Noisy-OR Models with Latent Confounding

Antti Hyttinen, Frederick Eberhardt, Patrik O. Hoyer
University of Helsinki & HIIT & Washington University in St. Louis

Summary

We examine the identifiability of causal models with latent confounding, given a set of experiments in which subsets of the observed variables are subject to interventions.

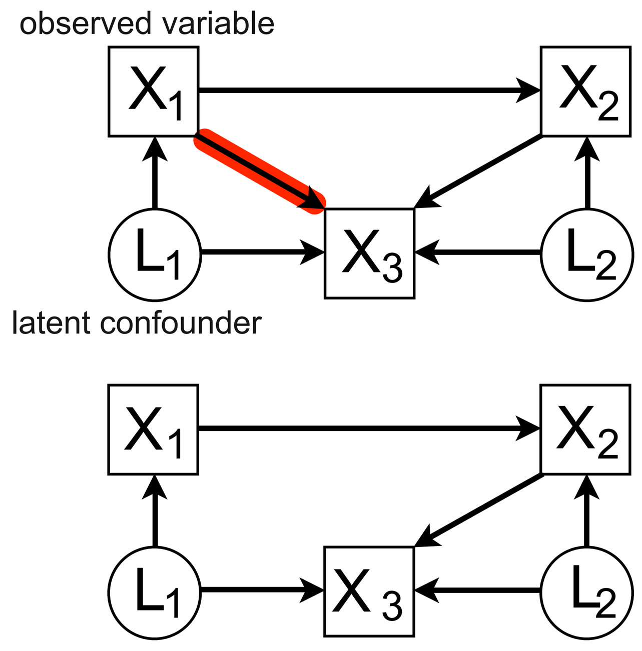
In general identifiability is impossible on the basis of experiments where only few variables are subject to intervention per experiment, which is often the case.

Identifiability is possible for a class of causal models whose conditional probability distributions are restricted to a 'noisy-OR' parameterization.

Identifiability is preserved under an extension of the noisy-OR CPD that allows for negative influences.

Several learning algorithms are introduced and tested for accuracy, scalability and robustness.

1. On the Identifiability of Causal Models with Latent Confounding



Passive observational data or experiments intervening on only a few variables at a time are generally insufficient to identify the **parameters** and the **structure** of a causal model with latent confounding.

For example, the two graphs on the left imply the exact same independences in single intervention experiments and when passively observed.

Furthermore, there exist parameterizations for the two graphs that produce the exact same distributions in those situations as well.

Thus, the presence of the **red** direct link cannot be determined unless both X_1 and X_2 are subject to an intervention in the same experiment.

2. Noisy-OR Model with Latent Confounding

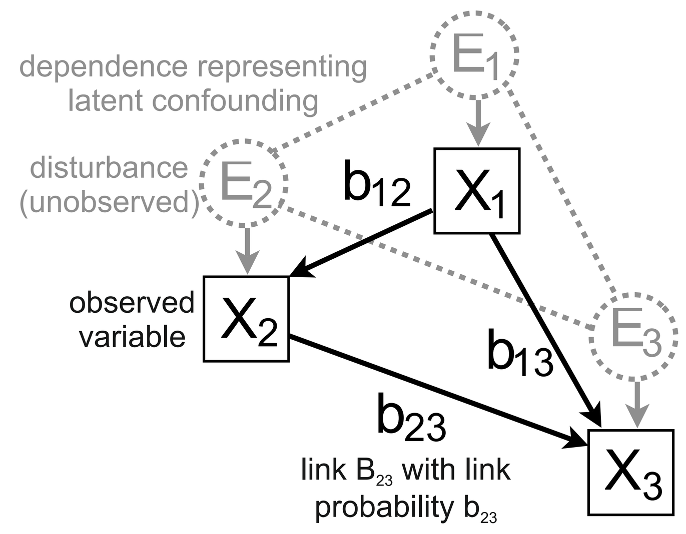
Structural Equation Model

$$X_1 := E_1$$
$$X_2 := (B_{12} \wedge X_1) \vee E_2$$
$$X_3 := (B_{13} \wedge X_1) \vee (B_{23} \wedge X_2) \vee E_3$$

Binary random variables X_1, X_2 and X_3 are observed. Links B_{12}, B_{23} and B_{13} and disturbances E_1, E_2 and E_3 are all unobserved binary random variables, introducing noise to the simple OR expressions.

Conditional Probability Distributions Links are independently distributed with model parameters $b_{12} = P(B_{12} = 1)$, b_{13} and b_{23}.

$$P(X_1 = 0|E_1) = (1 - E_1)$$
$$P(X_2 = 0|E_2, X_1) = (1 - E_2)(1 - b_{12})^{X_1}$$
$$P(X_3 = 0|E_3, X_1, X_2) = (1 - E_3)(1 - b_{13})^{X_1}(1 - b_{23})^{X_2}$$



Latent Confounding Latent confounding is represented by an arbitrary distribution $P(E_1^3)$ (total of 2^3 parameters). Any latent confounding (restricted by the noisy-OR CPD) can be presented through E_1, E_2 and E_3.

Joint Distribution

$$P(X_1^3) = \sum_{E_1^3} P(X_1|E_1)P(X_2|X_1, E_2)P(X_3|X_1, X_2, E_3)P(E_1^3)$$

Data Generation Draw a sample of disturbances E_1^3 from $P(E_1^3)$, links B_{12}, B_{13}, B_{23} from their independent distributions, and determine X_1, X_2 and X_3 from the SEM equations.

Context Specific Independence Property Noisy-OR CPDs have the following property.

$$(X_1 \perp\!\!\!\perp E_2 \,||\, X_1) \Rightarrow (X_1 \perp\!\!\!\perp E_2 \mid X_2 = 0 \,||\, X_1)$$

If parents X_1 and E_2 of variable X_2 are independent in some context (here when intervening on X_1), then additionally conditioning on their common child $X_2 = 0$ does not destroy this independence. This is evident from the SEM equations, if $X_2 = 0$, then $E_2 = 0$ and $(B_{12} \wedge X_1) = 0$, thus the value of E_2 does not provide any additional information about the value of X_1.

3. Identifiability

The parameters of any three variable model can be identified from single intervention experiments and passive observational data.

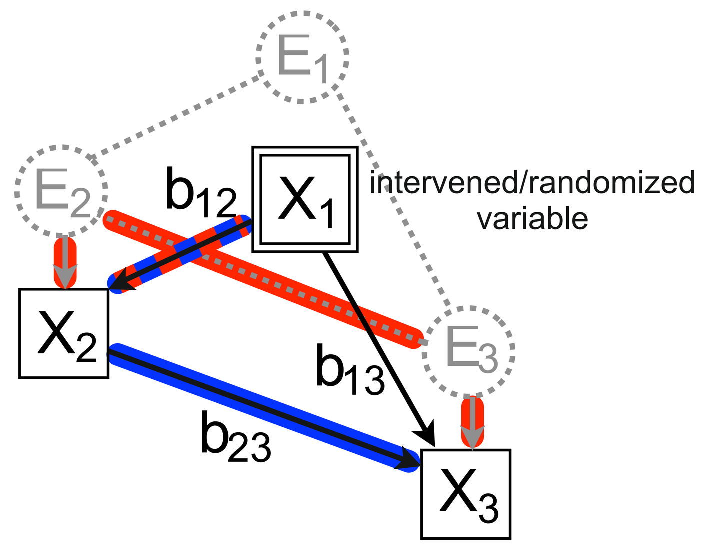
Step 1 Find a causal order from the ancestral relationships directly observed in the experiments and rename variables such that the causal order is X_1, X_2, X_3.

Step 2 Estimate link probability b_{12} by Cheng's causal power formula, using the intervention on X_1 to make E_2 independent of X_1.

$$b_{12} = \frac{P(X_2 = 1||X_1 = 1) - \overbrace{P(X_2 = 1||X_1 = 0)}^{X_2 = 1 \text{ caused by its other causes}}}{\underbrace{1 - P(X_2 = 1||X_1 = 0)}_{\text{renormalization}}}$$

Similarly, estimate b_{23} by intervening on X_2.

$$b_{23} = \frac{P(X_3 = 1||X_2 = 1) - P(X_3 = 1||X_2 = 0)}{1 - P(X_3 = 1||X_2 = 0)}$$



Step 3 Estimate the link probability b_{13} by additionally conditioning on $X_2 = 0$ s.t. the **blue** indirect path is intercepted.

$$b_{13} = \frac{P(X_3 = 1|X_2 = 0||X_1 = 1) - P(X_3 = 1|X_2 = 0||X_1 = 0)}{1 - P(X_3 = 1|X_2 = 0||X_1 = 0)}$$

The context specific independence property guarantees that the **red** path remains intercepted.

Step 4 Estimate the noise distribution from the passive observational data by solving a matrix equation:

$$\overbrace{\begin{bmatrix} \ddots & \vdots & \vdots & \vdots \\ \cdots & (1-b_{12}) & 0 & 0 \\ \cdots & 0 & (1-b_{13})(1-b_{23}) & 0 \\ \cdots & b_{12} & b_{13} + b_{23} - b_{13}b_{23} & 1 \end{bmatrix}}^{P(X_1^3|E_1^3)} \overbrace{\begin{bmatrix} \vdots \\ P(E_1^3 = 101) \\ P(E_1^3 = 110) \\ P(E_1^3 = 111) \end{bmatrix}}^{P(E_1^3)} = \overbrace{\begin{bmatrix} \vdots \\ P(X_1^3 = 101) \\ P(X_1^3 = 110) \\ P(X_1^3 = 111) \end{bmatrix}}^{P(X_1^3)}$$

The matrix on the left is lower triangular with a nonzero diagonal, and thus invertible.

All parameters of a noisy-OR model with latent confounding are identified from the combination of a **passive observational data set** and a set of experiments where **for each ordered variable pair (X_i, X_j) there is an experiment where X_i is randomized and X_j is observed**. This condition is often also necessary.

4. Learning Algorithms

Efficient Conditioning Conditioning reduces the effective sample size for estimating the link probabilities. However, if it happens in step 2 (above) that $b_{12} = 0$ or $b_{23} = 0$, then the **blue** path does not exist and conditioning on X_2 is unnecessary when estimating b_{13}. The correct conditioning sets for each link can always be determined based on links already estimated. In addition, the experimental data can also be taken into account when estimating $P(E_1^3)$.

EM-algorithm For up to eight variables, the model can also be learned using a version of the EM-algorithm.

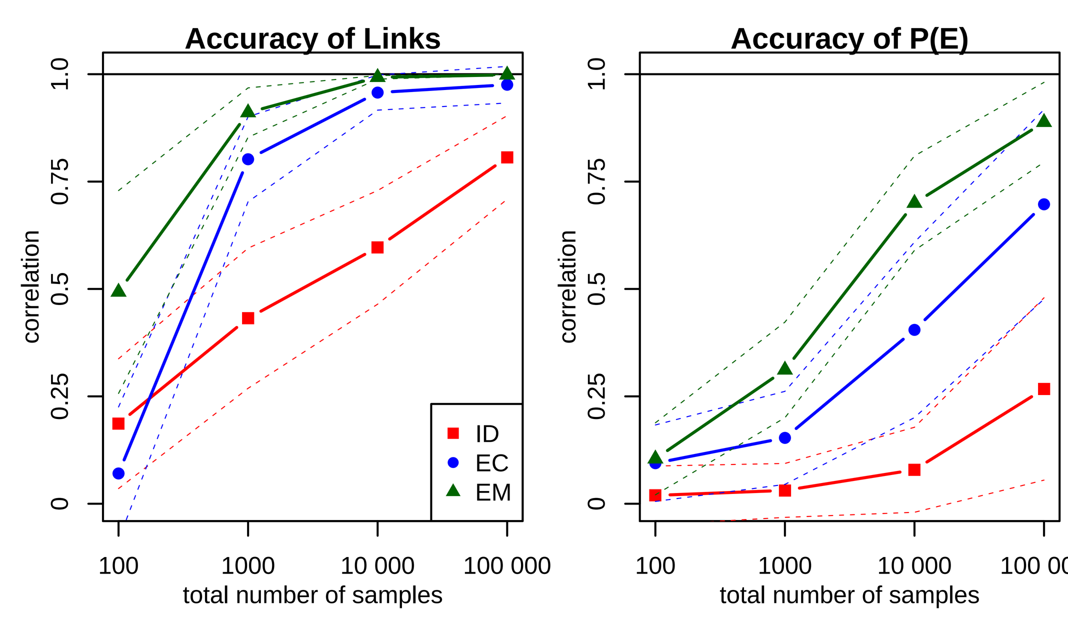
5. Extension to Negative Influences

In noisy-OR models, the parents X_1 and X_2 being ON has a positive effect on their child X_3 being ON. However, the noisy-OR parameterization can be extended to also allow for negative influences:

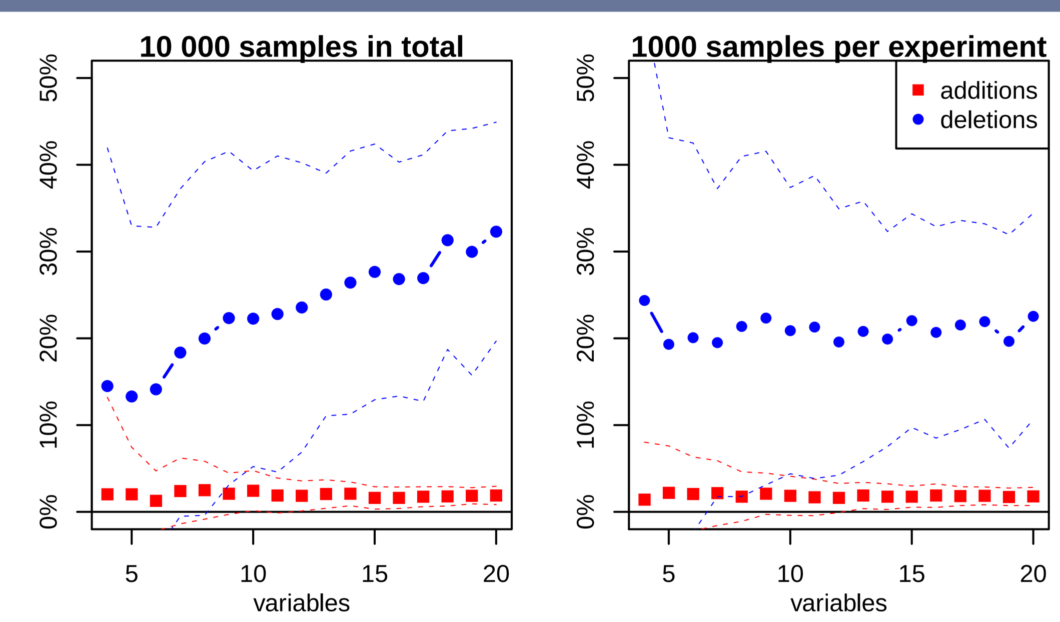
$$X_3 := E_3 \vee (B_{13} \wedge \widetilde{X}_1) \vee (B_{23} \wedge \widetilde{X}_2),$$

where for positive/generative causes $\widetilde{X}_i = X_i$ and for negative causes $\widetilde{X}_i = \neg X_i$. Now $X_1 = 0$ can cause $X_3 = 1$. The context specific independence property and the identifiability of the model are preserved.

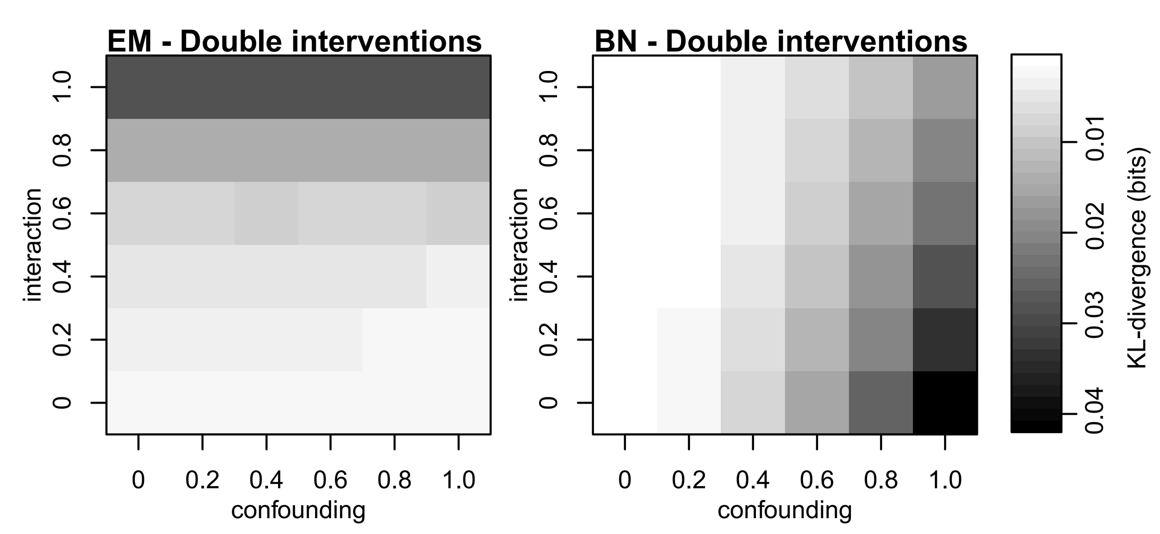
6. Simulations



Accuracy Accuracy of the learning algorithms with increasing sample sizes. EM is most accurate, EC beats the algorithm based on the identifiability proof (ID).



Scalability Structural errors when using the EC-algorithm on models with different sizes. Some statistically insignificant links are deleted.



Robustness Models were learned from single intervention and passive observational data, generated by a 'noisy-interactive-OR' model while the amount of latent confounding and interaction of the parents was varied. The shade of each square represents the average predictive accuracy in double intervention experiments. Lighter shades indicate better results. Standard Bayesian Network without hidden variables (BN) predicts accurately when there is little confounding, noisy-OR (EM) predicts accurately when there is only little interaction.