Causal Inference and Invariance

Qingyuan Zhao and Charles Zheng

Stanford University

February 21, 2016

(Part 2/2)

From Last Week: Causal Graph

Causal relationships in a system represented by a graph. The graph tells you:

I. which variables are affected by an intervention.
II. what conditional independence relationships exist in the joint distribution (*d-separation*.)
III. which sets of predictors and responses will have “invariant” optimal predictive rules.

This talk is restricted to directed acyclic graph (DAG), i.e. no feedback!

From Last Week: Three *Causal* Questions

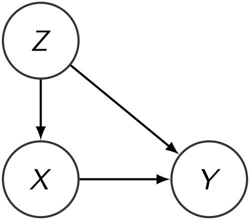
- Given a number of variables, which pairs are causally related?
  - Infer the *graph*.
- Given a number of variables and a fixed Y, which variables causally affect Y?
  - Infer the *invariance set*.
- Given a fixed X and a fixed Y, what is the causal effect of X on Y?
  - Infer the *causal effect*.

Why different languages? Convenience!

Section 1

Overview of Previous methods

Known Causal Structure



For example, suppose we want to estimate the causal effect of X on Y with known confounders Z.

- Graphical approach: the backdoor formula

$$\mathrm{P}(y|do(x)) = \sum_{z} \mathrm{P}(y|x,z)\mathrm{P}(z).$$

- Functional appraoch: outcome regression $Y \sim X + Z$.
- Potential outcome approach: estimate the propensity score.

Unknown Causal Structure

Conventional approach:

1. Estimate the Markov equivalence class of causal graphs via conditional independence relationships.
2. Infer or bound the identifiable causal effects.

More recent approach: impose additional functional/distributional assumptions to the structural equation model: for any variable Y,

$$Y = f(\text{parents}(Y); \epsilon_Y).$$

How should we think about the assumptions?

One thing for sure: They are no monsters!

How should we think about the assumptions?

- In statistics we make assumptions all the time: parametric, independence, function form, etc.
  - George Box: "All models are wrong but some are useful".
- To infer causation, we need to make different kinds of assumptions.
  - Problem statement: Can what we learned from this environment be generalized to another environment?
  - The ancient wisdom: "Correlation does not imply causation" (observational $\not\Rightarrow$ interventional).
  - Causal assumptions: causal graph, structural equation model, or invariant prediction.

What if we are willing to make both kinds of assumptions?

Section 2

Invariance

Assumed invariance

Focus: Given a number of variables and a fixed Y, which variables causally affect Y?
Data: i.i.d. samples of (X^e, Y^e) from different environments $e \in \mathcal{E}$.

Assumption (Invariant prediction)

There exists a vector of coefficients γ^* with support S^* such that for all $e \in \mathcal{E}$, X^e has an arbitrary distribution and

$$Y^e = \mu + X^e \gamma^* + \epsilon^e, \; \epsilon^e \sim F_\epsilon, \; \epsilon^e \perp\!\!\!\perp X^e_{S^*}.$$

Important:

- F_ϵ does not depend on e.
- ϵ is always independent of X.

This is essentially a single structural equation with $\mathrm{parents}(Y) = S^*$.

Building block

Testing the null hypothesis that (γ, S) satisfies the assumption.

$$\begin{aligned} H_{0,\gamma,S}(\mathcal{E}):\ & \gamma_k = 0 \text{ if } k \in S,\ \mathit{and}\ \exists F_\epsilon \text{ such that for all } e \in \mathcal{E}, \\ & Y^e = X^e\gamma + \epsilon^e,\ \epsilon^e \sim F_\epsilon,\ \epsilon^e \perp\!\!\!\perp X^e_S. \\ H_{0,S}(\mathcal{E}):\ & \exists\gamma \text{ such that } H_{0,\gamma,S}(\mathcal{E}) \text{ is true.} \end{aligned}$$