$N - 1$ Experiments Suffice to Determine the Causal Relations Among N Variables

F. Eberhardt   C. Glymour   R. Scheines

Carnegie Mellon University

27th of June 2006, Causality Reading Group

Presented by Simon Meier

meier-simon@student.ethz.ch

Outline

Inferring the causal structure

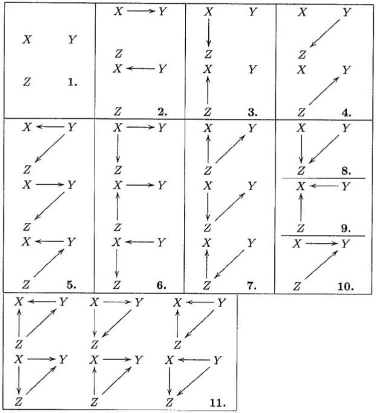
Under the following **assumptions**:

- Faithfulness The probability distribution over the variables is faithful to a directed acyclic graph on the variables, i.e., *no feedback relations* and *no hidden causes*
- Causal Markov Assumption Direct causes of a variable screen it off from variables that are not its effects
- Full scale D-Separation Any d-separation question can be decided
- Perfect Data The data is not supposed to be of any concern. In particular, *weak causal links*, *insufficient* or *missing data* is of no concern
- Interventions possible on every variable

Infer the **causal structure** of a system consisting of N variables using only interventions on single variables.

Clarifying The Problem

An example of inference on a system of three variables



Assume the true causal structure is the top graph in box 6.

Let us infer this structure under the assumptions from the slide before:

1. Randomize Y: PC-Algorithm gives

   Adjacencies $\{X, Y\}$
   Covariances None

   X is adjacent to Y, but it does not covary with Y. Thus X is a direct cause of Y.

2. Randomize Z: PC-Algorithm gives

   Adjacencies $\{X, Z\}$
   Covariances None

   X is adjacent to Z, but it does not covary with Z. Thus X is a direct cause of Z.

ALGORITHM (PC-Algorithm)

Assumptions *Faithfulness, full scale D-Separation*

Input *A set V of variables*

Output *An undirected graph $G = (V, E)$ with $E = \{ \{u, v\} \mid u, v \in V$, "$u$ directly causes v or v directly causes u" $\}$*

Definition (Direct Cause)

Given a set V of variables, for $X \in V$ and $Y \in V$ we say that X directly causes Y, *if there exists* some assignment of values v to the variables in $V - \{X, Y\}$ *such that* Y covaries with X when randomizing X while holding the variables in $V - \{X, Y\}$ fixed to the values v.

Inferring the Causal Structure

Algorithm (Infer Causal Structure)

Assumptions *Faithfulness, full scale D-Separation, causal Markov assumption, perfect data, interventions possible on every variable*

Input *A set V of variables, $|V| = N$*

Output *A directed acyclic graph $G = (V, E)$ with*
$E = \{ (u, v) \mid u, v \in V, \text{“}u \text{ directly causes } v\text{”} \}$

1. LET $V = \{X_1, \ldots, X_N\}$
2. FOR EACH $k \in \{1, \ldots, N-1\}$ *do the following experiment:*
   Randomize X_k
   Compute the “adjacency” graph G_k *of the resulting joint probability distribution over* V *using the PC-Algorithm*
3. LET $\tilde{G} = (V, \bigcup_{i=1}^{N-1} E(G_k))$ *and* $G = (V, \{\})$
4. FOR EACH $\{u, v\} \in \tilde{G}$ LET *w.l.o.g.* $u < v$:
   IF *v covaried with u in the u^{th} experiment* THEN
   add edge (u, v) to G, i.e., u is a direct cause of v
   ELSE
   add edge (v, u) to G, i.e., v is a direct cause of u

Definition (Experiment)

An experiment randomizes at most one variable and returns the resulting joint distribution of all variables.

Definition (Procedure)

A procedure is a sequence of experiments and a structure learning algorithm applied to the results of these experiments. When applied to a N variable problem, it is

- reliable *iff* it determines the correct graph uniquely for all DAGs on N vertices
- order reliable *iff* it is reliable for all orderings of experiments
- adaptive *iff* it chooses at each step the next experiment based on the results of the previous experiments

Proposition (Tight Bound)

No order reliable procedure randomizing a single variable at each step requires fewer than $N - 1$ experiments for an N variable problem in the worst case.

Proposition (Adaptive Procedure No Better)

Every reliable adaptive procedure for which each experiment randomizes a single variable requires, in the worst case, at least $N - 1$ experiments for an N variable problem.

Proposition (Passive Observation)

For $N \geq 3$, doing an experiment without an intervention, i.e., a passive observation, is of no use.

Proposition (Multiple Interventions)

Randomizing several variables in the same experiment can shorten the experimental sequence (cf. Section 3). But controlling for variables by experimentally holding their values constant is never an advantage.

Questions

Practical Relevance

1. What use does this algorithm have?
2. How could we weaken the assumptions?

Extensions

1. What would be the best procedure with respect to the expected number of experiments for a given probability distribution over the DAGs (e.g. the uniform distribution)?
2. How could we incorporate multiple interventions to shorten the experimental sequence?
3. How could we extend the algorithm to probability distributions over the variables that include common hidden causes?