A set of variables $Z$ satisfies the back-door criterion relative to an ordered pair of variables $(X_i, X_j)$ in a DAG $G$ if:

(i) no node in $Z$ is a descendant of $X_i$; and

(ii) $Z$ blocks every path between $X_i$ and $X_j$ that contains an arrow into $X_i$.

$$P(x_j|do(x_i)) = \sum_z P(x_j|x_i, z)P(z)$$

Diagram:

![Diagram showing a causal graph with nodes $X_1, X_2, X_3, X_4, X_5, X_6,$ and $X_i, X_j$. The diagram illustrates the back-door criterion with shaded nodes representing variables in $Z$ that block the paths between $X_i$ and $X_j$.]
Estimating causal effect: \( P(y|do(x)) \)

- Can you adjust for direct cause?

![Graphical Model](attachment:graph.png)
Estimating causal effect: $P(y|do(x))$

- Can you adjust for direct cause? NO!

Diagram:

```
(U) (Unobserved)

X -> Z -> Y
```
Estimating causal effect: \( P(y|do(x)) \)

- Can you apply backdoor criterion?

Graphical representation:

- \( U \) (Unobserved)
- \( X \)
- \( Z \)
- \( Y \)
Estimating causal effect: $P(y|do(x))$

- Can you apply backdoor criterion? NO!

```
X \rightarrow Z \rightarrow Y
U (Unobserved)
```

Karthika Mohan and Judea Pearl

Graphical Models for Causal Inference
Estimating causal effect: $P(y|do(x))$ 

- Is $P(y|do(x))$ identifiable?

![Diagram showing causal relationships between variables X, Z, Y, and an unobserved variable U.](image)
Estimating causal effect: $P(y|do(x))$

- Is $P(y|do(x))$ identifiable? YES!
Estimating causal effect: $P(y|do(x))$

Given: $P(y|\hat{x}) = \sum_z P(y|\hat{z})P(z|\hat{x})$
Estimating causal effect: \( P(y|do(x)) \)

Given: 
\[
P(y|\hat{x}) = \sum_z P(y|\hat{z})P(z|\hat{x})
\]
\[
P(z|\hat{x}) = P(z|x)
\]
\[
P(y|\hat{z}) = \sum_{x'} P(y|x',z)P(x')
\]

Therefore,
\[
P(y|\hat{x}) = \sum_z P(z|x) \sum_{x'} P(y|x',z)P(x')
\]
Front-door Criterion for Identifiability

Definition (Pearl-1995)

A set of variables $Z$ satisfies the front-door criterion relative to an ordered pair of variables $(X_i, X_j)$ in a DAG $G$ if:

(i) $Z$ intercepts all directed paths from $X$ to $Y$; and
(ii) there is no unblocked back-door path from $X$ to $Z$; and
(iii) all back-door paths from $Z$ to $Y$ are blocked by $X$

Figure: Frontdoor criterion is satisfied by $Z = \{Z_1, Z_2, Z_3\}$
If $Z$ satisfies the front door criterion relative to $(X, Y)$ and if $P(x, z) > 0$, then the causal effect of $X$ on $Y$ is identifiable and is given by:

$$P(y|\hat{x}) = \sum_z P(z|x) \sum_{x'} P(y|x', z) P(x')$$
Estimating causal effect $P(y|do(x))$

How can you syntactically derive claims about interventions?
Estimating causal effect $P(y|do(x))$

How can you syntactically derive claims about interventions?

▶ do-calculus

Figure: Subgraphs of $G$ used in the derivation of causal effects.
do-Calculus-[Pearl-1995]

**Rule-1** Insertion or deletion of observations

\[ P(y|\hat{x}, z, w) = P(y|\hat{x}, w) \text{ if } (Y \perp\!\!\!\!\!\!\!\perp Z|X, W)_{G_X} \]

\[ G \]

\[ G_Z = G_X \]

\[ G_{\overline{XZ}} \]

\[ G_{\overline{Z}} \]

\[ G_{\overline{XZ}} \]
Rule-2 Action/Observation exchange

\[ P(y|\hat{x}, \hat{z}, w) = P(y|\hat{x}, z, w) \text{ if } (Y \perp Z|X, W)_{G_{\overline{XZ}}} \]
**do-Calculus-[Pearl-1995]**

**Rule-3** Insertion or deletion of actions

\[ P(y|\hat{x}, \hat{z}, w) = P(y|\hat{x}, w) \] if \((Y \perp\!\!\!\perp Z|X, W)_{G_{\overline{X},Z(W)}}\)

where \(Z(W)\) is the set of \(Z\) nodes that are not ancestors of any \(W\) node in \(G_{\overline{X}}\)

---

**Diagram:**

- **G**
  - \(X\) → \(Z\) → \(Y\)
  - \(U\) (Unobserved)

- **G_{\overline{Z}} = G_{\overline{X}}**
  - \(X\) → \(Z\) → \(Y\)
  - \(U\) (Unobserved)

- **G_{\overline{XZ}}**
  - \(X\) → \(Z\) → \(Y\)

- **G_{\overline{Z}}**
  - \(X\) → \(Z\) → \(Y\)

- **G_{\overline{XZ}}**
  - \(X\) → \(Z\) → \(Y\)
Deriving causal effect using do-calculus

Compute: $P(y|\hat{z})$

$P(y|\hat{z}) = \sum_x P(y|x,\hat{z})P(x|\hat{z})$

$P(x|\hat{z}) = P(x)$ since $(Z \Perp X)_{G_{\overline{Z}}}$

$P(y|x,\hat{z}) = P(y|x,z)$ since $(Z \Perp Y|X)_{G_Z}$

$P(y|\hat{z}) = \sum_x P(y|x,z)P(x)$
Prove: \( P(y|\hat{x}) = \sum_z P(y|\hat{z})P(z|\hat{x}) \)
Prove: \( P(y|\hat{x}) = \sum_z P(y|\hat{z})P(z|\hat{x}) \)

\[ P(y|\hat{x}) = \sum_z P(yz|\hat{x}) = \sum_z P(y|\hat{z}z)P(z|\hat{x}) \]
Prove: \( P(y|\hat{x}) = \sum_z P(y|\hat{z})P(z|\hat{x}) \)

\[
P(y|\hat{x}) = \sum_z P(yz|\hat{x}) = \sum_z P(y|\hat{x}z)P(z|\hat{x})
\]

\( P(y|z\hat{x}) = P(y|\hat{z}\hat{x}) \) since \( Y \perp Z \) in \( G_{\overline{XZ}} \)
Prove: \( P(y|\hat{x}) = \sum_z P(y|\hat{z})P(z|\hat{x}) \)

\[
P(y|\hat{x}) = \sum_z P(yz|\hat{x}) = \sum_z P(y|\hat{z}z)P(z|\hat{x})
\]

\[
P(y|z\hat{x}) = P(y|\hat{z}\hat{x}) \text{ since } Y \perp Z \text{ in } G_{XXZ}
\]

\[
= P(y|\hat{z}) \text{ since } Y \perp X \text{ in } G_{ZXZ}
\]
Graphical Models in which $P(y|\hat{x})$ is Identifiable

(a) X \rightarrow Y

(b) X \rightarrow Y \rightarrow Z

(c) X \rightarrow Y

(d) X \rightarrow Y

(e) X \rightarrow Z \rightarrow Y

(f) X \rightarrow Z_1 \rightarrow Y

(g) X \rightarrow Z_3 \rightarrow Y
Graphical Models in which $P(y|\hat{x})$ is not Identifiable

(a) $X \rightarrow Y$
(b) $X \rightarrow Z \rightarrow Y$
(c) $X \rightarrow Y \rightarrow Z$
(d) $X \rightarrow Z \leftarrow Y$
(e) $X \rightarrow Y \rightarrow Z$
(f) $X \rightarrow Z \rightarrow Y$
(g) $X \rightarrow Z_1 \rightarrow Z_2 \rightarrow Y$
(h) $X \rightarrow Z \rightarrow W \rightarrow Y$

Karthika Mohan and Judea Pearl
Graphical Models for Causal Inference
C-components and C-factor

Two variables are said to be in the same C-component if they are connected by a path comprising of only bi-directional edges [Tian & Pearl, 2002].
C-components and C-factor

Two variables are said to be in the same C-component if they are connected by a path comprising of only bi-directional edges [Tian & Pearl, 2002].

\[ S_1 = \{ X, Y, W_1, W_2 \} \]
\[ S_2 = \{ Z_1 \} \]
\[ S_3 = \{ Z_2 \} \]
\[ S_4 = \{ T \} \]
C-components and C-factor

Two variables are said to be in the same C-component if they are connected by a path comprising of only bi-directional edges [Tian & Pearl, 2002].

\[ S_1 = \{X, Y, W_1, W_2\} \]
\[ S_2 = \{Z_1\} \]
\[ S_3 = \{Z_2\} \]
\[ S_4 = \{T\} \]

**C-factor**: \[ Q[S_i](v) = P_{v \mid s_i}(s_i) \]
Lemma (Tian & Pearl, 2002)

Let a topological order over \( V \) be \( V_1 < V_2 < ... < V_n \) and let 
\( V^{(i)} = \{V_1, V_2, ..., V_i\}, \ i = 1, ..., n \) and \( V^{(0)} = \phi \). For any set \( C \), 
let \( G_C \) denote the subgraph of \( G \) composed only of variables in \( C \). Then:

(i) Each C-factor \( Q_j, j = 1, ..., k \) is identifiable and is given by:

\[
Q_j = \prod_{\{i : V_i \in S_j\}} P(v_i | v^{(i-1)})
\]
Example: Identifiability of C-factor

Admissible order: $X_1 < X_2 < X_3 < X_4 < Y$

$Q_1 = P(x_4|x_1, x_2, x_3)P(x_2|x_1)$

$Q_2 = P(y|x_1, x_2, x_3, x_4)P(x_3|x_1, x_2)P(x_1)$
Necessary and Sufficient condition for identifiability of $P_x(v)$

Theorem (Tian & Pearl, 2002)

Let $X$ be a singleton. $P_x(v)$ is identifiable if and only if there is no bi-directed path connecting $X$ to any of its children.
Necessary and Sufficient condition for identifiability of $P_x(v)$

Theorem (Tian & Pearl, 2002)

Let $X$ be a singleton. $P_x(v)$ is identifiable if and only if there is no bi-directed path connecting $X$ to any of its children. When $P_x(v)$ is identifiable, it is given by:

$$P_x(v) = \frac{P(v)}{Q^X} \sum_x Q^X,$$

where $Q^X$ is the c-factor corresponding to the c-component $S^X$ that contains $X$. 
Example: Necessary and Sufficient condition for identifiability of $P_x(v)$

Admissible order: $X_1 < X_2 < X_3 < X_4 < Y$

$Q_1 = P(x_4| x_1, x_2, x_3)P(x_2| x_1)$

$Q_2 = P(y| x_1, x_2, x_3, x_4)P(x_3| x_1, x_2)P(x_1)$

$P_{x_1}(x_2, x_3, x_4, y) = Q_1 \sum_{x_1} Q_2$

$= P(x_4| x_1, x_2, x_3)P(x_2| x_1)$

$\sum_{x_1} P(y| x_1, x_2, x_3, x_4)P(x_3| x_1, x_2)P(x_1)$
Causal Effect Identifiability

Identification of $P_x(y|z)$ where $X \cap Y \cap Z = \phi$ and $X$ is not necessarily a singleton, [Shpitser & Pearl, 2006]

- Hedge Criterion
- IDC - Sound and Complete Algorithm
Counterfactuals

Query: Would the prisoner be dead had rifleman A not shot him, given that the prisoner is dead and rifleman A shot him?

\[
\begin{align*}
U \ (\text{Court order})
\end{align*}
\]

\[
\begin{align*}
C \ (\text{Captain})
\end{align*}
\]

\[
\begin{align*}
A
\end{align*}
\]

\[
\begin{align*}
B \ (\text{Riflemen})
\end{align*}
\]

\[
\begin{align*}
D \ (\text{Death})
\end{align*}
\]
Counterfactuals

Query: Would the prisoner be dead had rifleman A not shot him, given that the prisoner is dead and rifleman A shot him?

\[
\begin{align*}
U \quad \text{(Court order)} \\
C \quad \text{(Captain)} \\
A \\
B \quad \text{(Riflemen)} \\
D \quad \text{(Death)}
\end{align*}
\]
Counterfactuals

Query: Would the prisoner be dead had rifleman A not shot him, given that the prisoner is dead and rifleman A shot him?

\[ A \rightarrow C \rightarrow U \rightarrow B \leftarrow D \]

- Abduction
- Intervention

Karthika Mohan and Judea Pearl
Graphical Models for Causal Inference
Counterfactuals

Query: Would the prisoner be dead had rifleman A not shot him, given that the prisoner is dead and rifleman A shot him?

\[ U \text{ (Court order)} \]
\[ C \text{ (Captain)} \]
\[ A \]
\[ B \text{ (Riflemen)} \]
\[ D \text{ (Death)} \]

- Abduction
- Intervention
- Prediction
Counterfactuals

Query: Would the prisoner be dead had rifleman A not shot him, given that the prisoner is dead and rifleman A shot him?

\[
\begin{align*}
M & = C = U \\
M & = A = C \\
M & = B = C \\
M & = D = A \lor B
\end{align*}
\]

Facts: \(D\)

Conclusions: \(U, A, B, C, D\)

\[
\begin{align*}
M_{\neg A} & = C = U \\
M_{\neg A} & = \neg A \\
M_{\neg A} & = B = C \\
M_{\neg A} & = D = A \lor B
\end{align*}
\]

Facts: \(U\)

Conclusions: \(U, \neg A, B, C, D\)

Karthika Mohan and Judea Pearl

Graphical Models for Causal Inference
Markov Equivalence

Given 2 models, is there a test that would tell them apart?

Definition
Two graphs $G_1$ and $G_2$ are said to be Markov equivalent if every d-separation condition in one also holds in the other. 

\[\begin{align*}
A &\rightarrow Z \\
&\downarrow \quad \downarrow \\
B &\quad B \\
\end{align*}\]

\[\begin{align*}
A &\leftarrow Z \\
&\downarrow \quad \downarrow \\
B &\quad B \\
\end{align*}\]

$(G_1)$ $(G_2)$
Markov Equivalence

Given 2 models, is there a test that would tell them apart?

Definition
Two graphs $G_1$ and $G_2$ are said to be Markov equivalent if every d-separation condition in one also holds in the other.

Are these DAGs Markov Equivalent?

Hard to enumerate all separation conditions.
Observational Equivalence

Theorem (Verma & Pearl 1990)

Two DAGs are observationally equivalent iff they have the same sets of edges and the same sets of v-structures, that is, two converging arrows whose tails are not connected by an arrow.

Figure: Observationally Equivalent DAGs
Markov Equivalence and Observational Equivalence

If two DAGs are Markov Equivalent, then they are Observationally Equivalent as well. True/False?
Markov Equivalence and Observational Equivalence

If two DAGs are Markov Equivalent, then they are Observationally Equivalent as well. True/False? True if all variables are observed (i.e. no bi-directed edges) and False otherwise.
Markov Equivalence and Observational Equivalence

If two DAGs are Markov Equivalent, then they are Observationally Equivalent as well. True/False? True if all variables are observed (i.e. no bi-directed edges) and False otherwise.

Figure: DAGs that are Markov Equivalent but not Observationally Equivalent

How would you distinguish between the two?

- Verma Constraints (Refer slide:115)
Ancestral Graphs

Definition (Ancestral Graphs)

A graph which may contain directed or bi-directed edges is ancestral if:

(i) there are no directed cycles
(ii) whenever there is an edge $X \leftrightarrow Y$, then there is no directed path from $X$ to $Y$ or from $Y$ to $X$.

Figure: Ancestral graph

Figure: Not an Ancestral graph
Maximal Ancestral Graphs (MAGs)

Definition (Spirtes & Richardson, 2002)
An ancestral graph is said to be maximal if, for every pair of non-adjacent nodes $X$, $Y$ there exists a set $Z$ such that $X$ and $Y$ are d-separated conditional on $Z$.

Figure: DAG and its corresponding MAG
Construction of a MAG

Given: DAG $G$

Step-1: Construct a graph $M$ comprising of:
(i) all nodes in $G$
(ii) all uni-directional edges in $G$

Figure: DAG $G$

Figure: Graph $M$
**Construction of a MAG**

**Step-2:** For every bi-directed edge $A \leftrightarrow B$ in $G$,

(i) add $A \rightarrow B$ to $M$ if $A$ is an ancestor of $B$ in $G$

(ii) add $A \leftarrow B$ to $M$ if $B$ is an ancestor of $A$ in $G$

(iii) copy $A \leftrightarrow B$ to $M$ if (i) and (ii) do not hold true

![Figure: DAG G](image1)

![Figure: Graph M](image2)
Construction of a MAG

**Step-3:** For every pair of non-adjacent nodes $A$ and $B$ in $G$, connected by an *inducing path*,
(i) add $A \rightarrow B$ to $M$ if $A$ is an ancestor of $B$ in $G$
(ii) add $A \leftarrow B$ to $M$ if $B$ is an ancestor of $A$ in $G$
(iii) add $A \leftrightarrow B$ to $M$ if (i) and (ii) do not hold true

![DAG G and MAG M](image-url)
Markov Equivalence

Theorem

Two graphs $G_1$ and $G_2$ are said to be Markov equivalent if their MAGs are Markov Equivalent.

Are these MAGs Markov Equivalent?

Note: Markov Equivalence in MAGs are easier to check.

Complete criterion for determining Markov Equivalence of 2 MAGs: [Ali, Richardson and Spirtes, 2009]
Reversing an edge in a MAG

Definition (Screened Edge)

[Tian,2005] An edge $X \rightarrow Y$ is a screened edge in a MAG if $Pa(Y) = Pa(X) \cup \{X\}$ and $Sp(Y) = Sp(X)^1$.

Figure: MAG with Screened Edge: $X \rightarrow Y$

Footnote:

$^1$Nodes $X$ and $Y$ are spouses, if they are connected by a bi-directed edge.
Reversing an edge in a MAG

Theorem (Tian, 2005)

Let $M$ be a MAG with edge $X \rightarrow Y$ and $M'$ be a graph with edge $X \leftarrow Y$, otherwise identical to $M$. Then $M'$ is a MAG that is Markov Equivalent to $M$ if and only if $X \rightarrow Y$ is a screened edge in $M$.

Figure: Markov Equivalent MAGs
Confounding Equivalence

Definition (Pearl and Paz, 2009)

Define two sets, T and Z as c-equivalent (relative to X and Y), written $T \approx Z$, if the following equality holds for every x and y:

$$
\sum_t P(y|x, t)P(t) = \sum_z P(y|x, z)P(z) \quad \forall x, y
$$

Examples:

- $T = \{W_1, V_2\} \approx Z = \{W_2, V_1\}$
- $T = \{W_1, V_1\} \approx Z = \{W_2, V_2\}$
- $T = \{W_1, W_2\} \approx Z = \{W_1\}$
- $T = \{W_1, W_2\} \not\approx Z = \{W_2\}$

Note: C-equivalence is testable
Necessary and Sufficient Condition for C-Equivalence

Theorem (Pearl and Paz, 2009)

Let $Z$ and $T$ be two sets of variables containing no descendant of $X$. A necessary and sufficient condition for $Z$ and $T$ to be $c$-equivalent is that at least one of the following conditions hold:

- $X \perp (Z \cup T) \mid (Z \cap T)$ or
- $Z$ and $T$ are $G$-admissible

Examples:

- $T = \{W_1, V_2\} \approx Z = \{W_2, V_1\}$
- $T = \{W_1, V_1\} \approx Z = \{W_2, V_2\}$
- $T = \{W_1, W_2\} \approx Z = \{W_1\}$
- $T = \{W_1, W_2\} \not\approx Z = \{W_2\}$

\(^2\)satisfies back-door criterion
Assume all variables are normalized to have zero mean and unit variance.

\[
\begin{align*}
Z &= e_1 \\
W &= e_2 \\
X &= aZ + e_3 \\
Y &= bW + cX + e_4 \\
\text{Cov}(e_1, e_2) &= \alpha \neq 0 \\
\text{Cov}(e_2, e_3) &= \beta \neq 0 \\
\text{Cov}(e_3, e_4) &= \gamma \neq 0
\end{align*}
\]

Which parameters can be identified?
Definition
For any linear model for a causal diagram D that may include cycles and bi-directed arcs, the partial correlation \( \rho_{XY.Z} \) must vanish if and only if node X is d-separated from node Y by the variables of Z in D [Spirtes et al., 1997b].

\[ r_{TX.W_1Z_1} = 0 \]

Find more
Single Door Criterion for Direct Effects

**Theorem**

Let $G$ be any path diagram in which $\alpha$ is the path coefficient associated with link $X \rightarrow Y$ and let $G_\alpha$ denote the diagram that results when $X \rightarrow Y$ is deleted from $G$. The coefficient $\alpha$ is identifiable if there exists a set of variables $Z$ such that:

(i) $Z$ contains no descendant of $Y$ and
(ii) $Z$ d-separates $X$ from $Y$ in $G_\alpha$

Moreover, if $Z$ satisfies these two conditions, then $\alpha$ is equal to the regression coefficient $r_{YX|Z}$.
Instrumental Variables (IV)

Definition

A variable $Z$ is an instrument relative to a cause $X$ and an effect $Y$ if:

- $Z$ is independent of all error terms that have an influence on $Y$ when $X$ is held constant, and
- $Z$ is **not** independent of $X$.

In linear systems, Causal effect of $X$ on $Y = \frac{r_{ZY}}{r_{ZX}}$

Figure: $Z$ is an instrument in (a), (b) and (c) but not in (d)
Conditional Instrumental Variable

Definition (Brito & Pearl, 2002)

$Z$ is an instrumental variable if $\exists$ a set $W$ such that:

- $W$ contains only non-descendants of $Y$
- $W$ d-separates $Z$ from $Y$ in the sub-graph $G_\alpha$ obtained by removing the edge $X \rightarrow Y$
- $W$ does not d-separate $Z$ from $X$ in $G_\alpha$

Figure: Graph $G$ and corresponding subgraph $G_\alpha$
Conditional Instrumental Variable

- $Z$ is a conditional instrumental variable. Hence, $\alpha = \text{Causal effect of } X \text{ on } Y = \frac{r_{ZY,W}}{r_{ZX,W}}$.
- $W$ does not satisfy single-door criterion. So, $\alpha$ cannot be identified using single-door.

Figure: Graph $G$ and corresponding subgraph $G_\alpha$
Verma Constraints ([Tian and Pearl, 2002])

\[
Q[\{B, D\}] = \sum_u P(b|a, u)P(d|c, u)P(u)
\]

\[
P_{v\backslash d}(d) = \sum_u P(d|c, u)P(u)
\]

Also,

\[
Q[\{B, D\}] = P(d|a, b, c)P(b|a)
\]

\[
P_{v\backslash d}(d) = \sum_b P(d|a, b, c)P(b|a)
\]

\[
\sum_b P(d|a, b, c)P(b|a) \text{ is independent of } a.
\]

\[
Q[\{B, D\}] = \sum_u P(b|a, u)P(d|a, c, u)P(u)
\]

\[
P_{v\backslash d}(d) = \sum_u P(d|a, u, c)P(u)
\]

Also,

\[
Q[\{B, D\}] = P(d|a, b, c)P(b|a)
\]

\[
P_{v\backslash d}(d) = \sum_b P(d|a, b, c)P(b|a)
\]

\[
\sum_b P(d|a, b, c)P(b|a) \text{ is not independent of } a.
\]
Conclusions

Graphs are indispensable for:

- encoding causal assumptions
- identifying parameters and causal effects
- identifying testable implications

Go ahead and Exploit the Power of Graphs!
Thank You!