Causal Structure Search:
Philosophical Foundations and Problems

Richard Scheines & Peter Spirtes
Carnegie Mellon University
Outline

1. Causal Learning (vs. Predictive Learning)
2. Recent Successes
4. Problems with the Standard Set-up
Causal Discovery - Goals

1) Policy, Law, and Science: How can we use data to answer
   a) *subjunctive* questions (effects of future policy interventions), or
   b) *counterfactual* questions (what would have happened had things been done differently (law)?
   c) *scientific* questions (what mechanisms run the world)

2) Rumsfeld Problem: Do we know what we don’t know: Can we tell when there is or is not enough information in the data to answer causal questions?
Causal Learning is Harder than Prediction

Data($X,Y$)

Prediction

Statistical Machine Learning

Causal Structure Learning Algorithm

Causal Structure(s) (Graph)

$P(Y,X)$

$P(Y|X)$

$P(Y|X_{set})$
Causal Learning is Limited, but Rumsfeld

Population \((X, Y)\)

\(P(X, Y), \text{Causal Graph}(X, Y)\)

Data\((X, Y)\)

Background Knowledge

Causal Structure Learning Algorithm

Equivalence Class of Causal Structures

\(P(Y \mid X_{\text{set}})\)

Population \((X_1, X_2, X_3)\)

\(X_1 \rightarrow X_2 \rightarrow X_3\)

\(P(X_1, X_2, X_3): X_1 \perp\!
\perp X_3 \mid X_2\)

Data\((X_1, X_2, X_3)\)

BK: \(X_2\) prior to \(X_3\)

No confounders

Causal Structure Learning Algorithm

Equivalence Class

\(X_1 \rightarrow X_2 \rightarrow X_3\)

No

\(P(X_1 \mid X_{2\text{set}})\)

Yes

\(P(X_3 \mid X_{2\text{set}})\)
Recent Successes

(Partial List!)

• Do-Calculus
• Identification
• Bounding
• Bayesian Search
• Time-varying confounders and conditionally randomized treatment
  (Jamie Robins)
• Dynamic Bayes Nets
• Equivalence Classes
  (patterns, PAGs, Factor Analytic Measurement Models)
Recent Successes
(Partial List!)

• Pointwise Consistent Discovery Algorithms
  (patterns, PAGs, MMs, SEM with pure MM, Linear-Cyclic Models)

• Discovery in Time Series
  (Granger & Swanson, Hoover, Bessler, Moneta)

• Linear, non-Gaussian models (Shimizu, Hoyer, Hyvarinen)

• Active Search
  (Cooper, Eberhardt, Tong, Kohler, Murphy, He & Gong)

• Overlapping Sets of Variables (Tillman & Danks)

• Applications (Ed. Research, Biology, Economics, Sociology, etc.)

• Causality Challenge!!
Philosophical Foundations of Causal Structure Learning

\[ V = \{M, L\} \quad M \text{ measured, } L = \text{ unobserved (latent)} \]

Causal structure over \( V \) \( \Rightarrow \) Constraints in \( P(V) \)

- **Assumption 1**: Weak Causal Markov Assumption
  \[ V_1, V_2 \text{ causally disconnected} \Rightarrow V_1 \perp \!\!\!\!\!\!\perp V_2 \]

- **Assumption 2a**: Causal Markov Axiom

- **Assumption 2b**: Determinism, e.g., Structural Equations
  For each \( V_i \in V, V_i := f(\text{parents}(V_i)) \)
Causal structure over $V \Rightarrow$ Constraints in $P(V)$

Causal Markov Axiom:

If $G$ is a causal graph, and $P$ a probability distribution over the variables in $G$, then in $P$: every variable $V$ is independent of its non-effects, conditional on its immediate causes.
Faithfulness

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

\[ \text{Revenues} = a \text{Rate} + c \text{Economy} + \varepsilon_{\text{Rev.}} \]
\[ \text{Economy} = b \text{Rate} + \varepsilon_{\text{Econ.}} \]

Faithfulness: $a \neq -bc$
Modularity of Intervention/Manipulation

**Structural Equations:**
- Education = $\varepsilon_{ed}$
- Longevity = $f_1 \text{(Education)} + \varepsilon_{Longevity}$
- Income = $f_2 \text{(Education)} + \varepsilon_{income}$

**Manipulated Structural Equations:**
- Education = $\varepsilon_{ed}$
- Longevity = $f_1 \text{(Education)} + \varepsilon_{Longevity}$
- Income = $f_3 \text{(M1)}$
Structural Equations:
Education = \varepsilon_{ed}
Longevity = f_1 (Education) + \varepsilon_{Longevity}
Income = f_2 (Education) + \varepsilon_{Income}

Manipulated Structural Equations:
Education = \varepsilon_{ed}
Longevity = f_1 (Education) + \varepsilon_{Longevity}
Income = f_3 (M2,Education) + \varepsilon_{Income}
The Standard Set-up

- Measured Vars $M$ given
- $V = \{M, L\}$ satisfy Markov, Faithfulness, Modularity
- Tasks:
  - Discover structure (e.g., causal relations) among $M$
  - Estimate causal parameters
  - Less often:
    - Discover existence of $L$
    - Discover and estimate causal relations among $L$
Problems with the Standard Set-up

• Faithfulness in Redundant or Thermostatic Mechanisms

• Measurement
  • Classical Measurement Error
  • Coarsening
  • Aggregation

• Ambiguous Manipulations

• Modularity in Constraint Based, Reversible Systems

• Variable Construction / Decision Theory
Faithfulness

• Redundant Mechanisms

Gene A + Protein
   - Gene B +

Gene A _||_ Protein

• Thermostatic Equilibrium

Air Temp Target - Core
    + Sweat/Heatup

Core Temp

Air Temp _||_ Core Temp
Classical Measurement Error

\[ Z' = Z + \epsilon \]

Measurement Error: \[ Z' = Z + \epsilon \]

\[ X \perp\perp Y \mid Z \]

\[ X \perp\perp Y \mid Z' \] unless \( \text{Var}(\epsilon') = 0 \)
Coarsening

$Smoking_{coarse}$

- Ever smoked before age 50 [y,n]

$Smoking_{precise}$

- Exact amount smoked before age 50

$Tar\_stains\_precise$

- Exact amount of tar-stains on fingers at age 50

$Lung\_Cancer$

- By age 60

$Lung\_Cancer \parallel Tar\_stains\_precise \parallel Smoking\_precise$

$Lung\_Cancer \parallel Tar\_stains\_precise \parallel Smoking\_coarse$
TV → Obesity


Goals:
- Estimate the influence of TV on BMI
- Tease apart the mechanisms (diet, exercise)
Measures of Exercise, Diet

Exercise_M: L ← Calories expended in exercise in bottom two tertiles
Exercise_M: H ← Calories expended in exercise in top tertile

Diet_M: L ← Calories consumed in bottom two tertiles
Diet_M: H ← Calories consumed in top tertile
Measures of Exercise, Diet

Findings:
• TV and Obesity NOT screened off by Exercise_M & Diet_M
• Bias in mechanism estimation unknown
Microarrays: measured gene expressions are *sумs* of gene expression across all cells in tissue sample

\[ \forall \text{ Cells: } X \perp \perp Y \mid Z \]

\[ \Sigma_n X \perp \perp \Sigma_n Y \mid \Sigma_n Z \]

unless \( P(X,Y,Z) \) is special, e.g., Gaussian
Causal Discovery in fMRI

∀i,j : Xi _||_ Yj | {Z}

fMRI measures aggregate activity in a voxel

Variables aggregate activity over voxels

Σ X _||_ Σ Y | Σ Z
1960s: In RCTs, drugs that reduce $TC$ (total cholesterol), reduce the risk of $DH$ (Heart Disease).

- $P(DH \mid TC_{set})$ identifiable.

- $TC \equiv_{def} f(LDL, HDL)$, high-density & low-density cholesterol
Ambiguous Manipulations


- $HDL=L, \ LDL=L \implies TC=L$
- $HDL=L, \ LDL=H \implies TC=M$
- $HDL=H, \ LDL=L$
- $HDL=H, \ LDL=H \implies TC=H$

- Arrows in boldface are definitional links
Suppose HDL, LDL unobserved

* TC cannot be manipulated independently of both HDL and LDL

* “Set TC to M” is ambiguous over:
  - HDL = H and LDL = L
  - HDL = L and HDL = H
Suppose $HDL = H$ and $LDL = L$ prevents $H$, and $HDL = L$ and $HDL = H$ promotes $H$?

What is $P(DH \mid TC_{set} = M)$?

Can ambiguity be detected?

- Need additional assumptions? Yes, e.g., variability
- From observational data? Sometimes
- Will positive causal hypotheses be inferred involving variables whose effect is ambiguous? Probably not
Reversible/Constraint Systems

- $PV = nRT$
- Constraint persists, even with surgical interventions
- “joint” part of $P(V,T,P)$ remains unaltered by any intervention.
- Is there a causal graph and parameterization thereof such that the constraint holds for any permissable set of surgically altered equations?
- Can such systems be learned without intervention?
Variable construction can be framed as a search problem, thus a decision problem

Decision problem for prediction ? decision problem for causal learning
Variable Construction for Causal Learning

**Raw Data**
- Voxels in fMRI

**Features/Variables**
- Activity in a Brain Region
Variable Construction for Causal Learning

- Adjust for interunit – anatomical matching
- Correct for time lag of hemodynamic response & scan time
- Identify voxels with statistically improbable signals
- Cluster, usually by eyeball
- Variables constructed =
  - mean of signal intensity in cluster
  - one of the first 4 principal components
  - average intensity of top X% variance voxels
  - maximum variance voxel
  - non-contiguous regions
  - possibly overlapping

Raw Data
- Voxels in fMRI

Features/Variables
- Activity in a Brain Region
Decision Theory for Causal Learning

• Positive utility on increasing an output from baseline (e.g., learning in online course, brain activity in region associated with emotional intelligence among autistic children)

• Intervention on 1 variable, leave cost aside.

• Raw data $\rightarrow$ constructed variables $\rightarrow$ causal search algorithm

• Compute expected utility of intervention

• Uncertainty over:
  • causal structure
  • parameters in a given causal structure
Model Uncertainty for 1 set of constructed variables

\[
EU(\text{do}(X)) = EU(\text{do}(X) \text{ in } EC1) \cdot P(EC1) + EU(\text{do}(X) \text{ in } EC2) \cdot P(EC2)
\]

\[
EU(\text{do}(X) \text{ in } EC1) = EU(\text{do}(X) \text{ in } \text{DAG}_i \text{ in EC1}) \cdot P(\text{DAG}_i \text{ in EC1}) + ... 
\]

\[
EU(\text{do}(X) \text{ in } \text{DAG}_i \text{ in EC1}) = \int \text{EU}(\text{do}(X) \text{ in } \text{DAG}_i, \alpha = x) \, dx 
\]
Model Uncertainty for *many sets* of constructed variables

- EU(do(X)) vs. EU(do(X')) vs. EU(do(X''))
- Meaningful prior over models in output for each VC regime?
Thanks