

Deriving complete constraints in discrete hidden variable DAGs: some practical issues

Erin E Gabriel

Michael C. Sachs, Robin J Evans, Arvid
Sjölander

Pioneer Centre for SMARTbiomed
Section of Biostatistics, University of Copenhagen
Department of Statistics, University of Oxford
Institutionen för medicinsk epidemiologi och biostatistik, Karolinska
Institutet



Outline

Our method

- Preliminaries

- The Algorithm

- Properties

Practical Issues

- Determining nontriviality

- C-degree ≥ 2

- Scalability

Conclusion

Questions

1. Is there a systematic method for getting inequality and equality constraints in DAGs with hidden variables?
2. If yes, are there conditions under which we know we have the complete set of constraints?
3. Is there a way to identify which constraints are nontrivial?

Previous work

Inequalities

- Pearl [1995] derived the IV inequalities from “first principles”. They are complete only for the binary IV model.
- Bonet [2001] finds the complete set of constraints in almost all IV settings.
- Śliwa [2003] and Pironio et al. [2011] (and many others) derive inequalities for different variations of the Bell scenario.
- Kang and Tian [2006] generalize these approaches, but do not get the complete set.

Equalities

- Verma and Pearl [1991] early attempt; Shpitser et al. [2012] more generalized and they characterize all models with 4 observed variables
- [Evans and Richardson, 2019, Richardson et al., 2023] the nested-Markov characterization.

Corollary 5 *The extreme points of B are the columns of matrix A_2 .*

$$A_1 = \left[\begin{array}{cccc|cccc} 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 \\ \hline 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 \end{array} \right], \quad A_2 = \left[\begin{array}{cccc|cccc|cccc|cccc} 1 & \star & 0 & \star & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \star & \star \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 \\ 0 & 0 & 1 & 1 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 \\ \hline 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 \\ 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 \end{array} \right]$$

Figure 2: Example of matrices A_1 and A_2 for the binary case. A_1 is a $n^l m = 8$ times $m n^l = 8$ matrix and A_2 is a $m n^l = 8$ times $n^l m^n = 16$ matrix. The $\star$ mark repeated columns.

B is the space where the observed probabilities live under the IV model.

Notation

- Let $\mathcal{G}(\mathbf{W}, \mathbf{U})$ be a hidden variables DAG
- over variables $\mathbf{V} = \mathbf{W} \cup \mathbf{U}$ where $\mathbf{U}$ are unobserved and $\mathbf{W}$ are observed.
- All variables in $\mathbf{W}$ are discrete and finite, but no assumptions about $\mathbf{U}$.
- $\text{Paobs}(\mathbf{W}_i)$ = union of observed parents of variables in $\mathbf{W}_i$, likewise for $\text{Chobs}(\mathbf{W}_i)$, and $\text{Anobs}(\mathbf{W}_i)$ (which also includes $\mathbf{W}_i$ itself)
- Let $|\mathbf{W}|$ denote the cardinality of the set $\mathbf{W}$, and when used for a single variable, for example $|W_i|$, the cardinality of its domain.

Setting

We restrict our attention to DAGs meeting the following conditions:

Condition

All unobserved variables are exogenous, i.e., $Pa(U_i) = \emptyset$ for all $U_i \in \mathbf{U}$.

Condition

The child set of each unobserved variable correspond to a distinct set of observed vertices of at least size 2. That is, for each $U_i \in \mathbf{U}$, $|\text{Chobs}(U_i)| \geq 2$, $\text{Chobs}(U_i) \subseteq \mathbf{W}_i$, and there is no $U_j \in \mathbf{U}$ such that $\text{Chobs}(U_i) \subseteq \text{Chobs}(U_j)$.

Thanks to Evans [2016], we are not losing anything by enforcing Conditions 1 and 2: DAGs where these conditions do not hold can be transformed to DAGs where they do hold without changing the implied marginal model.

Definitions

Definition

Let $\mathcal{G}(\mathbf{W}, \mathbf{U})$ be a hidden variables DAG that obeys Conditions 1 and 2. Remove all edges that do not originate from variables in $\mathbf{U}$. The sets of variables in $\mathbf{W}$ that remain connected form a partition of $\mathbf{W}$ called districts. Denote $\mathcal{D}(\mathcal{G}) = \{D_1, \dots, D_k\}$ the set of k districts.

This same concept is called a c-component (short for confounded component) by Tian and Pearl [2002b], while the term district is used in, e.g., Richardson [2003], Evans [2018], and Duarte et al. [2023], which we will follow.

Definition

The c-degree of a district D_j is the maximum of 1 and the number of variables U_i in $\mathbf{U}$ for which there exists a $W_k \in D_j$ such that $U_i \in \text{Pa}(W_k)$ in $\mathcal{G}(\mathbf{W}, \mathbf{U})$.

The c-degree of $\mathcal{G}(\mathbf{W}, \mathbf{U})$ is the maximum of the c-degrees of its districts.

Definition: canonical partition

Definition

Let $\mathcal{G}(\mathbf{W}, \mathbf{U})$ be a hidden variables DAG that follows Conditions 1, 2, and has a c-degree of 1. The response variable transformation of $\mathcal{G}(\mathbf{W}, \mathbf{U})$ is a mixed graph $\mathcal{G}(\mathbf{W}, \mathbf{R})$ where:

- for each $W_i \in \mathbf{W}$, introduce a new unobserved variable R_{W_i} that is discrete with $|W_i|^{k_i}$ levels, where $k_i = \prod_{V \in \text{Pa}_{\text{obs}}(W_i)} |V|$;
- introduce an undirected edge $R_{W_i} - R_{W_j}$ for every pair of variables W_i, W_j that are in the same district.

The value of R_{W_i} determines the way that W_i responds to its parents. We remove $\mathbf{U}$ having arbitrary distribution and replaces them with categorical $\mathbf{R}$. These are called response function variables by Balke and Pearl [1994] and selectors by Bonet [2001].

Key point: This does not change the model.

Example

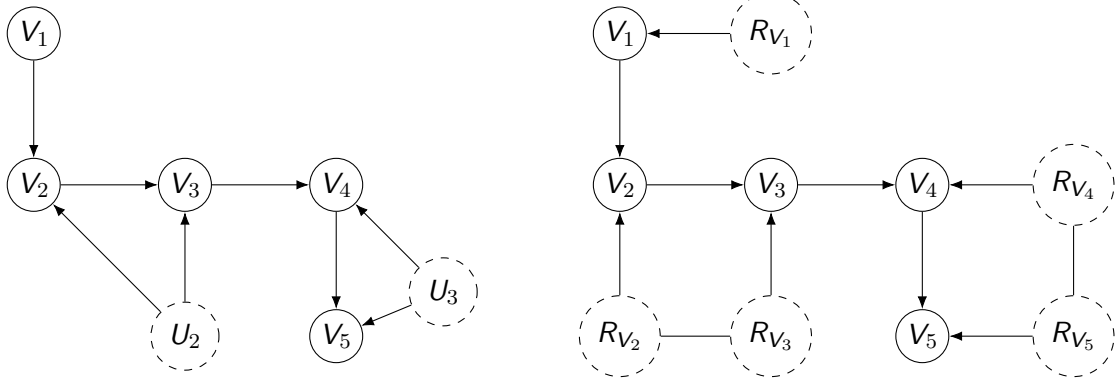


Figure: A sequential instrumental variable example, and its response function variable transformation.

Heading towards constraints

- We exploit the discreteness of the observations.
- Say $R_{W_i} = b$, there exists a response function $\phi_{W_i}(\text{Paobs}(W_i))$ that determines a value for W_i given $\text{Paobs}(W_i)$.
- So $\phi_{W_i} : \text{domain}(\text{Paobs}(W_i)) \rightarrow \text{domain}(W_i)$.
- Given values of the response function variables for W_i and all of its ancestors, the value of W_i is determined by recursive substitutions [Sachs et al., 2023, Duarte et al., 2023].

Remark

Therefore we can write observable probabilities in terms of probabilities of response function variables. Following Tian and Pearl [2002b], we will call the equations relating these two forms of probabilities functional constraints.

Functional constraints

Proposition

Let $\mathbf{W}_1$ be the vector of all variables in the same district D_1 , and let $\mathbf{W}_2 = \text{Paobs}(\mathbf{W}_1) \setminus \mathbf{W}_1$ be the vector of parents of $\mathbf{W}_1$ that are not part of $\mathbf{W}_1$. If D_1 has a c-degree 1, then for all possible values of the vectors

$w_1 = (w_{11}, \dots, w_{1k}), w_2 = (w_{21}, \dots, w_{2\ell}),$

1. define the probabilities $P^*(\mathbf{W}_1 = w_1 | \mathbf{W}_2 = w_2)$ as

$$\prod_{i: W_{1i} \in D_1} P(W_{1i} = w_{1i} | \{\text{Paobs}(\mathbf{W}_1) \cap \text{Anobs}(W_{1i})\} \setminus W_{1i}).$$

and

2. the functional constraints on $P^*(\mathbf{W}_1 = w_1 | \mathbf{W}_2 = w_2)$ under $\mathcal{G}(\mathbf{W}, \mathbf{U})$ are linear functions of response function variable probabilities of $\mathbf{R}_{\mathbf{W}_1}$.

Side note on causality

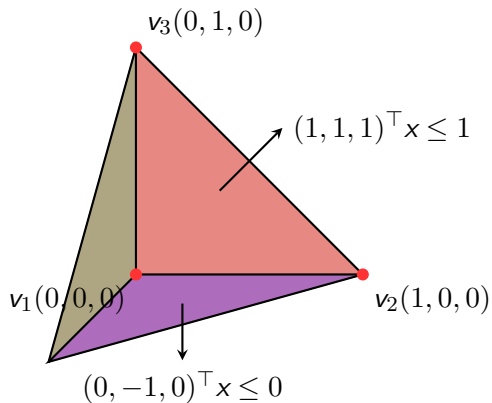
One could consider $\mathcal{G}(\mathbf{W}, \mathbf{U})$ a causal DAG and let $P(\mathbf{W}_1 = w_1 | \text{do}(\mathbf{W}_2 = w_2))$ denote the joint probability of $\mathbf{W}_1 = w_1$ under the intervention that sets $\mathbf{W}_2 = w_2$. All of the $P^*(\mathbf{W}_1 = w_1 | \mathbf{W}_2 = w_2)$ referenced in Proposition 1 can be given a causal interpretation using the do operator, such that $P^*(\mathbf{W}_1 = w_1 | \mathbf{W}_2 = w_2) = P(\mathbf{W}_1 = w_1 | \text{do}(\mathbf{W}_2 = w_2))$. Tian and Pearl [2002a]

One could also consider similar counterfactual quantities, e.g., $P(\mathbf{W}_1(w_2) = w_1)$, rather than the do-operator interventional quantities.

We have everything need, but in the wrong form

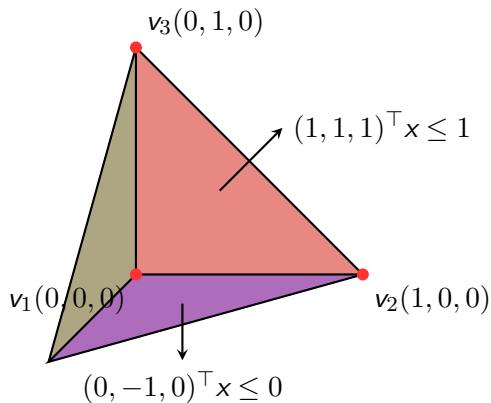
- Functional constraints describe the relationships between observable probabilities and joint probabilities of the response function variables.
- We carefully chose which quantities to look at so that they are:
 1. Linear in the joint response probabilities, and
 2. Cover the full joint observed distribution
- For each district, the proposition provides a matrix equation $\mathbf{p} = \mathbf{B}\mathbf{r}$, relating quantities involving only observed probabilities linearly to joint distribution of response variables for that district. $\mathbf{B}$ is a matrix of 0s and 1s.
- Lastly, we need a way to state the constraints in terms of only the observables. One could use quantifier elimination (e.g., CAD Geiger and Meek [1999]), but we use a geometric approach. They will provide the same result.

V-to-H conversion



- The matrix equation $\mathbf{p} = \mathbf{B}\mathbf{r}$ encodes the assumptions that the DAG imposes on the observed distribution.
- Bonet [2001] noted that the columns of $\mathbf{B}$ are the extreme points of the observables under the model or a V-presentation in terms of $\mathbf{p}$.
- A convex polytope can also be represented by describing the intersection of half-planes (H-representation). [Minkowski-Weyl Theorem]

V-to-H conversion



- There are good algorithms to do the V-to-H conversion (and vice versa), e.g., Fukuda [2018].
- Going from V-to-H gives us expressions involving only the $\mathbf{p}$ space.
- The H-representation consists of two rational matrices H_i, H_e and rational vectors b_i, b_e such that

$$H_i \mathbf{p} \leq b_i \text{ and } H_e \mathbf{p} = b_e.$$

Outline

Our method

Preliminaries

The Algorithm

Properties

Practical Issues

Determining nontriviality

C-degree ≥ 2

Scalability

Conclusion

Algorithm overview

Inputs: A hidden variables DAG $\mathcal{G}(\mathbf{W}, \mathbf{U})$, that meets Conditions 1, 2, and has c-degree 1.

1. List the conditional independencies using a standard algorithm based on checking for d-separation. **Output:** List of standard independence constraints.
2. Identify the districts $\mathcal{D} = \{D_1, \dots, D_k\}$.
3. Do the response function variable transformation of $\mathcal{G}$ into $\mathcal{G}^\dagger$.
4. For each district D_j :
 - 4.1 Construct the matrix representing the system of linear equations relating

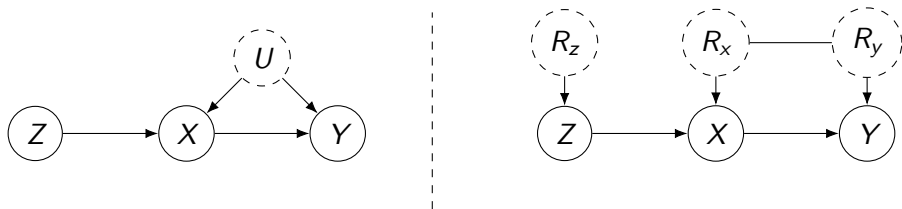
$$P^*(\mathbf{W}_1 = w_1 | \mathbf{W}_2 = w_2) = \sum_{b \in \mathcal{B}} P(\mathbf{R}_{\mathbf{W}_1} = b) \Leftrightarrow \mathbf{p} = \mathbf{B}r$$

for $\mathbf{W}_1 = D_j$ and $\mathbf{W}_2 = \text{Paobs}(\mathbf{W}_1) \setminus \mathbf{W}_1$

- 4.2 Convert to the H-representation

Output: The H-representation and the P^* components for each district

Example – The IV model



$$P^*(\mathbf{W}_1 = w_1 | \mathbf{W}_2 = w_2) = P(Y = y | X = x, Z = z)P(X = x | Z = z) = \\ \sum_{b \in \mathcal{B}} P((R_x, R_y) = b), \text{ for all } x, y, z,$$

where $\mathcal{B} = \{(b_x, b_y) : f_Y^{b_y}(f_X^{b_x}(z)) = y \text{ and } f_X^{b_x}(z) = x\}$. This can be represented by a matrix equation

$$\mathbf{p}_{xy \cdot z} = \mathbf{B} \mathbf{r}_{xy}.$$

IV example continued

H-representation defined by $b_i = (0, 0, 0, 0, 1, 1, 0, 1, 0, 1, 0, 0)^\top$,

$$H_i = \begin{pmatrix} 0 & 0 & 0 & 0 & -1 & 0 & 0 & 0 \\ -1 & -1 & -1 & 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & -1 & -1 & -1 & 0 \\ -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & -1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 & 1 & 0 \\ 0 & -1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & -1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & -1 & 0 \end{pmatrix},$$

$b_e = (-1, -1)^\top$, and

$$H_e = \begin{pmatrix} -1 & -1 & -1 & -1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & -1 & -1 & -1 & -1 \end{pmatrix}.$$

Interpretation

These expressions encode both inequality and equality constraints. For example, the 5th row of H_i and corresponding element of b_i encode the constraint

$$P(Y = 1, X = 0 | Z = 0) + P(Y = 0, X = 0 | Z = 1) \leq 1,$$

which is one of the classic instrumental inequalities as derived by Pearl [1995].

The equality constraints encoded by H_e are the standard probabilistic constraints:

$$\sum_{x,y \in \{0,1\}} P(Y = x, X = y | Z = z) = 1, \text{ for } z = 0, 1.$$

The district $\{Z\}$ does not admit any constraints other than the probabilistic ones.

Outline

Our method

Preliminaries

The Algorithm

Properties

Practical Issues

Determining nontriviality

C-degree ≥ 2

Scalability

Conclusion

Completeness

Proposition

Let $\mathcal{G}$ be a discrete hidden variable DAG satisfying Conditions 1, 2, and having c -degree 1. Then the constraints derived from the algorithm combined with any conditional independencies are complete.

Proof overview:

- The district factorization allows us to recover the joint observable distribution $P(\mathbf{W})$ from the joints within each district given its parents. Then within each district:
- The functional constraints cannot be further factorized because we limit to c -degree 1. Thus they are linear.
- The V-to-H conversion preserves all the functional constraints.
- The only cross-district constraints are conditional independencies, which are enumerated as a separate step.

Trivial constraints

In our IV example, there are some obviously trivial constraints like

$$\sum_{x,y \in \{0,1\}} P(Y = x, X = y | Z = z) = 1, \text{ for } z = 0, 1.$$
$$0 \leq P(Y = x, X = y | Z = z) \leq 1.$$

To identify the subset of potentially nontrivial constraints, consider an unconstrained $\check{\mathbf{p}}$ that lives in the Cartesian product of probability simplices, which is a convex polyhedron. Since the constraints are linear, violations can be witnessed at the extreme points of the unconstrained polyhedron (fundamental theorem of linear programming). If these extreme points do not produce violations, the constraints are definitely trivial.

Triviality

Proposition

Let $\check{\mathbf{P}}$ denote the Cartesian product of probability simplices, one for each unique value of w_2 , in the order corresponding to $\mathbf{p}$. Check whether $H_i \check{\mathbf{p}} > b_i$ or $H_e \check{\mathbf{p}} \neq b_e$ for each $\check{\mathbf{p}} \in V(\check{\mathbf{P}})$. Constraints encoded by H_i and H_e that are not violated by any of the extreme points of $\check{\mathbf{P}}$ are trivial.

We are careful with the phrasing here: No violation $\Rightarrow$ definitely trivial

Violation does not necessarily imply nontrivial and observable.

Outline

Our method

Preliminaries

The Algorithm

Properties

Practical Issues

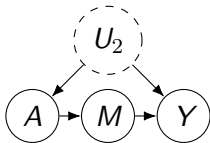
Determining nontriviality

C-degree ≥ 2

Scalability

Conclusion

Candidate nontrivial $\neq$ Actually nontrivial



When applying our algorithm to the front door model, we get this reported as potentially nontrivial:

$$P(A = 0) + P(A = 1) = 1.$$

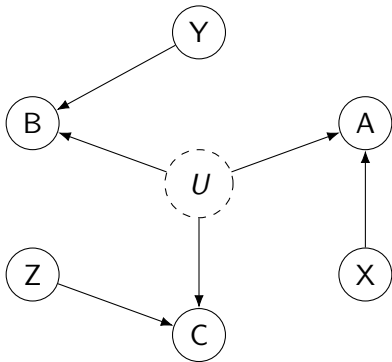
Written in terms of the interventional quantities that were tested by the algorithm:

$$P^*(A = 0|M = 0) = P^*(A = 0|M = 1).$$

Which is easily recognized via the definition,

$$P^*(A = a|M = m) = P(A = a)P(Y = 0|A = a, M = m) + P(A = a)P(Y = 1|A = a, M = m) \text{ as trivially } P(A = 0) = P(A = 0)$$

Tripartite Bell scenario



Our algorithm yields 53,894 constraints aside from the independences, out of which there are 53,856 inequality constraints, in agreement with Śliwa [2003] and Pironio et al. [2011]. Our method further identifies 32,866 of those as potentially nontrivial constraints, 6 of which are equalities. But how to reduce this number?

Outline

Our method

Preliminaries

The Algorithm

Properties

Practical Issues

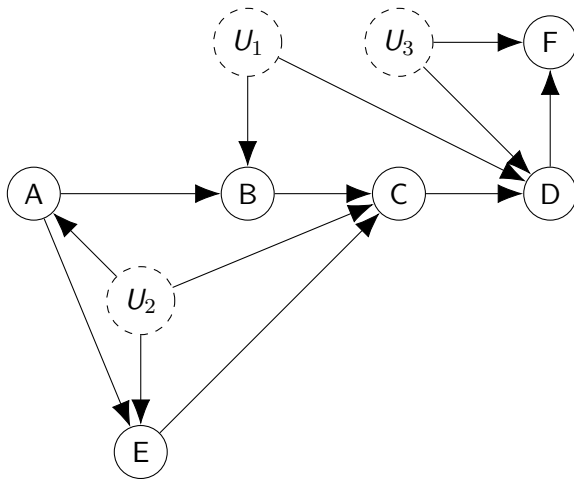
Determining nontriviality

C-degree ≥ 2

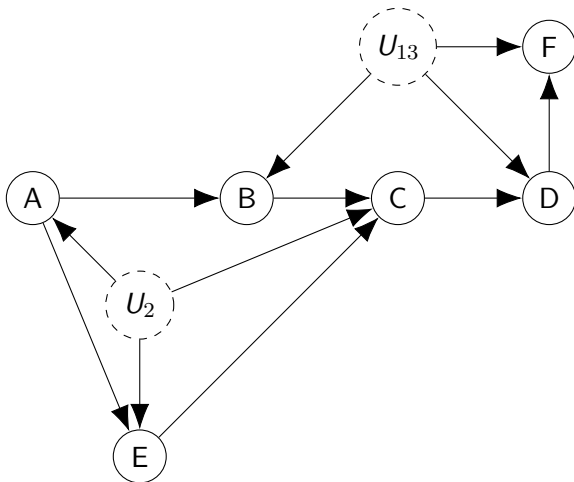
Scalability

Conclusion

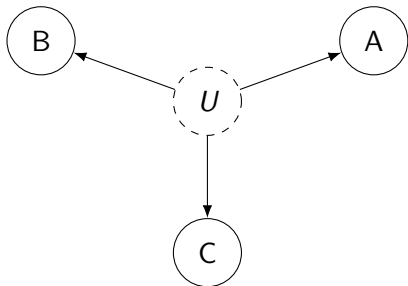
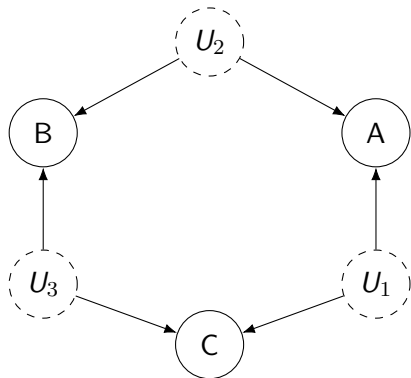
C-degree 2



Merging latents



Triangle scenario



Our procedure does not yield any constraints, but it is known that there are some [Wolfe et al., 2019].

General remark

Prior to applying the algorithm, for each district D_j :

1. check that the c-degree of D_j is at least 2, otherwise pass through;
2. create a new unobserved variable U_j^* ;
3. if there exists an edge from any unobserved variable to W_i in D_j , add an edge $U_j^* \rightarrow W_i$;
4. then remove every unobserved variable other than U_j^* from D_j .

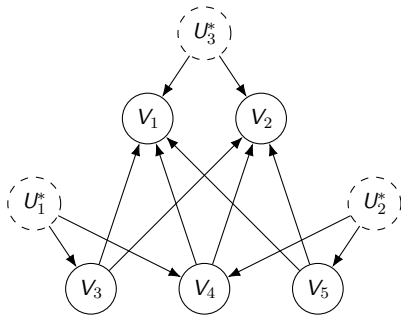
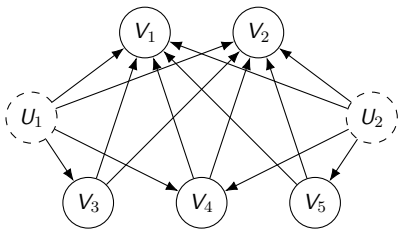
Remark

Let $\mathcal{G}'$ be a discrete hidden variables DAG satisfying Conditions 1, 2, with c-degree greater than 1. Let $\mathcal{G}$ be the graph obtained from by merging unobserved variables in $\mathcal{G}'$ so that the c-degree is 1. Then the constraints admitted by $\mathcal{G}$ hold under $\mathcal{G}'$.

... but they are not necessarily complete.

Other methods

Proposition 5 of Evans [2016] and its extension in [Ansanelli et al., 2025] strong face-splitting and the HLP Proposition (Theorem 26.1 and 26.4 of [Henson et al., 2014]) provide conditions under which one can add/remove an edge or split-up a latent from a DAG without changing the observable constraints.



There are potential ways forward Evans [2023]

Outline

Our method

Preliminaries

The Algorithm

Properties

Practical Issues

Determining nontriviality

C-degree ≥ 2

Scalability

Conclusion

Computation time excluding triviality checking

setting	time	memory allocated
iv-222	38.5ms	1016.2KB
iv-322	39.4ms	1.11MB
iv-332	245.4ms	2.31MB
iv-333	29.5s	6.05MB
seq-2-iv	65.9ms	2.57MB
seq-3-iv	338.2ms	32.45MB
med-222	39.7ms	1.13MB
bip-Bell-2222	39.7ms	1.11MB
bip-Bell-2233	787.9ms	2.36MB
trip-Bell	18.4m	56.84MB
quad-Bell	–	Out of memory (30GB)

Settings: iv-zxy= IV with number of levels for z,x,y; seq-2-iv=two sequential IVs; seq-3-iv=three sequential IVs; med-222=mediation all binary; bip-Bell-xyab=Bipartite Bell; trip-Bell=tripartite Bell all binary. Shridharan and Iyengar [2023] suggest methods for pruning.

Summary

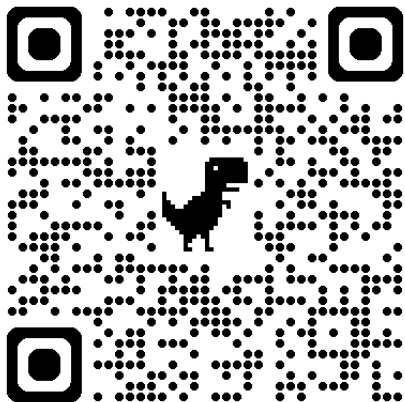
What we have done:

- We show that one can extend Bonet [2001] beyond IV provided the functional constraints are linear, thus providing the complete observable constraints
- We consider a factorization of the joint for which the functional constraints are always linear within a given class of discrete graphs, i.e. $c\text{-degree} = 1$.
- We provide algorithms to obtain all constraints and for selecting candidate nontrivial constraints
- Previous work can be used to increase the class of models by showing equivalence to models with $c\text{-degree} = 1$.

Limitations and future work:

- When can we not reduce the $c\text{-degree}$?
- Better automatic elimination of candidate nontrivials

Link to manuscript



<https://arxiv.org/abs/2601.11242>

References I

- Marina Maciel Ansanelli, Elie Wolfe, and Robert W. Spekkens. The observational partial order of causal structures with latent variables, 2025. URL <https://arxiv.org/abs/2502.07891>.
- Alexander Balke and Judea Pearl. Counterfactual probabilities: Computational methods, bounds and applications. In Uncertainty in artificial intelligence, pages 46–54. Elsevier, 1994.
- Blai Bonet. Instrumentality tests revisited. In Proceedings of the Seventeenth conference on Uncertainty in Artificial Intelligence, pages 48–55, 2001.
- Guilherme Duarte, Noam Finkelstein, Dean Knox, Jonathan Mummolo, and Ilya Shpitser. An automated approach to causal inference in discrete settings. Journal of the American Statistical Association, pages 1–16, 2023.
- Robin J Evans. Graphs for margins of Bayesian networks. Scandinavian Journal of Statistics, 43(3): 625–648, 2016.
- Robin J. Evans. Margins of discrete Bayesian networks. The Annals of Statistics, 46(6A):2623 – 2656, 2018. doi: 10.1214/17-AOS1631. URL <https://doi.org/10.1214/17-AOS1631>.
- Robin J Evans. Latent free equivalent mdags. Algebraic Statistics, 14(1):3–16, 2023.
- Robin J. Evans and Thomas S. Richardson. Smooth, identifiable supermodels of discrete DAG models with latent variables. Bernoulli, 25(2):848 – 876, 2019. doi: 10.3150/17-BEJ1005. URL <https://doi.org/10.3150/17-BEJ1005>.

References II

- K. Fukuda. cdd, cddplus and cddlib homepage. Swiss Federal Institute of Technology, Zurich., 2018. URL https://people.inf.ethz.ch/fukudak/cdd_home/.
- Dan Geiger and Christopher Meek. Quantifier elimination for statistical problems. In Proceedings of the Fifteenth conference on Uncertainty in artificial intelligence, pages 226–235, 1999.
- Joe Henson, Raymond Lal, and Matthew F Pusey. Theory-independent limits on correlations from generalized Bayesian networks. New Journal of Physics, 16(11):113043, 2014.
- Changsung Kang and Jin Tian. Inequality constraints in causal models with hidden variables. In Proceedings of the Twenty-Second Conference on Uncertainty in Artificial Intelligence, pages 233–240, 2006.
- Judea Pearl. On the testability of causal models with latent and instrumental variables. In Proceedings of the Eleventh conference on Uncertainty in artificial intelligence, pages 435–443, 1995.
- Stefano Pironio, Jean-Daniel Bancal, and Valerio Scarani. Extremal correlations of the tripartite no-signaling polytope. Journal of Physics A: Mathematical and Theoretical, 44(6):065303, jan 2011. doi: 10.1088/1751-8113/44/6/065303. URL <https://doi.org/10.1088/1751-8113/44/6/065303>.
- Thomas Richardson. Markov properties for acyclic directed mixed graphs. Scandinavian Journal of Statistics, 30(1):145–157, 2003.

References III

- Thomas S Richardson, Robin J Evans, James M Robins, and Ilya Shpitser. Nested Markov properties for acyclic directed mixed graphs. The Annals of Statistics, 51(1):334–361, 2023.
- Michael C Sachs, Gustav Jonzon, Arvid Sjölander, and Erin E Gabriel. A general method for deriving tight symbolic bounds on causal effects. Journal of Computational and Graphical Statistics, 32(2):567–576, 2023.
- Ilya Shpitser, Thomas S Richardson, James M Robins, and Robin Evans. Parameter and structure learning in nested Markov models. In 28th Conference on Uncertainty in Artificial Intelligence (Causal Structure Learning Workshop), 2012.
- Madhumitha Shridharan and Garud Iyengar. Scalable computation of causal bounds. Journal of Machine Learning Research, 24(237):1–35, 2023. URL <http://jmlr.org/papers/v24/22-1081.html>.
- Jin Tian and Judea Pearl. A general identification condition for causal effects. In Aaai/iaai, pages 567–573, 2002a.
- Jin Tian and Judea Pearl. On the testable implications of causal models with hidden variables. In Proceedings of the Eighteenth conference on Uncertainty in artificial intelligence, pages 519–527, 2002b.
- TS Verma and J Pearl. Equivalence and synthesis of causal models. Computer Science Department Technical Report, UCLA, page R150, 1991.

References IV

- Elie Wolfe, Robert W. Spekkens, and Tobias Fritz. The inflation technique for causal inference with latent variables. Journal of Causal Inference, 7(2), July 2019. ISSN 2193-3677. doi: 10.1515/jci-2017-0020. URL <http://dx.doi.org/10.1515/jci-2017-0020>.
- Cezary Śliwa. Symmetries of the Bell correlation inequalities. Physics Letters A, 317(3-4):165–168, 2003. doi: 10.1016/S0375-9601(03)01115-0.