# Stability of causal inference

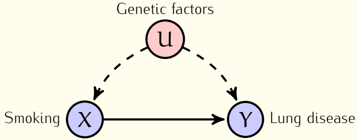
Leonard J. Schulman & Piyush Srivastava

California Institute of Technology

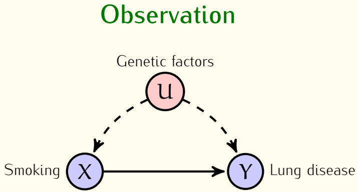
UAI 2016

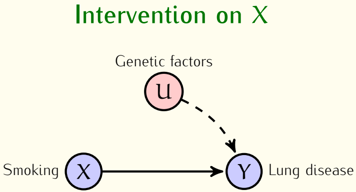
# Causal identification: Experimental intervention

## Observation



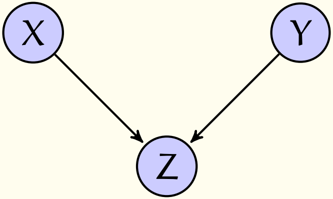
# Causal identification: Experimental intervention



# Directed graphical models

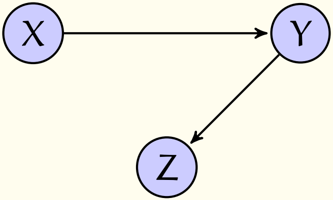
- A directed acyclic graph $G = (V, E)$ whose nodes are random variables
- Absent edges represent conditional independence assumptions




$$\begin{aligned} P(X, Y, Z) &= P(X)P(Y|X)P(Z|X, Y) \\ &= P(X)P(Y)P(Z|X, Y), \text{ due to model constraints} \end{aligned}$$

# Directed graphical models

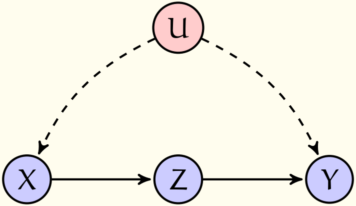
- A directed acyclic graph $G = (V, E)$ whose nodes are random variables
- Absent edges represent conditional independence assumptions




$$\begin{aligned} P(X, Y, Z) &= P(X)P(Y|X)P(Z|X, Y) \\ &= P(X)P(Y|X)P(Z|Y), \text{ due to model constraints} \end{aligned}$$

# Semi-Markovian models

- A Markovian model with some nodes hidden
- Hidden nodes have no parents

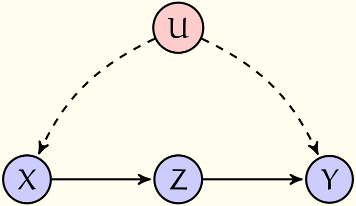



## Observed distribution

$$P(X, Y, Z) := \sum_{u} P(U = u)P(X|U = u)P(Z|X)P(Y|X, U = u)$$

# Semi-Markovian models

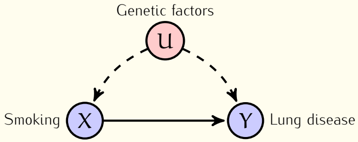
- A Markovian model with some nodes hidden
- Hidden nodes have no parents




## Observed distribution

$$
\begin{aligned}
P(X, Y, Z) &:= \sum_{u} P(U = u)P(X|U = u)P(Z|X)P(Y|X, U = u) \\
&= P(X)P(Z|X)P(Y|Z, X)
\end{aligned}
$$

# Interventions without experiments [Pearl, 1995]

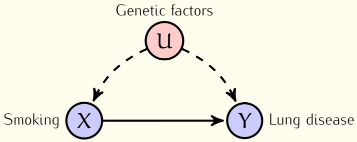



## Observational distribution

$$P(X, Y)$$

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

# Interventions without experiments [Pearl, 1995]

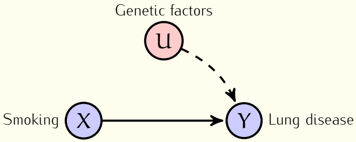



### Observational distribution

$$P(X, Y)$$

$$\sum_{u} P(U = u) P(X|u) P(Y|X, u)$$

$\longrightarrow$




### Intervention distribution

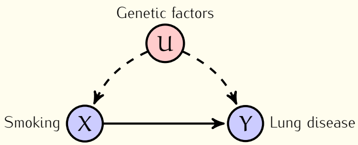
$$P(Y \mid do(X = x))$$

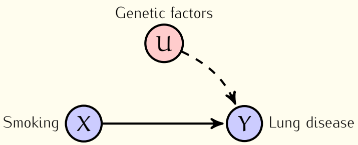
$$\sum_{u} P(U = u) P(Y|X = x, u)$$

### 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

$$P(X, Y)$$

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

$\longrightarrow$

## Intervention distribution

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

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

## Identification problem

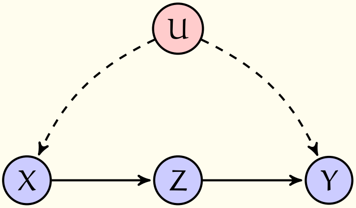
[**Pearl, 1995**]

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

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

## Condition number

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

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

# Perturbations in the input: Condition number

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

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

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

## Condition number

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

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

e.g., $\kappa$ 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

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

## 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

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

## 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

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

## 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]

# Results: Condition 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!

# Results: Condition 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

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

- 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

# Results: Condition of causal identification

**Theorem: An important class of well-conditioned examples**

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 $X$.

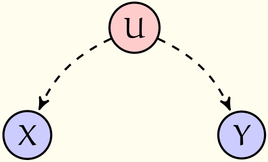
$$\kappa_{\mathsf{ID}(G,X,S)} = O(|V|).$$

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

Ill-conditioned examples

# Primitives of identifiability

# Easy cases: no directed edges

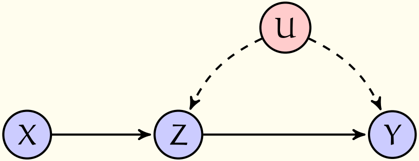



## Identification

$$P(Y \mid \mathrm{do}(X = x)) = \sum_x P(Y, X = x) = P(Y)$$

In general, if $X$ is not an ancestor of $Y$, it can be marginalized

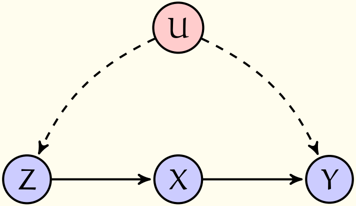
# Easy cases: no hidden edges




## Identification

$$P(YZ \mid do(X = x)) = P(YZ|X)$$

# Easy cases: no hidden edges (slightly more complicated)

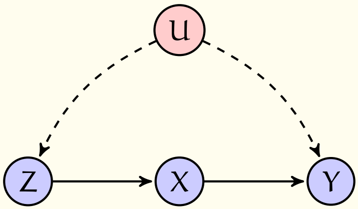



## Identification

$$P(YZ \mid do(X = x)) = P(Z)P(Y|Z, X)$$

- A generalization of this is the crucial tool in the identification algorithms described earlier

# Easy cases: no hidden edges (slightly more complicated)

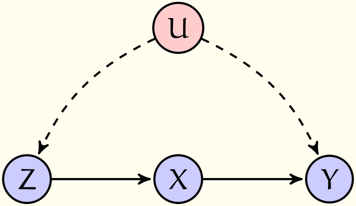



## Identification

$$P(YZ \mid do(X = x)) = P(Z)P(Y|Z, X)$$

- A generalization of this is the crucial tool in the identification algorithms described earlier
- ...and also, in connivance with the innocuous marginalization described above, the main source of ill-conditioning!

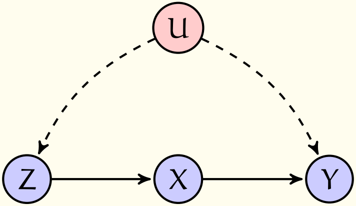
# C-components




## C-components [Tian and Pearl, 2002]

$\{Y, Z\}$ in the above graph is a C-component: a maximal connected component among observed nodes induced by the hidden edges

# C-components




## C-components [Tian and Pearl, 2002]

$\{Y, Z\}$ in the above graph is a C-component: a maximal connected component among observed nodes induced by the hidden edges

## C-components are identifiable [Tian and Pearl, 2002]

If $S \subseteq V$ is a C-component in $G = (V, E, U, D)$ then

$$P(S \mid \mathrm{do}(V - S)) = \prod_{A \in S} P(A \mid V_{\pi(<A)}),$$

where $\pi$ is a topological order on $V$ according to $E$

# C-components and general identifiability

## The hardest case

The “hardest” case for identifiability is $P(S|\mathsf{do}(X))$, where

- $X$ is an ancestor set for $S$ in $G$, and
- $S$ is a C-component in $G - X$

# C-components and general identifiability

## The hardest case

The "hardest" case for identifiability is $P(S|\mathbf{do}(X))$, where

- $X$ is an ancestor set for $S$ in $G$, and
- $S$ is a C-component in $G - X$

Case 1 **$S \cup X$ is a C-component in $G$:** $\mathbf{ID}(G, S, X)$ does not exist

# C-components and general identifiability

## The hardest case

The "hardest" case for identifiability is $P(S|\mathbf{do}(X))$, where

- $X$ is an ancestor set for $S$ in $G$, and
- $S$ is a C-component in $G - X$

Case 1 **$S \cup X$ is a C-component in $G$:** $\mathbf{ID}(G, S, X)$ does not exist

Case 2 **$S$ is a C-component in $G$:** Use C-component identifiability

# C-components and general identifiability

## The hardest case

The "hardest" case for identifiability is $P(S|\mathbf{do}(X))$, where

- $X$ is an ancestor set for $S$ in $G$, and
- $S$ is a C-component in $G - X$

Case 1 **$S \cup X$ is a C-component in $G$:** $\mathbf{ID}(G, S, X)$ does not exist

Case 2 **$S$ is a C-component in $G$:** Use C-component identifiability

Case 3 **$S \cup X'$ is a C-component in $G$, for some $X' \subsetneq X$:**

## Recursion

Call $\mathbf{ID}(S \cup X', X', S)$, but with $P$ replaced by

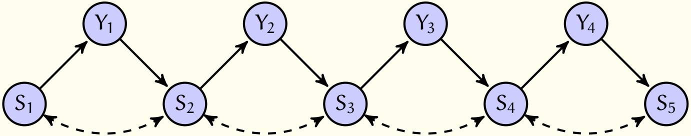
$$P'(S \cup X') := \prod_{A \in S \cup X'} P(A \mid V_{\pi(<A)}),$$

where $\pi$ is a topological order on $V$ according to $E$

Recursion will fail immediately unless some $X'$ is no more an ancestor of $S$!

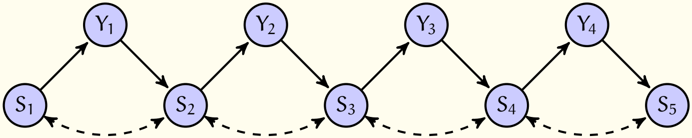
# The ill-conditioned examples

# A warm-up calculation: $\kappa$ is at least $\Omega(n)$




$$P(\cdot) \mapsto P(\boldsymbol{S} \mid \mathrm{do}(\boldsymbol{Y})) = P(.) \mapsto \prod_{i=1}^{n} P(S_i | S_{<i} Y_{<i})$$

# A warm-up calculation: $\kappa$ is at least $\Omega(n)$

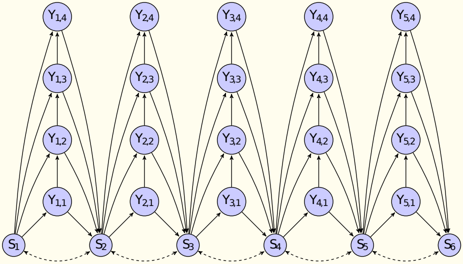



$$P(\cdot) \mapsto P(\boldsymbol{S} \mid \mathrm{do}(\boldsymbol{Y})) = P(.) \mapsto \prod_{i=1}^{n} P(S_i | S_{<i} Y_{<i})$$

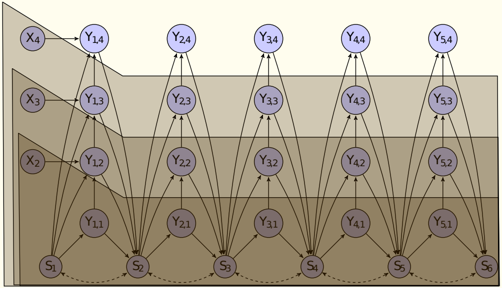
- When $P$ is uniform, the output of the map is the uniform distribution
- However, one can construct a $\tilde{P}$ that is $\epsilon$-close to $P$ and such that each conditional probability above has a positive $\Omega(\epsilon)$ relative error,
  - for a total relative error of $\Omega(n\epsilon)$.

No recursion was used here!

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

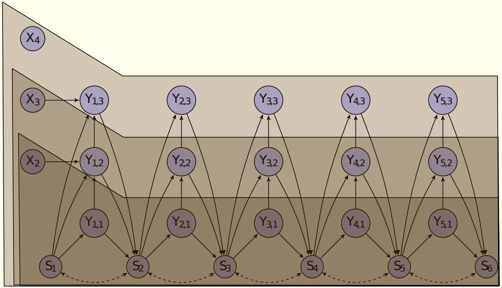


# The final gadget $(m = 6, k = 4)$: $P(\boldsymbol{S} \mid do(\boldsymbol{X}, \boldsymbol{Y}))$




$$P\left(X_{[k]}, S_{[m+1]}, Y_{[m],[k]}\right) := \quad P\left(X_k = x, X_{[k-1]}\right) \cdot \prod_{i=1}^{m} P\left(S_i, Y_{i,[k]} \mid \mathrm{pred}_i\right) \cdot P\left(S_{m+1} \mid \mathrm{pred}_{m+1}\right),$$

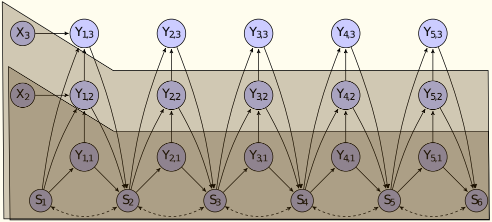
# The final gadget ($m = 6, k = 4$): $P(\boldsymbol{S} \mid \text{do}(\boldsymbol{X}, \boldsymbol{Y}))$




$$P'\left(X_{[k]}, S_{[m+1]}, Y_{[m],[k-1]}\right) := \quad P\left(X_k = x, X_{[k-1]}\right) \cdot \prod_{i=1}^{m} P\left(S_i, Y_{i,[k-1]} \mid \text{pred}_i\right) \cdot P\left(S_{m+1} \mid \text{pred}_{m+1}\right),$$

$\text{Rel}\, P = \epsilon \quad \rightsquigarrow \quad \text{Rel}\, P' \sim m \cdot \epsilon$

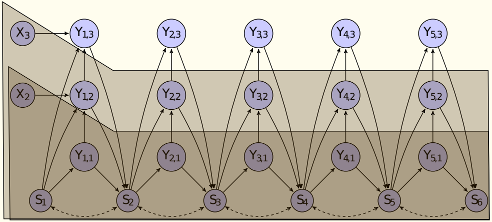
# The final gadget $(m = 6, k = 4)$: $P(\boldsymbol{S} \mid \mathrm{do}(\boldsymbol{X}, \boldsymbol{Y}))$




$$\pi(P)\left(X_{[k-1]}, S_{[m+1]}, Y_{[m],[k-1]}\right) := \sum_x P\left(X_k = x, X_{[k-1]}\right) \cdot \prod_{i=1}^{m} P\left(S_i, Y_{i,[k-1]} \mid \mathrm{pred}_i\right)$$
$$\cdot\, P\left(S_{m+1} \mid \mathrm{pred}_{m+1}\right),$$

$$\mathrm{Rel}\, P = \epsilon \quad \rightsquigarrow \quad \mathrm{Rel}\, P' \sim m \cdot \epsilon \quad \rightsquigarrow \quad \mathrm{Rel}\, \pi(P) \overset{?}{\sim} m \cdot \epsilon$$

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




$$\pi(P)\left(X_{[k-1]}, S_{[m+1]}, Y_{[m],[k-1]}\right) := \sum_{x} P\left(X_k = x, X_{[k-1]}\right) \cdot \prod_{i=1}^{m} P\left(S_i, Y_{i,[k-1]} \mid \text{pred}_i\right)$$
$$\cdot P\left(S_{m+1} \mid \text{pred}_{m+1}\right),$$

$\text{Rel}\, P = \epsilon \quad \rightsquigarrow \quad \text{Rel}\, P' \sim m \cdot \epsilon \quad \rightsquigarrow \quad \text{Rel}\, \pi(P) \overset{?}{\sim} m \cdot \epsilon$

Repeat $k$ times to get **Rel ID** $\sim m^k \cdot \epsilon$?

# Comments

- The marginalization operation can "eat up" the accumulated error if the underlying distribution is uniform

## Our proof

With appropriately chosen non-uniform distributions, the marginalization operation propagates errors

# Comments

- The marginalization operation can "eat up" the accumulated error if the underlying distribution is uniform

**Our proof**

With appropriately chosen non-uniform distributions, the marginalization operation propagates errors

- To get a condition number of $\sim \Omega(\exp(\sqrt{n}))$, choose $m \approxeq k \approxeq \sqrt{n}$

Details of analyzing this correctly are somewhat involved: please see paper

# Conclusion

**Condition number of causality**

**Highly ill-conditioned examples exist**

Very small uncertainties in the model or data can introduce very large errors in causal identification

**But not all instances are ill-conditioned**

A well studied class of examples indeed has small condition number: so numerically stable algorithms can be designed

# Conclusion

**Condition number of causality**

**Highly ill-conditioned examples exist**

Very small uncertainties in the model or data can introduce very large errors in causal identification

**But not all instances are ill-conditioned**

A well studied class of examples indeed has small condition number: so numerically stable algorithms can be designed

## Some future directions

- Algorithmically classify condition numbers for any given model and intervention
  - e.g., for comparing different models of the system being studied

# Conclusion

**Condition number of causality**

**Highly ill-conditioned examples exist**

Very small uncertainties in the model or data can introduce very large errors in causal identification

**But not all instances are ill-conditioned**

A well studied class of examples indeed has small condition number: so numerically stable algorithms can be designed

**Some future directions**

- Algorithmically classify condition numbers for any given model and intervention
  - e.g., for comparing different models of the system being studied
- Find ways to get around an ill-conditioned model by using more data
  - e.g. some measured intervention distributions? error correction?

# Conclusion

**Condition number of causality**

**Highly ill-conditioned examples exist**

Very small uncertainties in the model or data can introduce very large errors in causal identification

**But not all instances are ill-conditioned**

A well studied class of examples indeed has small condition number: so numerically stable algorithms can be designed

## Some future directions

- Algorithmically classify condition numbers for any given model and intervention
  - e.g., for comparing different models of the system being studied
- Find ways to get around an ill-conditioned model by using more data
  - e.g. some measured intervention distributions? error correction?
- Condition number for other causal inference problems, e.g. SEMs

# Conclusion

**Condition number of causality**

**Highly ill-conditioned examples exist**

Very small uncertainties in the model or data can introduce very large errors in causal identification

**But not all instances are ill-conditioned**

A well studied class of examples indeed has small condition number: so numerically stable algorithms can be designed

**Some future directions**

- Algorithmically classify condition numbers for any given model and intervention
  - e.g., for comparing different models of the system being studied
- Find ways to get around an ill-conditioned model by using more data
  - e.g. some measured intervention distributions? error correction?
- Condition number for other causal inference problems, e.g. SEMs

**Thank you!**

# Conclusion

**Condition number of causality**

**Highly ill-conditioned examples exist**

Very small uncertainties in the model or data can introduce very large errors in causal identification

**But not all instances are ill-conditioned**

A well studied class of examples indeed has small condition number: so numerically stable algorithms can be designed

**Some future directions**

- Algorithmically classify condition numbers for any given model and intervention
  - e.g., for comparing different models of the system being studied
- Find ways to get around an ill-conditioned model by using more data
  - e.g. some measured intervention distributions? error correction?
- Condition number for other causal inference problems, e.g. SEMs

**Thank you!**

# Conclusion

**Condition number of causality**

**Highly ill-conditioned examples exist**

Very small uncertainties in the model or data can introduce very large errors in causal identification

**But not all instances are ill-conditioned**

A well studied class of examples indeed has small condition number: so numerically stable algorithms can be designed

**Some future directions**

- Algorithmically classify condition numbers for any given model and intervention
  - e.g., for comparing different models of the system being studied
- Find ways to get around an ill-conditioned model by using more data
  - e.g. some measured intervention distributions? error correction?
- Condition number for other causal inference problems, e.g. SEMs

**Thank you!**

# 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

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

**$\varepsilon$ 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.