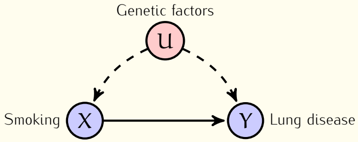
Stability of causal inference: Open problems

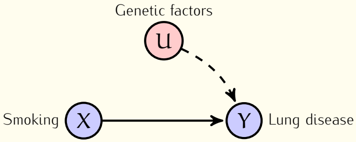
Leonard J. Schulman & Piyush Srivastava

California Institute of Technology

UAI 2016 Workshop on Causality

Interventions without experiments [Pearl, 1995]





Observational distribution

$$P(X, Y)$$

$$\sum_u P(U = u) P(X|u) P(Y|X, u)$$

→

Intervention distribution

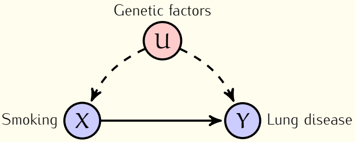
$$P(Y \mid do(X = x))$$

$$\sum_u P(U = u) P(Y|X = x, u)$$

Identification problem [Pearl, 1995]

When is $P(Y = y | do(X = x))$ computable given the observed distribution P?

Interventions without experiments [Pearl, 1995]

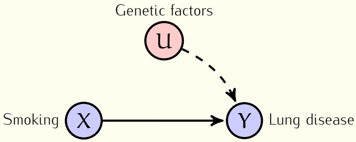


Observational distribution

$$P(X, Y)$$

$$\sum_u P(U = u)P(X|u)P(Y|X, u)$$

→



Intervention distribution

$$P(Y \mid do(X = x))$$

$$\sum_u P(U = u)P(Y|X = x, u)$$

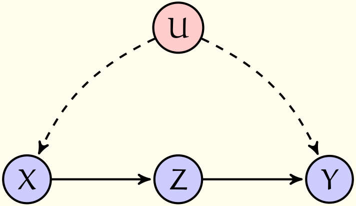
Identification problem [Pearl, 1995]

When is $P(Y = y| do(X = x))$ computable given the observed distribution P?

Not always!

Identifiable models

But sometimes it is...



Identification

$$P(Y \mid do(X = x)) = \sum_z P(Z = z | X = x) \cdot \sum_{x'} P(X = x') P(Y = y \mid Z = z, X = x').$$

Deciding identifiability

A long line of work culminated in the following striking result

Complete Identification [Huang and Valtorta, 2008; Shpitser and Pearl, 2006, ...]

An efficient algorithm with the following characteristics exists:

Input: Semi-Markovian graph $G = (V, E, U, D)$, disjoint subsets X, Y of V

Output: Either

- A rational map

$$\mathsf{ID}(G, X, Y) : P(V) \mapsto P(Y \mid \mathsf{do}(X)), \text{ or}$$

- A certificate of non-existence of such a map

Note

- The observed distribution P is not an input to the algorithm
- The output is not numerical, but a symbolic, exact description of the map

Deciding identifiability

A long line of work culminated in the following striking result

Complete Identification [Huang and Valtorta, 2008; Shpitser and Pearl, 2006, ...]

An efficient algorithm with the following characteristics exists:

Input: Semi-Markovian graph $G = (V, E, U, D)$, disjoint subsets X, Y of V

Output: Either

- A rational map

$$\mathsf{ID}(G, X, Y) : P(V) \mapsto P(Y \mid \mathsf{do}(X)), \text{ or}$$

- A certificate of non-existence of such a map

ID assumes...

- Exact knowledge of observed distribution P
- Exact knowledge of the model G (no "missing" edges)

Stability of the identification map

$$G = (V, E, U, D) \text{ is a semi-Markovian graph}$$
$$\mathsf{ID}(G, X, Y) : P(V) \mapsto P(Y \mid \mathsf{do}(X))$$

Statistical stability

How sensitive is $\mathsf{ID}(G, X, Y)$ to small perturbations in the input P?

Stability of the identification map

$$G = (V, E, U, D) \text{ is a semi-Markovian graph}$$
$$\mathsf{ID}(G, X, Y) : P(V) \mapsto P(Y \mid \mathsf{do}(X))$$

Statistical stability

How sensitive is $\mathsf{ID}(G, X, Y)$ to small perturbations in the input P?

Model Stability

How sensitive is $\mathsf{ID}(G, X, Y)$ to extra assumptions (missing edges) in G?

Perturbations in the input: Condition number

$$G = (V, E, U, D) \text{ is a semi-Markovian graph}$$
$$\mathsf{ID}(G, X, Y) : P(V) \mapsto P(Y \mid \mathsf{do}(X))$$

Suppose instead of P, we get $\tilde{P}$ as input to $\mathsf{ID}(G, X, Y)$, such that

$$\left| \log \frac{\tilde{P}(\cdot)}{P(\cdot)} \right| \leqslant \epsilon \quad \equiv \quad \mathsf{Rel}\, P \leqslant \epsilon, \text{ in } \| \cdot \|_\infty \text{ norm}$$

Condition number

$$\kappa_{\mathsf{ID}(G,X,Y)} = \sup \frac{\mathsf{Rel}\, P(Y | \mathsf{do}(X))}{\mathsf{Rel}\, P}$$

How large is the relative error in the output compared to that in the input?

Perturbations in the input: Condition number

$$G = (V, E, U, D) \text{ is a semi-Markovian graph}$$
$$\mathsf{ID}(G, X, Y) : P(V) \mapsto P(Y \mid \mathsf{do}(X))$$

Suppose instead of P, we get $\tilde{P}$ as input to $\mathsf{ID}(G, X, Y)$, such that

$$\left|\log \frac{\tilde{P}(\cdot)}{P(\cdot)}\right| \leqslant \epsilon \quad \equiv \quad \mathsf{Rel}\, P \leqslant \epsilon, \text{ in } \|\cdot\|_\infty \text{ norm}$$

Condition number

$$\kappa_{\mathsf{ID}(G,X,Y)} = \lim_{\epsilon \downarrow 0} \sup_{\mathsf{Rel}\, P \leqslant \epsilon} \frac{\mathsf{Rel}\, P(Y | \mathsf{do}(X))}{\mathsf{Rel}\, P}$$

How large is the relative error in the output compared to that in the input?

e.g., κ for computing conditional probabilities from P is at most 2.

Perturbations in the input: Condition number

$$G = (V, E, U, D) \text{ is a semi-Markovian graph}$$
$$\mathsf{ID}(G, X, Y) : P(V) \mapsto P(Y \mid \mathsf{do}(X))$$

Suppose instead of P, we get $\tilde{P}$ as input to $\mathsf{ID}(G, X, Y)$, such that

$$\left|\log \frac{\tilde{P}(\cdot)}{P(\cdot)}\right| \leqslant \epsilon \quad \equiv \quad \mathsf{Rel}\, P \leqslant \epsilon, \text{ in } \|\cdot\|_\infty \text{ norm}$$

Sources of perturbations

- Standard model for floating-point round off in numerical analysis

Perturbations in the input: Condition number

$$G = (V, E, U, D) \text{ is a semi-Markovian graph}$$

$$\mathsf{ID}(G, X, Y) : P(V) \mapsto P(Y \mid \mathsf{do}(X))$$

Suppose instead of P, we get $\tilde{P}$ as input to $\mathsf{ID}(G, X, Y)$, such that

$$\left| \log \frac{\tilde{P}(\cdot)}{P(\cdot)} \right| \leqslant \epsilon \quad \equiv \quad \mathsf{Rel}\, P \leqslant \epsilon, \text{ in } \|\cdot\|_\infty \text{ norm}$$

Sources of perturbations

- Standard model for floating-point round off in numerical analysis
- Statistical sampling errors: usually additive (even worse)

Perturbations in the input: Condition number

$$G = (V, E, U, D) \text{ is a semi-Markovian graph}$$
$$\mathsf{ID}(G, X, Y) : P(V) \mapsto P(Y \mid \mathsf{do}(X))$$

Suppose instead of P, we get $\tilde{P}$ as input to $\mathsf{ID}(G, X, Y)$, such that

$$\left|\log \frac{\tilde{P}(\cdot)}{P(\cdot)}\right| \leqslant \epsilon \quad \equiv \quad \mathsf{Rel}\, P \leqslant \epsilon, \text{ in } \|\cdot\|_\infty \text{ norm}$$

Sources of perturbations

- Standard model for floating-point round off in numerical analysis
- Statistical sampling errors: usually additive (even worse)
- Intentionally introduced errors: e.g. by some differential privacy mechanisms

Perturbations in the input: Inaccurate models

$$G = (V, E, U, D) \text{ is a semi-Markovian graph}$$
$$\mathsf{ID}(G, X, Y) : P(V) \mapsto P(Y \mid \mathsf{do}(X))$$

Suppose instead of P, we get $\tilde{P}$ as input to $\mathsf{ID}(G, X, Y)$, such that

$$(1 - \epsilon) \leqslant \frac{\tilde{P}(\cdot)}{P(\cdot)} \leqslant (1 + \epsilon) \quad \equiv \quad \mathsf{Rel}\, P \leqslant \epsilon, \text{ in } \|\cdot\|_\infty \text{ norm}$$

Ignoring "weak" edges

The same framework of perturbations to P can handle "model stability" as well!

[see paper for details]

Condition number of causal identification

Theorem: There exist highly ill-conditioned examples!

There exists an infinite sequence of semi-Markovian graphs G_n with n observed vertices and disjoint subsets S_n and T_n of the observed vertices such that

$$\kappa_{\mathsf{ID}(G_n, T_n, S_n)} = \exp\left(\Omega\left(n^{0.49}\right)\right)$$

- This is a property of the ID map itself, not of an algorithm computing it!

Condition vs. Stability

Paper available at http://www.caltech.edu/~piyushs and in UAI 2016 proceedings

Condition number of causal identification

Theorem: There exist highly ill-conditioned examples!

There exists an infinite sequence of semi-Markovian graphs G_n with n observed vertices and disjoint subsets S_n and T_n of the observed vertices such that

$$\kappa_{\mathsf{ID}(G_n, T_n, S_n)} = \exp\left(\Omega\left(n^{0.49}\right)\right)$$

- This is a property of the **ID** map itself, not of an algorithm computing it!

On these examples, any algorithm computing **ID** may lose
$\Omega\left(n^{0.49}\right)$
bits of precision

Condition vs. Stability

Paper available at http://www.caltech.edu/~piyushs and in UAI 2016 proceedings

Condition number of causal identification

Theorem: Good examples exist as well

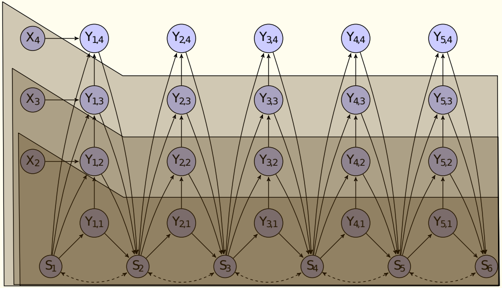
Let G be a semi-Markovian graph and let X be an observed node in G such that it is not possible to reach a child of X from X using only the hidden edges. Then, for any subset S of V not containing X.

$$\kappa_{\mathrm{ID}(G,X,S)} = O(|V|).$$

- Identifiability under the above condition was proved by Tian and Pearl [2002]

Paper available at http://www.caltech.edu/~piyushs and in UAI 2016 proceedings

The bad example ($m = 6, k = 4$): $P(\boldsymbol{S} \mid do(\boldsymbol{X}, \boldsymbol{Y}))$



Problem: Characterize models based on condition number

Question

Given a semi-Markovian model G and subsets S and T,
output tight bounds on the condition number of $\mathsf{ID}(\mathsf{G}, S, \mathsf{T})$

Problem: Characterize models based on condition number

Question

Given a semi-Markovian model G and subsets S and T,
output tight bounds on the condition number of $\mathbf{ID}(G, S, T)$

Why?

- If an identification problem is very badly conditioned, one needs very precise data for causal identification to be useful

Problem: Characterize models based on condition number

Question

Given a semi-Markovian model G and subsets S and T,
output tight bounds on the condition number of $\mathbf{ID}(G, S, T)$

Why?

- If an identification problem is very badly conditioned, one needs very precise data for causal identification to be useful
- Evaluate different models w.r.t. utility for causal identification

Problem: Characterize models based on condition number

Question

Given a semi-Markovian model G and subsets S and T, output tight bounds on the condition number of $\mathsf{ID}(\mathsf{G}, S, \mathsf{T})$

Why?

- If an identification problem is very badly conditioned, one needs very precise data for causal identification to be useful
- Evaluate different models w.r.t. utility for causal identification
- A more indirect reason...

Approximate causal identification through stability

Suppose $P(S \mid do(T))$ is not identifiable in G,
but becomes identifiable if some "weak edges" are removed

Approximate causal identification through stability

Suppose $P(S \mid \mathsf{do}(T))$ is not identifiable in G,
but becomes identifiable if some "weak edges" are removed

ϵ-weak edge

$e = (A, B)$ is ϵ-weak if for all a, a',

$$\left| \log \frac{P(B = \cdot \mid \mathsf{parents}(B) = \cdot, A = a)}{P(B = \cdot \mid \mathsf{parents}(B) = \cdot, A = a')} \right| \leqslant \epsilon$$

Approximate causal identification through stability

Suppose $P(S \mid do(T))$ is not identifiable in G,
but becomes identifiable if some “weak edges” are removed

ϵ-weak edge

$e = (A, B)$ is ϵ-weak if for all a, a',

$$\left|\log \frac{P(B = \cdot \mid \text{parents}(B) = \cdot, A = a)}{P(B = \cdot \mid \text{parents}(B) = \cdot, A = a')}\right| \leqslant \epsilon$$

Observation: Approximate identification of unidentifiable models

If $P(S \mid do(T))$ is identifiable in $G' = G - \{e\}$, and

$$\kappa_{\mathsf{ID}(G',T,S)} \leqslant \alpha,$$

then

$$\left|\log \frac{P_G(S|do(X))}{\mathsf{ID}(G', T, S)(P)}\right| \leqslant O(\alpha \cdot \epsilon)$$

Appendix

Condition number and numerical stability

Condition number is a property of the function
Numerical stability is a property of a floating point algorithm

$$\mathsf{ADD} : (x_1, x_2, \ldots, x_n) \mapsto x_1 + x_2 \ldots x_n$$

Condition number

$$\kappa = \frac{\sum_{i=1}^{n} |x_i|}{|\sum_{i=1}^{n} x_i|} = 1, \text{ for positive } x_i$$

Numerical stability: Naive linear summation

$$O(n \cdot \varepsilon \cdot \kappa)$$

Numerical stability: Kahan summation

$$O(\varepsilon \cdot \kappa), \text{ to first order in } \varepsilon$$

ε is the "machine epsilon"

Bibliography I

Yimin Huang and Marco Valtorta. On the completeness of an identifiability algorithm for semi-Markovian models. *Ann. Math. Artif. Intell.*, 54(4):363–408, December 2008. ISSN 1012-2443, 1573-7470. doi: 10.1007/s10472-008-9101-x. URL http://link.springer.com/article/10.1007/s10472-008-9101-x.

Judea Pearl. Causal diagrams for empirical research. *Biometrika*, 82(4):669–688, December 1995. ISSN 0006-3444, 1464-3510. doi: 10.1093/biomet/82.4.669. URL http://biomet.oxfordjournals.org/content/82/4/669.

Ilya Shpitser and Judea Pearl. Identification of joint interventional distributions in recursive semi-Markovian causal models. In *Proc. 20th AAAI Conference on Artifical Intelligence*, pages 1219–1226. AAAI Press, July 2006. URL http://www.aaai.org/Papers/AAAI/2006/AAAI06-191.pdf.

Jin Tian and Judea Pearl. A general identification condition for causal effects. In *Proceedings of the Eighteenth National Conference on Artificial Intelligence*, pages 567–573, 2002.