
Identifying Causal Effects via Context-specific Independence Relations (*Supplementary Material*)

Santtu Tikka

Department of Mathematics and Statistics
University of Jyväskylä, Finland
santtu.tikka@jyu.fi

Antti Hyttinen

HIIT, Department of Computer Science
University of Helsinki, Finland
antti.hyttinen@helsinki.fi

Juha Karvanen

Department of Mathematics and Statistics
University of Jyväskylä, Finland
juha.t.karvanen@jyu.fi

A Proofs

Proof of Theorem 1 without positivity. For pedagogical purposes, we begin with a proof without the positivity assumption. The subsequent proof, where positivity is assumed, is obtained via a slight modification.

Define the corresponding LDAG to the 3-SAT instance as follows (see Fig. 1) [2]. Let nodes $\mathbf{Z} = \{Z_1, \dots, Z_l\}$ correspond to variables in the 3-SAT instance and U_1, \dots, U_k represent the clauses of the 3-SAT instance. Let the parents of node U_i ($i \geq 1$) be U_{i-1} , and the Z s appearing in i th clause. The label L_i on the edge $U_{i-1} \rightarrow U_i$ consists of assignments to other parents of U_i that are in \mathbf{Z} : let it be exactly the set of assignments that *do not* satisfy the i th clause. Let the parent of X be U_0 and the parents of Y be X, U_k . We will show that $P(Y | \text{do}(X))$ is identifiable from $P(X, Y)$ if and only if the 3-SAT instance is unsatisfiable.

Identifiability when UNSAT Suppose the 3-SAT instance is unsatisfiable. Then for any assignment \mathbf{z} to \mathbf{Z} , edge $U_{i-1} \rightarrow U_i$ is absent in the context \mathbf{z} , and according to CSI-separation: $I_X \perp\!\!\!\perp Y | X, \mathbf{Z} = \mathbf{z}$. It follows that $I_X \perp\!\!\!\perp Y | X, \mathbf{Z}$. By d-separation $I_X \perp\!\!\!\perp \mathbf{Z}$, by contraction $I_X \perp\!\!\!\perp Y, \mathbf{Z} | X$, by decomposition $I_X \perp\!\!\!\perp Y | X$. Thus the causal effect is identifiable:

$$P(Y | \text{do}(X)) = P(Y | X, I_X = 1) = P(Y | X, I_X = 0) = P(Y | X).$$

Non-identifiability when SAT Suppose the 3-SAT instance is satisfiable. We define 2 parameterizations for the LDAG, M^1 and M^2 that agree with $P(X, Y)$ but disagree on $P(Y | \text{do}(X))$. First let us describe the common part of both models. Let \mathbf{Z} be distributed uniformly mutually independently. Let $U_i = U_{i-1}$ if Z_a, Z_b, Z_c satisfy the clause i and 0 otherwise. Let U_0 be uniform, and let X be equal to U_0 .

Same observed distribution We can calculate the observed joint $P(X, Y)$ as follows:

$$\begin{aligned}
P(X, Y) &= \sum_{U_k, U_0, \mathbf{Z}} P(X, Y, U_k, U_0, \mathbf{Z}) \\
&= \sum_{U_k, U_0, \mathbf{Z}} P(Y | X, U_k) P(X | U_0) P(U_k | U_0, \mathbf{Z}) P(U_0) P(\mathbf{Z}) \quad || P(\mathbf{Z}) = 2^{-l}, P(U_0) = 2^{-1} \\
&= 2^{-l-1} \sum_{U_k, U_0, \mathbf{Z}} P(Y | X, U_k) P(X | U_0) P(U_k | U_0, \mathbf{Z}) \quad || P(X | U_0 = 1 - X) = 0 \\
&= 2^{-l-1} \sum_{U_k, \mathbf{Z}} P(Y | X, U_k) P(U_k | U_0 = X, \mathbf{Z}) \\
&= 2^{-l-1} \sum_{U_k, \mathbf{z} \in \text{SAT}} P(Y | X, U_k) P(U_k | U_0 = X, \mathbf{z}) \quad || U_k = U_0 = X \text{ when SAT} \\
&\quad + 2^{-l-1} \sum_{U_k, \mathbf{z} \in \text{UNSAT}} P(Y | X, U_k) P(U_k | U_0 = X, \mathbf{z}) \quad || U_k = 0 \text{ when UNSAT} \\
&= 2^{-l-1} \left[\sum_{\mathbf{z} \in \text{SAT}} P(Y | X, U_k = X) + \sum_{\mathbf{z} \in \text{UNSAT}} P(Y | X, U_k = 0) \right].
\end{aligned}$$

In order for this to be equal for M^1 and M^2 it suffices that

$$\begin{aligned}
P^1(Y | X, U_k = 0) &= P^2(Y | X, U_k = 0), \\
P^1(Y | X = 1, U_k = 1) &= P^2(Y | X = 1, U_k = 1).
\end{aligned}$$

In addition, the probabilities should be non-zero for $P(X, Y)$ to be positive for all 3-SAT instances. Then

$$P^1(X, Y) = P^2(X, Y) > 0.$$

Different causal effects The parameter $P(Y | X = 0, U_k = 1)$ does not affect the passive observational distribution in any way, so let $P^1(Y | X = 0, U_k = 1) \neq P^2(Y | X = 0, U_k = 1)$. This difference alters the causal effect:

$$\begin{aligned}
P(Y | \text{do}(X = 0)) &= 2^{-l-1} \sum_{U_k, U_0, \mathbf{Z}} P(Y | X = 0, U_k) P(U_k | U_0, \mathbf{Z}) \\
&= 2^{-l-1} \sum_{U_k, U_0, \mathbf{z} \in \text{SAT}} P(Y | X = 0, U_k) P(U_k | U_0, \mathbf{z}) \quad || U_k = U_0 \text{ when SAT} \\
&\quad + 2^{-l-1} \sum_{U_k, U_0, \mathbf{z} \in \text{UNSAT}} P(Y | X = 0, U_k) P(U_k | U_0, \mathbf{z}) \quad || U_k = 0 \text{ when UNSAT} \\
&= 2^{-l-1} \sum_{\mathbf{z} \in \text{SAT}} P(Y | X = 0, U_k = U_0) + 2^{-l-1} \sum_{\mathbf{z} \in \text{UNSAT}} P(Y | X = 0, U_k = 0).
\end{aligned}$$

and thus

$$\begin{aligned}
&P^1(Y | \text{do}(X = 0)) - P^2(Y | \text{do}(X = 0)) = \\
&2^{-l-1} \sum_{\mathbf{Z} \in \text{SAT}} [P^1(Y | X = 0, U_k = 1) - P^2(Y | X = 0, U_k = 1)] \neq 0.
\end{aligned}$$

This shows that causal effect is non-identifiable if the 3-SAT instance is satisfiable. Next, we provide a similar construction for the case where positivity is assumed. \square

Proof of Theorem 1. Define the corresponding LDAG to the 3-SAT instance as follows (see Fig. 1) [2]. Let nodes $\mathbf{Z} = \{Z_1, \dots, Z_l\}$ correspond to variables in the 3-SAT instance and U_1, \dots, U_k represent the clauses of the 3-SAT instance. Let the parents of node U_i ($i \geq 1$) be U_{i-1} , and the Z s appearing in i th clause. The label \mathbf{L}_i on the edge $U_{i-1} \rightarrow U_i$ consists of assignments to other parents of U_i that are in \mathbf{Z} : let it be exactly the set of assignments that *do not* satisfy the i th clause. Let the parent of X be U_0 and the parents of Y be X, U_k . We will show that $P(Y | \text{do}(X))$ is identifiable from $P(X, Y)$ if and only if the 3-SAT instance is unsatisfiable.

Identifiability when UNSAT Suppose the 3-SAT instance is unsatisfiable. Then for any assignment \mathbf{z} to \mathbf{Z} , edge $U_{i-1} \rightarrow U_i$ is absent in the context \mathbf{z} , and according to CSI-separation: $I_X \perp\!\!\!\perp Y | X, \mathbf{Z} = \mathbf{z}$. It follows that $I_X \perp\!\!\!\perp Y | X, \mathbf{Z}$. By d-separation $I_X \perp\!\!\!\perp \mathbf{Z}$, by contraction $I_X \perp\!\!\!\perp Y, \mathbf{Z} | X$, by decomposition $I_X \perp\!\!\!\perp Y | X$. Thus the causal effect is identifiable:

$$P(Y | \text{do}(X)) = P(Y | X, I_X = 1) = P(Y | X, I_X = 0) = P(Y | X).$$

Non-identifiability when SAT Suppose the 3-SAT instance is satisfiable. We define 2 parameterizations for the LDAG, M^1 and M^2 that agree on $P(X, Y)$ but disagree on $P(Y | \text{do}(X))$.

First let us describe the common part of both models. Let \mathbf{Z} be distributed uniformly mutually independently. If Z_a, Z_b, Z_c do not satisfy i :th clause we let U_i to be distributed uniformly. If they do let $P(U_i = 1 | U_{i-1} = 1) = p$ and $P(U_i = 1 | U_{i-1} = 0) = 1 - p$ such that $p \neq 1/2$. For M^1 let $P^1(U_0 = 1) = 1/2$ and let $P^1(X = 1 | U_0 = 1) = q$ and $P^1(X = 1 | U_0 = 0) = 1 - q$ such that $q \neq 1/2$. For M^2 let $P^2(U_0 = 0) = q$ and $P^2(X | U_0) = 1/2$ irrespective of the value of U_0 .

First we observe that $P^1(X) = P^2(X)$. For M^1 we have

$$\begin{aligned} P^1(X = 0) &= P^1(X = 0 | U_0 = 0)P^1(U_0 = 0) + P^1(X = 0 | U_0 = 1)P^1(U_0 = 1) \\ &= q \frac{1}{2} + (1 - q) \frac{1}{2} = \frac{1}{2}, \end{aligned}$$

and for M^2 we have

$$\begin{aligned} P^2(X = 0) &= P^2(X = 0 | U_0 = 0)P^2(U_0 = 0) + P^2(X = 0 | U_0 = 1)P^2(U_0 = 1) \\ &= \frac{1}{2}q + \frac{1}{2}(1 - q) = \frac{1}{2}. \end{aligned}$$

If some clause j is unsatisfied, the distributions all up to k are uniform regardless whether the remaining clauses are satisfied or not:

$$\begin{aligned} P(U_j | U_0 = 1, \mathbf{z}) &= \sum_{U_{j-1}} P(U_j | U_{j-1}, \mathbf{z})P(U_{j-1} | U_0 = 1, \mathbf{z}) = \frac{1}{2} \sum_{U_{j-1}} P(U_{j-1} | U_0, \mathbf{z}) = \frac{1}{2}. \\ P(U_{j+1} | U_0 = 1, \mathbf{z}) &= \sum_{U_j} P(U_{j+1} | U_j, \mathbf{z})P(U_j | U_0 = 1, \mathbf{z}) = (1 - p) \frac{1}{2} + p \frac{1}{2} = \frac{1}{2}. \end{aligned}$$

If all clauses are satisfied by \mathbf{z} we can define the following random variable:

$$F = \sum_{i=1}^k I(U_i \neq U_{i-1}, \mathbf{z} \in \text{SAT}),$$

where I is the indicator function. Based on the parametrization of the U_i 's we know that $P(F | \mathbf{z}) = \text{Bin}(k, 1 - p)$. Furthermore, $U_k = U_0$ if $F = 2t$ for any $t \in \{0, \dots, \lfloor k/2 \rfloor\}$. Similarly, $U_k \neq U_0$ if $F = 2t + 1$ for any $t \in \{0, \dots, \lfloor \frac{k-1}{2} \rfloor\}$. From this we obtain the following distributions

$$\begin{aligned} P(U_k = 0 | U_0 = 0, \mathbf{z}) &= P(U_k = 1 | U_0 = 1, \mathbf{z}) = \sum_{t=0}^{\lfloor k/2 \rfloor} \binom{k}{2t} (1 - p)^{2t} p^{k-2t} := a. \\ P(U_k = 1 | U_0 = 0, \mathbf{z}) &= P(U_k = 0 | U_0 = 1, \mathbf{z}) = \sum_{t=0}^{\lfloor \frac{k-1}{2} \rfloor} \binom{k}{2t+1} (1 - p)^{2t+1} p^{k-2t-1} = 1 - a. \end{aligned}$$

Same observed distribution We can calculate the observed joint $P(X, Y)$ for both models as follows:

$$\begin{aligned}
P(X, Y) &= \sum_{U_k, U_0, \mathbf{Z}} P(X, Y, U_k, U_0, \mathbf{Z}) \\
&= \sum_{U_k, U_0, \mathbf{Z}} P(Y | X, U_k) P(X | U_0) P(U_k | U_0, \mathbf{Z}) P(U_0) P(\mathbf{Z}) \quad || P(\mathbf{Z}) = 2^{-l} \\
&= 2^{-l} \sum_{U_k, U_0, \mathbf{Z}} P(Y | X, U_k) P(X | U_0) P(U_k | U_0, \mathbf{Z}) P(U_0) \\
&= 2^{-l} \sum_{U_k, U_0, \mathbf{z} \in \text{SAT}} P(Y | X, U_k) P(X | U_0) P(U_k | U_0, \mathbf{z}) P(U_0) \\
&\quad + 2^{-l} \sum_{U_k, U_0, \mathbf{z} \in \text{UNSAT}} P(Y | X, U_k) P(X | U_0) P(U_k | U_0, \mathbf{z}) P(U_0).
\end{aligned}$$

For M^1 and M^2 we have

$$\begin{aligned}
P(X = 0, Y) &= 2^{-l} \sum_{\mathbf{z} \in \text{SAT}} \left(P(Y | X = 0, U_k = 0) qa \frac{1}{2} + P(Y | X = 0, U_k = 0) (1 - q) (1 - a) \frac{1}{2} \right. \\
&\quad \left. + P(Y | X = 0, U_k = 1) q (1 - a) \frac{1}{2} + P(Y | X = 0, U_k = 1) (1 - q) a \frac{1}{2} \right) \\
&\quad + 2^{-l} \sum_{\mathbf{z} \in \text{UNSAT}} \left(P(Y | X = 0, U_k = 0) q \frac{1}{4} + P(Y | X = 0, U_k = 0) (1 - q) \frac{1}{4} \right. \\
&\quad \left. + P(Y | X = 0, U_k = 1) q \frac{1}{4} + P(Y | X = 0, U_k = 1) (1 - q) \frac{1}{4} \right) \\
&= 2^{-l-1} \sum_{\mathbf{z} \in \text{SAT}} (P(Y | X = 0, U_k = 0) qa + P(Y | X = 0, U_k = 0) (1 - q) (1 - a) \\
&\quad + P(Y | X = 0, U_k = 1) q (1 - a) + P(Y | X = 0, U_k = 1) (1 - q) a) \\
&\quad + 2^{-l-2} \sum_{\mathbf{z} \in \text{UNSAT}} (P(Y | X = 0, U_k = 0) + P(Y | X = 0, U_k = 1)).
\end{aligned}$$

When $X = 1$, the expression for $P(X = 1, Y)$ is different for the models. For M^1 it is

$$\begin{aligned}
P^1(X = 1, Y) &= 2^{-l} \sum_{\mathbf{z} \in \text{SAT}} \left(P^1(Y | X = 1, U_k = 0) (1 - q) a \frac{1}{2} + P^1(Y | X = 1, U_k = 0) q (1 - a) \frac{1}{2} \right. \\
&\quad \left. + P^1(Y | X = 1, U_k = 1) (1 - q) (1 - a) \frac{1}{2} + P^1(Y | X = 1, U_k = 1) qa \frac{1}{2} \right) \\
&\quad + 2^{-l} \sum_{\mathbf{z} \in \text{UNSAT}} \left(P^1(Y | X = 1, U_k = 0) (1 - q) \frac{1}{4} + P^1(Y | X = 1, U_k = 0) q \frac{1}{4} \right. \\
&\quad \left. + P^1(Y | X = 1, U_k = 1) (1 - q) \frac{1}{4} + P^1(Y | X = 1, U_k = 1) q \frac{1}{4} \right) \\
&= 2^{-l-1} \sum_{\mathbf{z} \in \text{SAT}} (P^1(Y | X = 1, U_k = 0) (1 - q) a + P^1(Y | X = 1, U_k = 0) q (1 - a) \\
&\quad + P^1(Y | X = 1, U_k = 1) (1 - q) (1 - a) + P^1(Y | X = 1, U_k = 1) qa) \\
&\quad + 2^{-l-2} \sum_{\mathbf{z} \in \text{UNSAT}} (P^1(Y | X = 1, U_k = 0) + P^1(Y | X = 1, U_k = 1)),
\end{aligned}$$

and for M^2 it is

$$\begin{aligned}
P^2(X = 1, Y) &= 2^{-l} \sum_{\mathbf{z} \in \text{SAT}} \left(P^2(Y | X = 1, U_k = 0)qa\frac{1}{2} + P^2(Y | X = 1, U_k = 0)(1-q)(1-a)\frac{1}{2} \right. \\
&\quad \left. + P^2(Y | X = 1, U_k = 1)q(1-a)\frac{1}{2} + P^2(Y | X = 1, U_k = 1)(1-q)a\frac{1}{2} \right) \\
&\quad + 2^{-l} \sum_{\mathbf{z} \in \text{UNSAT}} \left(P^2(Y | X = 1, U_k = 0)q\frac{1}{4} + P^2(Y | X = 1, U_k = 0)(1-q)\frac{1}{4} \right. \\
&\quad \left. + P^2(Y | X = 1, U_k = 1)q\frac{1}{4} + P^2(Y | X = 1, U_k = 1)(1-q)\frac{1}{4} \right) \\
&= 2^{-l-1} \sum_{\mathbf{z} \in \text{SAT}} (P^2(Y | X = 1, U_k = 0)qa + P^2(Y | X = 1, U_k = 0)(1-q)(1-a) \\
&\quad + P^2(Y | X = 1, U_k = 1)q(1-a) + P^2(Y | X = 1, U_k = 1)(1-q)a) \\
&\quad + 2^{-l-2} \sum_{\mathbf{z} \in \text{UNSAT}} (P^2(Y | X = 1, U_k = 0) + P^2(Y | X = 1, U_k = 1)).
\end{aligned}$$

In order to ensure that $P^1(X = 1, Y) = P^2(X = 1, Y)$ we let $P^1(Y | X = 1, U_k) = P^2(Y | X = 1, U_k) = 1/2$. For $P(Y | X = 0, U_k)$ we let $P^1(Y | X = 0, U_k) = P^2(Y | X = 0, U_k) > 0$. The considered parametrization ensures that $P(X, Y, \mathbf{Z}, U_0, \dots, U_k)$ is positive for all 3-SAT instances, and

$$P^1(X, Y, \mathbf{Z}, U_0, \dots, U_k) = P^2(X, Y, \mathbf{Z}, U_0, \dots, U_k) > 0.$$

Different causal effects The different parameterizations for U_0 and the conditional distribution $P(X | U_0)$ in M^1 and M^2 induce a difference in the causal effects. The causal effect can be computed as

$$\begin{aligned}
P(Y | \text{do}(X = 0)) &= 2^{-l} \sum_{U_k, U_0, \mathbf{z} \in \text{SAT}} P(Y | X, U_k)P(U_k | U_0, \mathbf{z})P(U_0) \\
&\quad + 2^{-l} \sum_{U_k, U_0, \mathbf{z} \in \text{UNSAT}} P(Y | X, U_k)P(U_k | U_0, \mathbf{z})P(U_0).
\end{aligned}$$

For model M^1 this is

$$\begin{aligned}
P^1(Y | \text{do}(X = 0)) &= 2^{-l} \sum_{\mathbf{z} \in \text{SAT}} \left(P^1(Y | X = 0, U_k = 0)a\frac{1}{2} + P^1(Y | X = 0, U_k = 0)(1-a)\frac{1}{2} \right. \\
&\quad \left. + P^1(Y | X = 0, U_k = 1)(1-a)\frac{1}{2} + P^1(Y | X = 0, U_k = 1)a\frac{1}{2} \right) \\
&\quad + 2^{-l} \sum_{\mathbf{z} \in \text{UNSAT}} \left(P^1(Y | X = 0, U_k = 0)\frac{1}{4} + P^1(Y | X = 0, U_k = 0)\frac{1}{4} \right. \\
&\quad \left. + P^1(Y | X = 0, U_k = 1)\frac{1}{4} + P^1(Y | X = 0, U_k = 1)\frac{1}{4} \right) \\
&= 2^{-l-1} \sum_{\mathbf{z} \in \text{SAT}} (P^1(Y | X = 0, U_k = 0) + P^1(Y | X = 0, U_k = 1)) \\
&\quad + 2^{-l-1} \sum_{\mathbf{z} \in \text{UNSAT}} (P^1(Y | X = 0, U_k = 0) + P^1(Y | X = 0, U_k = 1)).
\end{aligned}$$

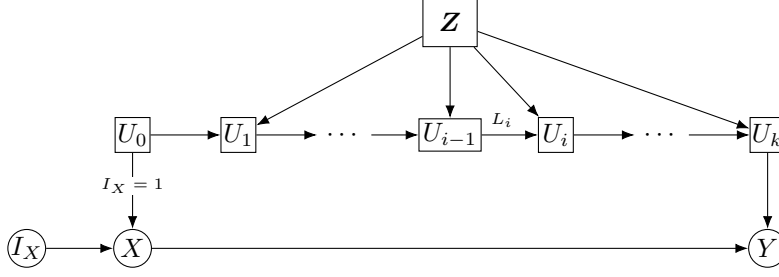


Figure 1: LDAG for the proof of Theorem 1.

For model M^2 the causal effect is

$$\begin{aligned}
& P^2(Y \mid \text{do}(X = 0)) \\
&= 2^{-l} \sum_{z \in \text{SAT}} (P^2(Y \mid X = 0, U_k = 0)aq + P^2(Y \mid X = 0, U_k = 0)(1-a)(1-q) \\
&\quad + P^2(Y \mid X = 0, U_k = 1)(1-a)q + P^2(Y \mid X = 0, U_k = 1)a(1-q)) \\
&+ 2^{-l} \sum_{z \in \text{UNSAT}} \left(P^2(Y \mid X = 0, U_k = 0)\frac{1}{2}q + P^2(Y \mid X = 0, U_k = 0)\frac{1}{2}(1-q) \right. \\
&\quad \left. + P^2(Y \mid X = 0, U_k = 1)\frac{1}{2}q + P^2(Y \mid X = 0, U_k = 1)\frac{1}{2}(1-q) \right) \\
&= 2^{-l} \sum_{z \in \text{SAT}} (P^2(Y \mid X = 0, U_k = 0)aq + P^2(Y \mid X = 0, U_k = 0)(1-a)(1-q) \\
&\quad + P^2(Y \mid X = 0, U_k = 1)(1-a)q + P^2(Y \mid X = 0, U_k = 1)a(1-q)) \\
&+ 2^{-l-1} \sum_{z \in \text{UNSAT}} (P(Y \mid X = 0, U_k = 0) + P(Y \mid X = 0, U_k = 1)).
\end{aligned}$$

Thus

$$\begin{aligned}
& P^1(Y \mid \text{do}(X = 0)) - P^2(Y \mid \text{do}(X = 0)) \\
&= 2^{-l} \sum_{z \in \text{SAT}} \left(\left(\frac{1}{2} - aq - (1-a)(1-q) \right) P(Y \mid X = 0, U_k = 0) \right. \\
&\quad \left. + \left(\frac{1}{2} - (1-a)q - a(1-q) \right) P(Y \mid X = 0, U_k = 1) \right) \\
&= 2^{-l} \sum_{z \in \text{SAT}} (rP(Y \mid X = 0, U_k = 0) - rP(Y \mid X = 0, U_k = 1)).
\end{aligned}$$

where $r = 1/2 - aq - (1-a)(1-q)$. It is apparent that we can always choose q and p (i.e., a) in such a way that the causal effects will differ as long as $P(Y \mid X = 0, U_k = 0) \neq P(Y \mid X = 0, U_k = 1)$. This shows that the causal effect is non-identifiable if the 3-SAT instance is satisfiable. \square

Proof of Theorem 2. Note that this proof follows closely to the original soundness proofs of do-calculus [3]. Any formula that is derivable with do-calculus is also derivable here with rule 1 of the calculus presented in the main paper. This is because conditioning on $I_X = 1$ is exactly equivalent to removing all incoming edges of a node. The equivalent inference rules to the do-calculus formulation in [1] are presented below, the notation $\parallel \mathbf{X}$ means that \mathbf{X} are intervened and all edges into them are cut.

Rule 1 (insertion/deletion of observations) originally

$$P(\mathbf{Y} \mid \text{do}(\mathbf{X}), \mathbf{Z}, \mathbf{W}) = P(\mathbf{Y} \mid \text{do}(\mathbf{X}), \mathbf{W}) \text{ if } \mathbf{Y} \perp\!\!\!\perp \mathbf{Z} \mid \mathbf{X}, \mathbf{W} \parallel \mathbf{X}$$

is just

$$P(\mathbf{Y} \mid \mathbf{X}, I_{\mathbf{X}} = 1, \mathbf{Z}, \mathbf{W}) = P(\mathbf{Y} \mid \mathbf{X}, I_{\mathbf{X}} = 1, \mathbf{W}) \text{ if } \mathbf{Y} \perp\!\!\!\perp \mathbf{Z} \mid \mathbf{X}, \mathbf{W}, I_{\mathbf{X}} = 1$$

Rule 2 (action/observation exchange) originally

$$P(Y | \text{do}(X), \text{do}(Z), W) = P(Y | \text{do}(X), Z, W) \text{ if } Y \perp\!\!\!\perp I_Z | X, Z, W \parallel X$$

is simply

$$\begin{aligned} P(Y | X, I_X = 1, Z, I_Z = 1, W) &= P(Y | X, I_X = 1, Z, I_Z = 0, W) \\ &\quad \text{if } Y \perp\!\!\!\perp I_Z | X, Z, W, I_X = 1 \end{aligned}$$

Rule 3 (Insertion/deletion of actions) originally in form

$$P(Y | \text{do}(X), \text{do}(Z), W) = P(Y | \text{do}(X), W) \text{ if } Y \perp\!\!\!\perp I_Z | X, W \parallel X$$

is a bit more complicated.

One way of deriving the equivalence with the presented rule is to first drop conditioning on Z and then change conditioning on $I_Z = 1$.

$$\begin{aligned} P(Y | X, I_X = 1, Z, I_Z = 1, W) &= P(Y | X, I_X = 1, I_Z = 1, W) \\ &\quad \text{if } Y \perp\!\!\!\perp Z | X, W, I_X = 1, I_Z = 1 \\ P(Y | X, I_X = 1, I_Z = 1, W) &= P(Y | X, I_X = 1, I_Z = 0, W) \\ &\quad \text{if } Y \perp\!\!\!\perp I_Z | X, W, I_X = 1 \end{aligned}$$

In the lemma below we show the equivalence of the original condition to the conjunction of the new conditions. Anything that can be derived with rule 3 of do-calculus can also be derived with rule 1 of the paper. \square

Lemma 1. *D-separation relation $Y \perp\!\!\!\perp I_Z | X, W \parallel X$ is equivalent to $[Y \perp\!\!\!\perp I_Z | X, I_X = 1, W] \wedge [Y \perp\!\!\!\perp Z | X, W, I_X = 1, I_Z = 1]$.*

Proof. The original graphical condition $Y \perp\!\!\!\perp I_Z | X, W \parallel X$ is clearly equivalent to the first condition $Y \perp\!\!\!\perp I_Z | X, I_X = 1, W$, the only difference is notational. This shows one direction and the first part of the other direction. What remains to be shown is that $Y \perp\!\!\!\perp Z | X, W, I_X = 1, I_Z = 1$ follows from $Y \perp\!\!\!\perp I_Z | X, W \parallel X$.

We are not assuming any other labels or CSIs on the graph except for the ones brought by the use of intervention variables. We assume the form of d-separation which allows for repeated edges [4]. We also leave out the intervening and conditioning on X in both conditions as we can just delete edges into X in all considered graphs. Proving here the contrapositive, so assume $Y \not\perp\!\!\!\perp Z | W, I_Z = 1$. Thus there is a d-connecting path between some $Y_j \in Y$ and some $Z_i \in Z$ active when conditioning on $W, I_Z = 1$. Due to the nature of intervention and intervention variables, conditioning on $I_Z = 1$ can only break d-connecting paths, so we can drop the conditioning on I_Z . Then, continuing the path with the arc $I_{Z_i} \rightarrow Z_i$ gives us a d-connecting path between I_{Z_i} and Y_j . It is not intercepted at Z_i , since Z_i was originally intervened on ($I_{Z_i} = 1$), and thus the path to Y_j must be out of Z_i . Hence $Y \not\perp\!\!\!\perp I_Z | W$. This shows that $Y \perp\!\!\!\perp I_Z | W \Rightarrow Y \perp\!\!\!\perp Z | W, I_Z = 1$ and adding intervention on X back in we have $Y \perp\!\!\!\perp I_Z | X, W \parallel X \Rightarrow Y \perp\!\!\!\perp Z | X, W, I_X = 1, I_Z = 1$. \square

Proof of Theorem 3. Suppose that there exists a set C such that $Y \perp\!\!\!\perp Z | X, w, C$ is implied by an LDAG G . If $Y \perp\!\!\!\perp C | X, w$, then $Y \perp\!\!\!\perp C, Z | X, w$ by contraction and $Y \perp\!\!\!\perp Z | X, w$ by decomposition. If $C \not\perp\!\!\!\perp Z | X, w$, the CSI is obtained similarly by deriving $Z \perp\!\!\!\perp Y, C | X, w$ by contraction. If $Y \perp\!\!\!\perp C | X, Z, w$ in G then we can write

$$\begin{aligned} P(Y | X, w) &= \sum_C P(Y | X, w, c) P(c | X, w) \\ &= \sum_C P(Y | X, Z, w, c) P(c | X, w) \\ &= \sum_C P(Y | X, Z, w) P(c | X, w) \\ &= P(Y | X, Z, w). \end{aligned}$$

If $Z \perp\!\!\!\perp C | X, Y, w$, the result follows by swapping the roles of Y and Z in the above derivation. \square

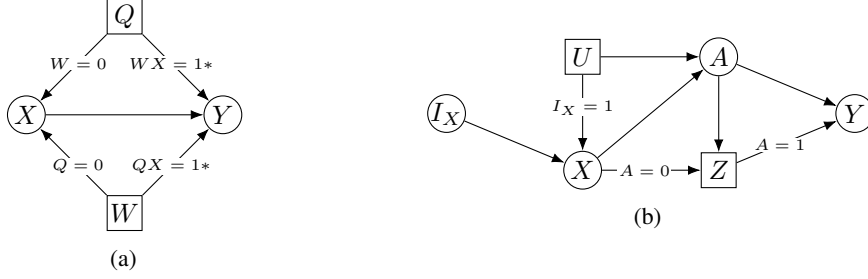


Figure 2: LDAGs for the proof of Theorem 6.

Proof of Theorem 4. Let $c_i \in \text{val}(\mathcal{C})$ and let R_i be a representative of $[c_i]_{\sim}$. Then Y and Z are CSI-separated by X in the context w, c_j in G for every $c_j \in \text{val}(\mathcal{C})$ such that $[c_j]_{\sim} = [c_i]_{\sim}$ because their context-specific DAGs are identical. Since the choice of c_i is arbitrary and the equivalence classes partition $\text{val}(\mathcal{C})$, the claim holds for all $c \in \text{val}(\mathcal{C})$. \square

Proof of Theorem 5. Each new term is identified by using only valid rules of the calculus described in the paper and by ensuring that the required CSIs are implied by G when rule 1 is applied. Theorem 2 shows that CSI-separation is complete for determining CSIs resulting from intervention nodes. It follows that if the search terminates and returns a formula for the target distribution, it was reached from $P(W)$ through a chain of valid manipulations. Suppose now that the search returns NA. By definition, Algorithm 1 enumerates every rule in every valid way. Furthermore, the order in which the distributions are expanded is irrelevant when identifiability is concerned, since every possible manipulation will still be carried out. The search will eventually terminate, since we only identify any specific term once and the set of valid manipulations through the rules for any given term is clearly finite. \square

Proof of Theorem 6. First we show that identifiability of $P(Y | \text{do}(X))$ in every context specific DAG does not imply identifiability in the LDAG itself. Consider LDAG of Figure 2(a). Clearly for any context over $S = Q, W$ there are no latent confounders between X and Y and therefore the effect $P(Y | \text{do}(X))$ is identifiable for all context-specific DAGs.

Let unobserved Q, W be uniformly distributed. Let $X := Q \wedge W$ with a flip with probability .1. Let $Y := Q \vee W$, with a flip with probability .1. The positive joint can be described by:

Q	W	$X = 0$	$X = 1$	$Y = 0$	$Y = 1$
0	0	9/10	1/10	9/10	1/10
0	1	9/10	1/10	1/10	9/10
1	0	9/10	1/10	1/10	9/10
1	1	1/10	9/10	1/10	9/10

Which induces marginal:

$P(X, Y)$	$Y = 0$	$Y = 1$
$X = 0$	1/4	9/20
$X = 1$	1/20	1/4

The marginal distributions are $P(X) = (7/10, 3/10)$ and $P(Y) = (3/10, 7/10)$. Conditionals of Y are $P(Y | X = 0) = (5/14, 9/14)$ and $P(Y | X = 1) = (1/6, 5/6)$. The causal effect $P(Y | \text{do}(X)) = P(Y)$ as forcing X does not influence Y . Let the second model be such that Q, W are uniform, X is distributed as $P(X)$ in the first model, and Y as $P(Y | X)$. The marginal of X, Y is the same, labels are respected, but causal effect $P(Y | \text{do}(X)) = P(Y | X) \neq P(Y)$.

Next we show that identifiability of $P(Y | \text{do}(X), c)$ does not imply the identifiability of $P(Y | \text{do}(X))$ in the LDAG. Consider the LDAG of Fig. 2(b). It is possible to identify the conditional causal effects $P(Y | \text{do}(X), A = 0)$ and $P(Y | \text{do}(X), A = 1)$ from $P(X, Y, A)$ since

$$P(Y | \text{do}(X), A) = P(Y | A, X, I_X = 1) = P(Y | A, X),$$

as Y is independent of I_X in the LDAG. However, the causal effect $P(Y | \text{do}(X))$ is not identifiable. To show this, we define 2 parameterizations, M^1 and M^2 , for the LDAG, that agree on $P(X, Y, A)$

but disagree on $P(Y | \text{do}(X))$. In the first model we let $P^1(U) = 1/2$, $P^1(X = 1 | U = 1) = q$ and $P(X = 1 | U = 0) = 1 - q$. In the second model we let $P^2(U = 1) = q$ and $P^2(X | U) = 1/2$. We also let $P^1(A | X = 1, U) = P^2(A | X = 1, U) = P(A | X = 1, U)$.

We have that

$$\begin{aligned} P(X, A) &= \sum_U P(X, A, U) \\ &= P(U = 1)P(X | U = 1)P(A | X, U = 1) + P(U = 0)P(X | U = 0)P(A | X, U = 0). \end{aligned}$$

For $X = 1$ and we have

$$P^1(X = 1, A) = P^2(X = 1, A) = \frac{1}{2}qP(A | X = 1, U = 1) + \frac{1}{2}(1 - q)P(A | X = 1, U = 0),$$

for $X = 0$ we have

$$\begin{aligned} P^1(X = 0, A) &= \frac{1}{2}(1 - q)P^1(A | X = 0, U = 1) + \frac{1}{2}qP^1(A | X = 0, U = 0), \\ P^2(X = 0, A) &= \frac{1}{2}qP^2(A | X = 0, U = 1) + \frac{1}{2}(1 - q)P^2(A | X = 0, U = 0), \end{aligned}$$

thus we set $P(A | X = 0, U = 1) = P(A | X = 0, U = 0) = 1/2$ for both M^1 and M^2 . Furthermore, we let $P^1(Y, Z | X, A) = P^2(Y, Z | X, A) = P(Y, Z | X, A)$. This parametrization produces the same observed joint distribution $P^1(Y, X, A) = P^2(Y, X, A)$. Computing the causal effects result in

$$\begin{aligned} P^1(Y | \text{do}(X = 1)) &= \sum_{U, A, Z} P^1(U)P^1(A | U, X = 1)P^1(Y, Z | X = 1, A) \\ &= \sum_{A, Z} \left(\frac{1}{2}P(A | X = 1, U = 1)P(Y, Z | X = 1, A) \right. \\ &\quad \left. + \frac{1}{2}P(A | X = 1, U = 0)P(Y, Z | X = 1, A) \right), \end{aligned}$$

and

$$\begin{aligned} P^2(Y | \text{do}(X = 1)) &= \sum_{U, A, Z} P^2(U)P^2(A | U, X = 1)P^2(Y, Z | X = 1, A) \\ &= \sum_{A, Z} (qP(A | X = 1, U = 1)P(Y, Z | X = 1, A) \\ &\quad + (1 - q)P(A | X = 1, U = 0)P(Y, Z | X = 1, A)). \end{aligned}$$

The difference $P^1(Y | \text{do}(X = 1)) - P^2(Y | \text{do}(X = 1))$ is

$$\sum_{A, Z} \left[P(Y, Z | X = 1, A) \left(\left(\frac{1}{2} - q \right) P(A | X = 1, U = 1) - \left(\frac{1}{2} - q \right) P(A | X = 1, U = 0) \right) \right],$$

which is non-zero as long as $P(A | X = 1, U = 1) \neq P(A | X = 1, U = 0)$.

□

B Heuristics

In order to guide the search to identify the most promising terms, we relate the source distributions to the target quantity through a proximity function and always expand the closest term first. We construct our proximity function h by comparing the variables and value assignments that appear on the left-hand and right-hand sides of the target $P^t = P(\mathbf{Y}_1, \mathbf{y}_2 | \mathbf{X}_1, \mathbf{x}_2)$ and the source term $P^s = P(\mathbf{A}_1, \mathbf{a}_2 | \mathbf{B}_1, \mathbf{b}_2)$. Let \mathcal{S} be the set of variables where value assignments between the terms agree and let \mathcal{T} be the complement where the values disagree. We also denote $\mathbf{Y} = \mathbf{Y}_1 \cup \mathbf{Y}_2$, $\mathbf{X} = \mathbf{X}_1 \cup \mathbf{X}_2$, $\mathbf{A} = \mathbf{A}_1 \cup \mathbf{A}_2$ and $\mathbf{B} = \mathbf{B}_1 \cup \mathbf{B}_2$. Following the approach of [5], we define the proximity function as:

$$h(P^t, P^s) = 10|\mathbf{Y} \cap \mathbf{A}| + 5|\mathbf{X} \cap \mathbf{B}| + 3|\mathcal{S}| - 2|\mathbf{Y} \setminus \mathbf{A}| - 2|\mathbf{A} \setminus \mathbf{Y}| - 2|\mathbf{X} \setminus \mathbf{B}| - 2|\mathbf{B} \setminus \mathbf{X}| - |\mathcal{T}|.$$

Note that we also penalize the term $|\mathbf{A} \setminus \mathbf{Y}|$ even though extra variables on the left-hand side of a term do not hinder identifiability. Having a penalty for this term seems to have the tendency to produce simpler formulas for the causal effects.

References

- [1] A. Hyttinen, F. Eberhardt, and M. Järvisalo. Do-calculus when the true graph is unknown. In *Proceedings of the 31st Conference on Uncertainty in Artificial Intelligence*, pages 395–404. AUAI Press, 2015.
- [2] D. Koller and N. Friedman. *Probabilistic Graphical Models: Principles and Techniques*. MIT Press, 2009.
- [3] J. Pearl. Causal diagrams for empirical research. *Biometrika*, 82(4):669–688, 1995.
- [4] M. Studený. Bayesian networks from the point of view of chain graphs. In *Proceedings of the 14th conference on Uncertainty in artificial intelligence*, pages 496–503. Morgan Kaufmann, 1998.
- [5] S. Tikka, A. Hyttinen, and J. Karvanen. Causal effect identification from multiple incomplete data sources: A general search-based approach. <https://arxiv.org/abs/1902.01073>, 2019.