Center for Causal Discovery

Day 2: Search

June 14, 2016

Carnegie Mellon University
Outline

Models → Data
1) Representing/Modeling Causal Systems
2) Estimation and Model fit
3) Hands on with Real Data

Models ← Data
1) Bridge Principles: Markov Axiom and D-separation
2) Model Equivalence
3) Model Search
Causal Structure

Testable Statistical Predictions

Causal Graphs

\[
\begin{align*}
&\forall x, y, z \quad P(X = x, Z = z \mid Y = y) = \\
&\quad P(X = x \mid Y = y) \cdot P(Z = z \mid Y = y)
\end{align*}
\]
**Bridge Principles:**

*Acyclic Causal Graph over* \( V \) \( \Rightarrow \) *Constraints on* \( P(V) \)

**Weak Causal Markov Assumption**

\( V_1, V_2 \) causally disconnected \( \Rightarrow \) \( V_1 \perp\!\!\!\perp V_2 \)

\( V_1, V_2 \) causally disconnected \( \iff \)

i. \( V_1 \) not a cause of \( V_2 \), and

ii. \( V_1 \) not an effect of \( V_2 \), and

iii. No common cause \( Z \) of \( V_1 \) and \( V_2 \)
Bridge Principles: 
Acyclic Causal Graph over V $\Rightarrow$ Constraints on P(V)

Weak Causal Markov Assumption
$V_1, V_2$ causally disconnected $\Rightarrow$ $V_1 \perp\!\!\!\!\!\!\perp V_2$

Determinism
(Structural Equations)

Causal Markov Axiom
If $G$ is a causal graph, and $P$ a probability distribution over the variables in $G$, then in $<G,P>$ satisfy the Markov Axiom iff:

*every variable $V$ is independent of its non-effects,*
*conditional on its immediate causes.*
Bridge Principles: **Acyclic Causal Graph** over $V \Rightarrow$ Constraints on $P(V)$

- **Causal Markov Axiom**
- **Acyclicity**
- **d-separation criterion**

**Causal Graph**

- $Z \longrightarrow X \longrightarrow Y_1$
- $Y_2 \longrightarrow X$

**Graphical Independence Oracle**

- $Z \perp\!\!\!\!\!\!\perp Y_1 \mid X$
- $Z \perp\!\!\!\!\!\!\perp Y_2 \mid X$
- $Z \perp\!\!\!\!\!\!\perp Y_2 \mid X,Y_1$
- $Y_1 \perp\!\!\!\!\!\!\perp Y_2 \mid X$
- $Y_1 \perp\!\!\!\!\!\!\perp Y_2 \mid X,Z$
Equivalence Classes

Equivalence:

• Independence Equivalence: \( M_1 \models (X \perp \!\!\!\!\perp Y \mid Z) \iff M_2 \models (X \perp \!\!\!\!\perp Y \mid Z) \)

• Distribution Equivalence: \( \forall \theta_1 \exists \theta_2 M_1(\theta_1) = M_2(\theta_2) \), and vice versa

• Independence (d-separation equivalence)
  - DAGs: Patterns
  - PAGs: Partial Ancestral Graphs
  - Intervention Equivalence Classes

• Measurement Model Equivalence Classes
• Linear Non-Gaussian Model Equivalence Classes
• Etc.
d-separation/Independence Equivalence

D-separation Equivalence Theorem (Verma and Pearl, 1988)

Two acyclic graphs over the same set of variables are d-separation equivalent iff they have:

• the same adjacencies
• the same unshielded colliders
Colliders

Y: Collider

Y: Non-Collider

Shielded

Unshielded
d-separation/Independence Equivalence

D-separation Equivalence Theorem (Verma and Pearl, 1988)

Two acyclic graphs over the same set of variables are d-separation equivalent iff they have:

• the same adjacencies
• the same unshielded colliders

Exercises
1) Create a 4-variable DAG
2) Specify a 1-edge variant that is equivalent
3) Specify a 1-edge variant that is not
4) Show with IM and Estimators that you have succeeded in steps 2 and 3
Independence Equivalence Classes: Patterns & PAGs

• **Patterns** (Verma and Pearl, 1990): graphical representation of d-separation equivalence class (among models with no latent common causes)

• **PAGs**: (Richardson 1994) graphical representation of a d-separation equivalence class that includes models with latent common causes and sample selection bias that are d-separation equivalent over a set of measured variables X
Patterns

Possible Edges

\[
\begin{align*}
X_1 & \quad X_2 \\
X_1 & \quad X_2 \\
X_1 & \rightarrow X_2
\end{align*}
\]

Example

\[
\begin{align*}
X_1 & \rightarrow X_2 \\
X_3 & \rightarrow X_4
\end{align*}
\]
Patterns: What the Edges Mean

- $X_1$ and $X_2$ are not adjacent in any member of the equivalence class.

- $X_1 \rightarrow X_2$ ($X_1$ is a cause of $X_2$) in every member of the equivalence class.

- $X_1 \rightarrow X_2$ in some members of the equivalence class, and $X_2 \rightarrow X_1$ in others.
Patterns

Pattern

Represents

X_1 \quad X_2

\quad X_3 \quad X_4

\quad X_1 \quad X_2

\quad X_3 \quad X_4
Patterns

Specify all the causal graphs represented by the Pattern:

1) 

2) 

??

??
Patterns

Specify all the causal graphs represented by the Pattern:

1)

2)
Tetrad Demo: Generating Patterns

![Tetrad Graph Interface](image)
Causal Search Spaces are Large

• Directed Acyclic Graphs (between $2^{\binom{N}{2}}$ and $3^{\binom{N}{2}}$) ... $\binom{N}{2}$ is $O(N^2)$

• Directed Graphs (4\binom{N}{2})

• Markov Equivalence Class of DAGs (patterns): DAGs / 3.7

• Markov Equivalence Class of DAGs with confounders (roughly PAGs)

• Equivalence Class of “Linear Measurement Models”

• Equivalence Class of Directed Graphs with confounders

• Relative to: Experimental Setup $V = \{\text{Obs}, \text{Manip}\}$
Causal Search as a Method

Causal Knowledge
- e.g.,
  Markov Equivalence Class of Causal Graphs

Experimental Setup($V$)
- $V = \{\text{Obs, Manip}\}$
- $P(\text{Manip})$

Discovery Algorithm

Background Knowledge
- Salary $\rightarrow$ Gender
- Infection $\rightarrow$ Symptoms

General Assumptions
- Markov,
- Faithfulness
- Linearity
- Gaussianity
- Acyclicity

Data

Statistical Inference

P_{\text{Manip}}(V)
For Example

**Markov Equivalence Class of Causal Graphs (Pattern)**

- $X_1 \rightarrow X_2 \rightarrow X_3$
- $X_1 \leftrightarrow X_3$
- $X_1 \rightarrow X_2 \rightarrow X_3$

**Passive Observation**

$P(V) : X_1 \parallel X_3 \mid X_2$

**Discovery Algorithm**

**General Assumptions**
- Markov, Faithfulness, No latents, no cycles,

**Background Knowledge**
- $X_2$ prior in time to $X_3$

**Statistical Inference**

**Data**
Faithfulness

Constraints on a probability distribution $P$ generated by a causal structure $G$ hold for all parameterizations of $G$.

\[
\text{Revenues} := \beta_1 \text{Rate} + \beta_2 \text{Economy} + \varepsilon_{\text{Rev}}
\]

\[
\text{Economy} := \beta_3 \text{Rate} + \varepsilon_{\text{Econ}}
\]

\[
\text{Faithfulness:}
\]
\[
\beta_1 \neq -\beta_3 \beta_2
\]
\[
\beta_2 \neq -\beta_3 \beta_1
\]
Faithfulness

Constraints on a probability distribution \( P \) generated by a causal structure \( G \) hold for all parameterizations of \( G \).

All and only the constraints that hold in \( P(V) \) are entailed by the causal structure \( G(V) \), rather than lower dimensional surfaces in the parameter space.

Causal Markov Axiom:

\[ X \text{ and } Y \text{ causally disconnected} \quad \models \quad X \perp\!\!\!\!\!\!\perp Y \]

Faithfulness:

\[ X \text{ and } Y \text{ causally disconnected} \quad \not\models \quad X \perp\!\!\!\!\!\!\perp Y \]
Challenges to Faithfulness

By evolutionary design:
Gene A \(\|\|\) Protein 24

By evolutionary design:
Air temp \(\|\|\) Core Body Temp

Sampling Rate vs. Equilibration rate
Search Methods

• Constraint Based Searches
  • PC, FCI
  • Pointwise, but not uniformly consistent

• Scoring Searches
  • GES, FGS
  • Scores: BIC, AIC, etc.
  • Search: Hill Climb, Genetic Alg., Simulated Annealing
  • Difficult to extend to latent variable models
  • Meek and Chickering Greedy Equivalence Class (GES)
  • Pointwise, but not uniformly consistent

• Latent Variable Psychometric Model Search
  • BPC, MIMbuild, etc.

• Linear non-Gaussian models (Lingam)
• Models with cycles
• And more!!!
Score Based Search

Data

Model Score

Equivalence Class of Causal Graphs

X₁ → X₂ → X₃

X₁ ← X₂ ← X₃

X₁ ← X₂ → X₃

X₁ → X₂ ← X₃

Equivalence Class of Causal Graphs

Model Scores:
AIC, BIC, etc.

Background Knowledge

e.g., X₂ prior in time to X₃
Tetrad Demo and Hands On
Tetrad Demo and Hands-on

1) Go to “estimation1.tet”

2) Add Search node (from Data1)
   - Choose and execute one of the
     “Pattern searches”

3) Add a “Graph Manipulation” node to search
   result: “choose Dag in Pattern”

4) Add a PM to GraphManip

5) Estimate the PM on the data

6) Compare model-fit to model fit for true mode
Background Knowledge
Tetrad Demo and Hands-on

1) Create new session
2) Select “Search from Simulated Data” from Template menu
3) Build graph below, PM, IM, and generate sample data N=1,000.
4) Execute PC search, $\alpha = .05$
Background Knowledge
Tetrad Demo and Hands-on

1) Add “Knowledge” node
2) Create “Tiers” as shown below.
3) Execute PC search again, $\alpha = .05$
4) Compare results (Search2) to previous search (Search1)
Backround Knowledge
Direct and Indirect Consequences

True Graph

PC Output
Background Knowledge

PC Output
No Background Knowledge
Background Knowledge
Direct and Indirect Consequences

Direct Consequence
Of Background Knowledge

Indirect Consequence
Of Background Knowledge

PC Output
Background Knowledge

PC Output
No Background Knowledge
Charitable Giving (Search)

1) Load in charity data
2) Add search node
3) Enter Background Knowledge:
   - Tangibility is exogenous
   - Amount Donated is endogenous only
   - Tangibility $\rightarrow$ Imaginability is required
4) Choose and execute one of the “Pattern searches”
5) Add a “Graph Manipulation” node to search result: “choose Dag in Pattern”
6) Add a PM to GraphManip
7) Estimate the PM on the data
8) Compare model-fit to hypothetical model
Lead-IQ Search

1) Load in lead-iq data
2) Add search node
3) Enter Background Knowledge:
   • Ciq is endogenous
4) Choose and execute one of the “Pattern searches”
5) Add a “Graph Manipulation” node to search result: “choose Dag in Pattern”
6) Add a PM to GraphManip
7) Estimate the PM on the data
Extra Slides:
D-separation
D-separation

- Undirected Paths
- Colliders vs. Non-Colliders
D-separation: Undirected Paths

Undirected Path from X to Y:
- any sequence of edges beginning with X and ending at Y in which no edge repeats

Paths from X to Y:

Undirected Paths from X to Y:
D-separation: Undirected Paths

Undirected Path from X to Y:
• any sequence of edges beginning with X and ending at Y in which no edge repeats

Paths from X to Y:

1) \( X \leftrightarrow V \rightarrow Y \)
D-separation: Undirected Paths

Undirected Path from X to Y:
- any sequence of edges beginning with X and ending at Y in which no edge repeats

Paths from X to Y:

1) $X \leftarrow V \rightarrow Y$
2) $X \rightarrow Y$
D-separation: Undirected Paths

Undirected Path from X to Y:
- any sequence of edges beginning with X and ending at Y in which no edge repeats

Paths from X to Y:

1) X \leftarrow V \rightarrow Y

2) X \rightarrow Y

3) X \rightarrow Z1 \leftarrow W \rightarrow Y
D-separation: Undirected Paths

Undirected Path from X to Y:
- any sequence of edges beginning with X and ending at Y in which no edge repeats

Paths from X to Y:
1) $X \leftarrow V \rightarrow Y$
2) $X \rightarrow Y$
3) $X \rightarrow Z_1 \leftarrow W \rightarrow U \rightarrow Y$
4) $X \rightarrow Z_1 \leftarrow W \rightarrow U \rightarrow Y$
D-separation: Undirected Paths

Undirected Path from X to Y:
- any sequence of edges beginning with X and ending at Y in which no edge repeats

Paths from X to Y:
1) X ← V → Y
2) X → Y
3) X → Z1 ← W → Y
4) X → Z1 ← W → U → Y
5) X → Z1 → Z2 → U → Y
D-separation: Undirected Paths

Undirected Path from X to Y:
- any sequence of edges beginning with X and ending at Y in which no edge repeats

Paths from X to Y:
1) $X \leftarrow V \rightarrow Y$
2) $X \rightarrow Y$
3) $X \rightarrow Z_1 \leftarrow W \rightarrow Y$
4) $X \rightarrow Z_1 \leftarrow W \rightarrow U \rightarrow Y$
5) $X \rightarrow Z_1 \rightarrow Z_2 \rightarrow U \rightarrow Y$
6) $X \rightarrow Z_1 \rightarrow Z_2 \rightarrow U \leftarrow W \rightarrow Y$
D-separation: Undirected Paths

Undirected Path from X to Y:
- any sequence of edges beginning with X and ending at Y in which no edge repeats

Illegal Path from X to Y:

1) X ← Z₁ → Z₂ → U ← W → Z₁ → Z₂ → U → Y
Colliders

Y: *Collider*

Y: *Non-Collider*

Shielded

Unshielded
A variable is or is not a collider on a path

Variable: U

Paths from X to Y

\[ X \rightarrow Z1 \leftarrow W \rightarrow U \rightarrow Y \]

Paths on which U is a non-collider:
Colliders – *a variable on a path*

Variable: U

Paths from X to Y

\[ X \rightarrow Z_1 \leftarrow W \rightarrow U \rightarrow Y \]

Paths on which U is a non-collider:

\[ X \rightarrow Z_1 \rightarrow Z_2 \rightarrow U \rightarrow Y \]

Path on which U is a collider:
Colliders – a variable on a path

Variable: U

Paths from X to Y

$X \rightarrow Z_1 \leftarrow W \rightarrow U \rightarrow Y$

Paths on which U is a non-collider:

$X \rightarrow Z_1 \rightarrow Z_2 \rightarrow U \rightarrow Y$

Path on which U is a collider:

$X \rightarrow Z_1 \rightarrow Z_2 \rightarrow U \leftarrow W \rightarrow Y$
Conditioning on Colliders

*induce* Association

- Gas [\(y, n\)]
- Battery [\(\text{live, dead}\)]

Car Starts [\(y, n\)]

Gas \(\parallel\) Battery

Gas \(\parallel\) Battery | Car starts = no

Conditioning on Non-Colliders

*screen-off* Association

- Exp [\(y, n\)]
- Symptoms [\(\text{yes, no}\)]

Infection [\(y, n\)]

Exp \(\parallel\) Symptoms

Exp \(\parallel\) Symptoms | Infection
D-separation

X is \(d\)-separated from Y by \(Z\) in \(G\) iff
Every undirected path between X and Y in \(G\) is \(inactive\) relative to \(Z\)

An undirected path is \(inactive\) relative to \(Z\) iff
\(any\) node on the path is \(inactive\) relative to \(Z\)

A node \(N\) (on a path) is \(inactive\) relative to \(Z\) iff
a) \(N\) is a non-collider in \(Z\), or
b) \(N\) is a collider that is not in \(Z\), and has no descendant in \(Z\)

A node \(N\) (on a path) is \(active\) relative to \(Z\) iff
a) \(N\) is a non-collider not in \(Z\), or
b) \(N\) is a collider that is in \(Z\), or has a descendant in \(Z\)

\(X\) \(d\)-sep \(Y\) relative to \(Z\) = \(\emptyset\) ?

Undirected Paths between \(X\) , \(Y\):
1) \(X \rightarrow Z_1 \leftarrow W \rightarrow Y\)
2) \(X \leftarrow V \rightarrow Y\)
D-separation

X is \textit{d-separated} from Y by Z in G iff
Every undirected path between X and Y in G is \textit{inactive} relative to Z

An undirected path is \textit{inactive} relative to Z iff
\textit{any} node on the path is \textit{inactive} relative to Z

A node N (on a path) is \textit{inactive} relative to Z iff
\begin{enumerate}
\item N is a non-collider in Z, or
\item N is a collider that is not in Z, and has no descendant in Z
\end{enumerate}

A node N (on a path) is \textit{active} relative to Z iff
\begin{enumerate}
\item N is a non-collider not in Z, or
\item N is a collider that is in Z, or has a descendant in Z
\end{enumerate}

X \textit{d-Sep} Y relative to Z = $\emptyset$ ?

\begin{align*}
X \to Z_1 & \leftrightarrow W \to Y \text{ active? No} \\
1) \quad \text{Z1 active? No} \\
2) \quad \text{W active? Yes}
\end{align*}
D-separation

X is *d-separated* from Y by Z in G iff
Every undirected path between X and Y in G is *inactive* relative to Z

An undirected path is *inactive* relative to Z iff
any node on the path is *inactive* relative to Z

A node N (on a path) is *inactive* relative to Z iff
a) N is a non-collider in Z, or
b) N is a collider that is not in Z, and has no descendant in Z

A node N (on a path) is *active* relative to Z iff
a) N is a non-collider not in Z, or
b) N is a collider that is in Z, or has a descendant in Z

X d-sep Y relative to Z = ?  No

X ← V → Y active?  Yes

1) V active?  Yes
D-separation

X is \textit{d-separated} from Y by Z in G iff

Every undirected path between X and Y in G is inactive relative to Z

An undirected path is inactive relative to Z iff

any node on the path is inactive relative to Z

A node N is inactive relative to Z iff

a) N is a non-collider in Z, or
b) N is a collider that is not in Z, and has no descendant in Z

A node N (on a path) is \textit{active} relative to Z iff

a) N is a non-collider not in Z, or
b) N is a collider that is in Z, or has a descendant in Z

\[ X \text{ d-} \text{sep} \ Y \text{ relative to } Z = \{ W, Z_2 \} ? \]

Undirected Paths between X, Y:

1) \( X \to Z_1 \leftarrow W \to Y \)

2) \( X \leftarrow V \to Y \)
D-separation

X is *d-separated* from Y by Z in G iff
Every undirected path between X and Y in G is inactive relative to Z

An undirected path is inactive relative to Z iff
any node on the path is inactive relative to Z

A node N is inactive relative to Z iff
a) N is a non-collider in Z, or
b) N is a collider that is not in Z, and has no descendant in Z

A node N (on a path) is *active* relative to Z iff
a) N is a non-collider not in Z, or
b) N is a collider that is in Z, or has a descendant in Z

X d-sep Y relative to Z = \{W, Z_2\}?  No

1) X \rightarrow Z_1 \leftarrow W \rightarrow Y

Z1 active?  Yes
W active?  No
D-separation

X \rightarrow Z_1 \rightarrow Y
X \text{ d-sep } Y \text{ given } \emptyset \, ? \quad \text{No}
X \text{ d-sep } Y \text{ given } \{Z_1\} \, ? \quad \text{No}

X \downarrow \quad Y \downarrow
Z_1 \quad Z_2
X \text{ d-sep } Z_2 \text{ given } \emptyset \, ? \quad \text{No}
X \text{ d-sep } Z_2 \text{ given } \{Z_1\} \, ? \quad \text{No}
D-separation + Intervention:
Statistical Control ≠ Experimental Control

Question: Does $X_1$ directly cause $X_3$?

Truth: No, $X_2$ mediates

How to find out?

Experimentally control for $X_2$
D-separation + Intervention:

Statistical Control $\neq$ Experimental Control

Experimentally control for $X_2$

$X_3$ d-sep $X_1$ by \{$X_2$ set\} $\Leftrightarrow$

Yes: $X_3 \perp\!\!\!\!\perp X_1 \mid X_2$(set)

Statistically control for $X_2$

$X_3$ d-sep $X_1$ by \{$X_2$\} $\Leftrightarrow$

No! $X_3 \not\perp\!\!\!\!\perp X_1 \mid X_2$
Extra Slides:
Constraint based search
Constraint-based Search for Patterns

1) Adjacency phase

2) Orientation phase
Constraint-based Search for Patterns: Adjacency phase

X and Y are not adjacent if they are independent conditional on any subset that doesn’t X and Y

1) Adjacency
   • Begin with a fully connected undirected graph
   • Remove adjacency X-Y if X _||_ Y | any set S
Causal Graph

\[
\begin{align*}
X2 &\rightarrow X3 \\
X1 &\rightarrow X3
\end{align*}
\]

Independences

\[
\begin{align*}
X1 &\perp\!\!\!\perp X2 \\
X1 &\perp\!\!\!\perp X4 \mid \{X3\} \\
X2 &\perp\!\!\!\perp X4 \mid \{X3\}
\end{align*}
\]

Begin with:

\[
\begin{align*}
X1 &
\end{align*}
\]

From

\[
\begin{align*}
X1 &\perp\!\!\!\perp X2 \\
X1 &\perp\!\!\!\perp X4 \mid \{X3\}
\end{align*}
\]

From

\[
\begin{align*}
X2 &\perp\!\!\!\perp X4 \mid \{X3\}
\end{align*}
\]
2) Orientation

• Collider test:
  Find triples X – Y – Z, orient according to whether the set that separated X-Z contains Y

• Away from collider test:
  Find triples X → Y – Z, orient Y – Z connection via collider test

• Repeat until no further orientations

• Apply Meek Rules
Search: Orientation

Patterns

Y Unshielded

Test: $X \perp \perp Z \mid S$, is $Y \in S$

X \quad Y \quad Z

Collider

X \quad Y \quad Z

No

X \quad Y \quad Z

Yes

Non-Collider

X \quad Y \quad Z

X \quad Y \quad Z

X \quad Y \quad Z

X \quad Y \quad Z
Search: Orientation

Away from Collider

Test Conditions

1) $X_1 - X_2$ adjacent, and *into* $X_2$.
2) $X_2 - X_3$ adjacent
3) $X_1 - X_3$ not adjacent

Test

$X_1 \parallel X_3 \mid S, X_2 \in S$

No

Yes

$X_1 \rightarrow X_2 \rightarrow X_3$

$X_1 \rightarrow X_3 \rightarrow X_2$
Search: Orientation

After Adjacency Phase

Collider Test: X1 – X3 – X2

Away from Collider Test:

X1 → X3 – X4   X2 → X3 – X4

X1 _||_ X2

X1 _||_ X4 | X3
X2 _||_ X4 | X3
Away from Collider Power!

X₁ → X₂ ─── X₃  \[ X₁ \perp\perp X₃ \mid S, X₂ \in S \]

X₁ → X₂ → X₃

X₂ – X₃ oriented as X₂ → X₃

Why does this test also show that X₂ and X₃ are not confounded?

C

X₁ → X₂ → X₃

\[ X₁ \perp\perp X₃ \mid S, X₂ \in S \]

\[ X₁ \perp\perp X₃ \mid S, X₂ \in S, C \not\in S \]