Proceedings of Machine Learning Research 343:153–166, 2026 Probabilistic Graphical Models (PGM)

Accounting for Data Uncertainty in Counterfactual Reasoning

Rafael Cabañas RCABANAS@UAL.ES
Department of Mathematics
Center for the Development and Transfer of Mathematical Research to Industry
University of Almería
Almería, Spain

Helge Langseth HELGE.LANGSETH@NTNU.NO
Department of Computer Science
Norwegian University of Science and Technology
Trondheim, Norway

Thomas D. Nielsen TDN@CS.AAU.DK
Department of Computer Science
Aalborg University
Aalborg, Denmark

Antonio Salmerón ANTONIO.SALMERON@UAL.ES
Department of Mathematics
Center for the Development and Transfer of Mathematical Research to Industry
University of Almería
Almería, Spain

Editors: Gustau Camps-Valls, Manuele Leonelli and Gherardo Varando

Abstract

Structural causal models with discrete variables are a class of probabilistic graphical models used for causal and counterfactual reasoning. In the presence of observational (i.e., non-experimental) data, some queries in such models are non-identifiable and therefore cannot be computed exactly. Instead, several methods derive bounds on non-identifiable queries by solving linear programming problems. However, these methods usually assume a precise specification of the data distribution. In this work, we consider a more realistic scenario in which uncertainty in the data is represented through an interval-valued specification of the data distribution. We formulate the corresponding linear programming problem under this setting, which covers both identifiable and non-identifiable counterfactual queries under data uncertainty. We show how a recently proposed divide-and-conquer algorithm for bounding counterfactual probabilities (Bjøru et al., 2025) can be adapted to such imprecise specifications of the data distribution without a substantial increase in computational complexity.

Keywords: Counterfactual Reasoning, Structural Causal Models, Data Uncertainty

© 2026 R. Cabañas, H. Langseth, T.D. Nielsen & A. Salmerón.

1. Introduction

Structural causal models (SCMs) with discrete variables (Bareinboim et al., 2022; Pearl, 2009) are probabilistic graphical models (PGMs) for causal analysis that enable reasoning about hypothetical scenarios. SCMs consist of endogenous (observable) and exogenous (usually latent) variables, with endogenous values determined from the exogenous ones through structural equations. *Pearl's ladder* (Pearl and Mackenzie, 2018) is often used to identify the three types of queries typically associated with causal analysis. At the first rung are the observational queries (e.g., "What if I *see* this?"), which can be answered using standard statistical models learnt from observational data, based on the calculation of conditional distributions. At the second rung we find interventional queries (e.g., "What if I *do* this?"), which can be solved by, for instance, causal Bayesian models (Pearl, 2009) and are represented by conditional distributions after interventions based on so-called *do-calculus* (Pearl, 2009). The last rung holds counterfactual queries that address hypothetical scenarios (e.g., "What if we had done this instead of passively observing that?"). They are calculable by SCMs. However, when the SCM model is learned from observational data, the result of such queries are not necessarily precisely defined, and the answer to the query will then be given as an imprecise probability, and represented through a (non-degenerate) interval. Unfortunately, the computational complexity of calculating these intervals exactly is overwhelming even for medium-sized models, and this has motivated research into computationally efficient approaches.

The *Divide and Conquer for Causal Computation* (DCCC) method (Bjøru et al., 2024, 2025) can bound non-identifiable queries by, for each exogenous variable, solving a linear programming problem that characterizes all exogenous distributions consistent with the available data. Its main advantage over other approaches is that DCCC reduces the computational complexity of the problem by removing certain exogenous states, thereby transforming the model into a collection of simpler SCMs, each of which admits a unique solution to the corresponding linear program.

The DCCC as well as previous methods, rely on the empirical data distribution, which must be precisely defined, meaning that the estimated probability of each configuration of the endogenous variables is estimated by maximum likelihood. In this paper we are interested in accounting for uncertainty in the data distribution, rather than relying on point estimates. More precisely, we will assume that the value of the empirical distribution for each configuration of the endogenous variables is given by a hard interval. Such intervals can be regarded both from a frequentist or from a Bayesian point of view. In the frequentist case, the interval is just a traditional confidence interval, while in the Bayesian case, one could assume a Dirichlet prior over the probabilities of the exogenous variables that, combined with a multinomial likelihood, results in a Dirichlet posterior. Either way the complexity of solving counterfactual queries grows with respect to the case in which the empirical distribution is defined in terms of point estimates. More precisely, the answer to counterfactual queries in the point case is computed by exploring the set of distributions of the exogenous variables that are compatible with the data, which amounts to guaranteeing that the marginal over the endogenous variables matches the empirical distribution. In the interval case, the set of compatible distributions is much larger, since it includes all those

whose marginal over the endogenous variables is compatible with the intervals specified by the empirical distribution.

2. Background

Random variables are denoted by uppercase letters and lower-case is used for their possible values (or states), while Ω_V denotes the state space of variable V. Similarly, $\mathbf{V} = \{V_1, V_2, \ldots, V_n\}$ denotes a set of variables and $\mathbf{v}$ a joint state of its domain $\Omega_{\mathbf{V}} = \times_{V \in \mathbf{V}} \Omega_V$. Variables are omitted from assignments when their context is clear, and we may for instance simply denote $P(V = v)$ as $P(v)$. Finally, $\tilde{P}$ is used to denote empirical distributions.

Structural causal models (SCMs) (Pearl, 2009) are a class of PGMs used for causal and counterfactual reasoning, consisting of two types of nodes: *endogenous* nodes, which represent the internal variables of the modelled problem, and *exogenous* nodes, which represent factors outside the model. Formally, an SCM $\mathcal{M}$ is a 4-tuple $\langle \mathbf{U}, \mathbf{V}, \mathcal{F}, \mathcal{P} \rangle$, where (Bareinboim et al., 2022)

- $\mathbf{U}$ is a set of exogenous variables that are determined by factors outside the model,
- $\mathbf{V}$ is a set of variables $\{V_1, V_2, \ldots, V_n\}$, called endogenous, that are determined by other (exogenous and endogenous) variables in the model, i.e. by variables in $\mathbf{U} \cup \mathbf{V}$,
- $\mathcal{F}$ is a set of functions $\{f_{V_1}, f_{V_2}, ..., f_{V_n}\}$ called *structural equations* (SE), such that each of them is a function $f_{V_i} : \Omega_{\mathrm{Pa}_{V_i}} \to \Omega_{V_i}$, where Pa_{V_i} are the variables directly determining V_i, including the exogenous parent $U_i \subseteq \mathbf{U}$,
- $\mathcal{P}$ is a set of probability distributions $P(U)$ for each $U \in \mathbf{U}$.

The structural equations $\mathcal{F}$ define a directed acyclic graph (DAG) $\mathcal{G}$, referred to as the *causal graph* of the model, whose nodes correspond to the variables in $\mathbf{U} \cup \mathbf{V}$. All endogenous and exogenous variables are assumed to have finite, discrete domains. Note that a discrete exogenous variable, taking values on a finite set of unobserved states, is enough to represent all counterfactual distributions (Zhang et al., 2022). When the distributions for the exogenous variables are not known, we refer to the model as a *partially* specified SCM, denoted by a calligraphic letter, such as $\mathcal{M}$. In contrast, if all such distributions are provided, the model is considered fully specified and is denoted as, e.g., M.

An SCM in which each exogenous variable influences exactly one endogenous variable is called *Markovian* (Avin et al., 2005). It is important to note that each endogenous variable must have exactly one exogenous parent. The cardinality of each exogenous variable U is at most $|\Omega_Y|^{|\Omega_{\mathbf{X}}|}$, where Y denotes its only endogenous child and $\mathbf{X}$ the set of endogenous parents of Y.

Inference in a partial SCM given a dataset $\mathcal{D}$ consists of estimating a set of fully-specified SCMs that are compatible with the data (Bjøru et al., 2024). SCMs can be represented as *credal networks* (Zaffalon et al., 2020, 2024) which are Bayesian networks with imprecise conditional distributions (Cozman, 2000). In credal networks, each node is associated with a set of probability mass functions known as a *credal set*. When transforming SCMs into credal networks, the endogenous observations impose linear constraints on the probabilities of the exogenous variables, so that a separate credal set for each exogenous variable can be

obtained, called the *solution set*. The solution set is the convex set of all the distributions over an exogenous variable U leading to the same distribution over the endogenous children after marginalizing out the exogenous parents. Assuming that the empirical distribution provides point probability values, the solution set is defined as

$$\mathcal{K}(U) := \left\{ P(U) : \sum_{u \in \Omega_U} P(u) \cdot P(Y|\mathbf{X}, u) = \tilde{P}(Y|\mathbf{X}) \right\}, \tag{1}$$

where Y is the only child of U, $\mathbf{X}$ the set of endogenous parents of Y and $\tilde{P}(Y|\mathbf{X})$ is the (point) empirical distribution computed from a dataset $\mathcal{D}$. Hence, the goal of inference is to identify all such distributions. We define the solution set for a partially-specified SCM $\mathcal{M}$ given a dataset $\mathcal{D}$, denoted by $\mathcal{S}_{\mathcal{M},\mathcal{D}}$, as the collection of all fully-specified SCMs satisfying $P(U) \in \mathcal{K}(U)$ for every $U \in \mathbf{U}$ (we will sometimes also denote the solution set as $\mathcal{S}_{\mathcal{M},\tilde{P}}$ to emphasize the empirical distribution). Once $\mathcal{S}_{\mathcal{M},\mathcal{D}}$ has been identified, it can be employed to bound any causal or counterfactual query Q by evaluating the query within each model in the set and subsequently aggregating the results.

The DCCC method (Bjøru et al., 2024, 2025) operates by transforming a complex SCM into a set of simpler ones, each with exogenous variables of smaller cardinality. The transformation is done through the so-called *reduction* operation, which involves the removal of some states from the exogenous domains. It is formally defined as follows:

Definition 1 (Reduction operator) *Let $\mathcal{M}$ be a partially-specified SCM whose set of exogenous variables is $\mathbf{U}$ and let $u \in \Omega_U$ with $U \in \mathbf{U}$. Then the reduction operation, denoted $R(\mathcal{M}, u)$, produces a new partially-specified SCM $\mathcal{M}'$ by removing assignments from $\mathcal{F}$ and $\mathcal{P}$ in $\mathcal{M}$ that are consistent with u.*

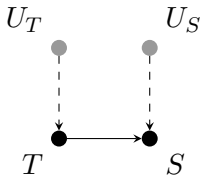
Reduction applies to various states of the different exogenous variables $\mathcal{A}_{\mathbf{U}} := \left\{u^{(i)}\right\}_{i=1}^{m}$ s.t. $u^{(i)} \in \Omega_U$ and $U \in \mathbf{U}$, which can be recursively defined as

$$R(\mathcal{M}, \mathcal{A}_{\mathbf{U}}) = R\left(R(\mathcal{M}, u^{(1)}), \mathcal{A}_{\mathbf{U}} \setminus \{u^{(1)}\}\right). \tag{2}$$

The reduction operation $R(\mathcal{M}, u)$ is equivalent to imposing the constraint $P(u) = 0$ on $\mathcal{M}$. Thus, when looking for the solution set of a reduced SCM, we can equivalently look for all the solutions from the original solution set that are consistent with that constraint.

The answer for any counterfactual query can be bounded after a solution has been found. The bounds are computed by building a fully specified SCM for each extreme point in the solution and computing the answer for that SCM. The lower and upper bounds are then obtained from the minimum and maximum values across all such computations.

The DCCC method for Markovian models proceeds by computing the set $\mathcal{K}(U)$ in Equation (1) for each exogenous variable U in a model. Each variable U has a single endogenous child Y, with endogenous parents $\mathbf{X} = \text{Pa}_Y \setminus \{U\}$. If we let $p = |\Omega_{\mathbf{X}}|$ and $q = |\Omega_Y|$, the canonical domain size of U is $|\Omega_U| = q^p$, which means that the solution space is determined by $q \cdot p$ linear constraints of the form $\sum_{u \in \Omega_U} P(u) \cdot P(Y|\mathbf{X}, u) = \tilde{P}(Y|\mathbf{X})$, of which $p \cdot (q-1)+1$ are independent under $\sum_{u \in \Omega_U} P(u) = 1$. The extreme points in $\mathcal{K}(U)$ are thus distributions $P(U)$ for which at most $p \cdot (q-1) + 1$ of the states $u \in \Omega_U$ have non-zero



f_T

	U_T	
	v_0	v_1
$T=0$	1	0
$T=1$	0	1

$\tilde{P}(T)$	
$T=0$	0.337
$T=1$	0.663

f_S

		U_S			
T	S	u_0	u_1	u_2	u_3
0	0	1	1	0	0
0	1	0	0	1	1
1	0	1	0	1	0
1	1	0	1	0	1

$\tilde{P}(S\|T)$	$T=0$	$T=1$
$S=0$	0.462	0.323
$S=1$	0.538	0.677

Figure 1: Elements of an SCM: (left) causal graph, (center) structural equations and (right) empirical distribution. Adapted from (Bjøru et al., 2025, Figure 1).

probability (Bjøru et al., 2025), and they correspond to reductions R of the original model $\mathcal{M}$ such that the image U' of U under R has domain size at most $|\Omega_{U'}| = p \cdot (q-1) + 1$.

The number of possible reductions is $\binom{q^p}{p\cdot(q-1)+1}$, and for each one of them, DCCC fixes the probability of the states in the complement set to 0, and solves for the unique solution of the resulting linear system. If this solution is a probability distribution such that $P(u) \geq 0, \forall u \in \Omega_U$, $\sum_{u\in\Omega_U} P(u) = 1$, it is retrieved as an extreme point of $\mathcal{K}(U)$. Once all possible reductions are explored, a complete collection of extreme points is retrieved and exact query bounds may be calculated.

3. Problem formulation

As shown by Zaffalon et al. (2020), any partial SCM can be represented by a credal network. Furthermore, all queries in an SCM, counterfactual or not, are linear-fractional in the probabilities allocated to the exogenous variables (Zaffalon et al., 2020, 2024). Therefore, any probabilistic query in the SCM resulting in a (potentially degenerate) interval can be calculated by visiting the vertices defining the probability distribution over the exogenous variables that define the partial SCM. We will therefore consider how to enumerate these vertices. We start the discussion by re-visiting the running example used by Bjøru et al. (2025), see Figure 1. This SCM models a drug study involving 700 patients (Mueller et al., 2022). The causal graph on the left includes the endogenous variables $\mathbf{V} = \{T, S\}$, representing the *treatment* and *survival*, respectively. The goal is to analyze whether receiving treatment ($T = 1$) contributes to survival ($S = 1$). Meanwhile, $\mathbf{U} = \{U_T, U_S\}$ represents the set of exogenous variables, which are assumed to be root nodes, with endogenous variables as their children. Next, the center of the figure shows the SEs $f_T(U_T)$ and $f_S(T, U_S)$ as deterministic CPTs of the form $\tilde{P}(T|U_T)$ and $\tilde{P}(S|T, U_S)$, characterized by containing only ones and zeros. If not provided from expert knowledge, SEs can be automatically inferred from the causal graph, without any loss of generality, via a *canonical specification*. This is the case of the SEs shown in the example, where the states of an exogenous variable will then represent all possible deterministic mechanisms between its children and their respective endogenous parents. Conversely, under a non-canonical specification, some of the exogenous states are assumed to be impossible and directly omitted from the CPTs. In a Markovian model, the cardinality of each exogenous variable U is at most $|\Omega_Y|^{|\Omega_{\mathbf{X}}|}$, where

Y is its only child and $\mathbf{X}$ the set of endogenous parents of Y. For the canonical models we consider on in this paper, we have that $|\Omega_U| = |\Omega_Y|^{|\Omega_\mathbf{X}|}$.

We read these tables as follows: Consider the table for f_S, and as an example, focus on the column for which $U_S = u_1$. It has a 1 in the rows $\{T = 0, S = 0\}$ and $\{T = 1, S = 1\}$. This means that in the event that $U_S = u_1$, $S = 0$ whenever $T = 0$ and $S = 1$ if $T = 1$; both events with probability 1. In other words, u_1 represents the structural equation "$S = 1$ if and only if $T = 1$". Finally, the right part of the figure gives the empirical distributions $\tilde{P}(T)$ and $\tilde{P}(S|T)$. Bjøru et al. (2025) showed that $\tilde{P}(T)$ and $\tilde{P}(S|T)$ resulted in a solution that is a *partial* SCM. For variable U_T they found that $\mathcal{K}(U_T) = \{P(U_T) = [0.337, 0.663]\}$. They defined $P_1(U_S) = [0.323,\ 0.139, 0,\ 0.538]$ and $P_2(U_S) = [0, 0.462,\ 0.323,\ 0.215]$ and showed that solutions for U_S are the convex combination of $P_1(U_S)$ and $P_2(U_S)$, $\mathcal{K}(U_S) = \{P(U_S) = \lambda P_1(U_S) + (1 - \lambda)P_2(U_S);\ \lambda \in [0, 1]\}$.

Let us now extend this example by using imprecise descriptions of the empirical data distributions in the form of intervals. Such imprecise descriptions can be defined either from a frequentist point of view as confidence intervals or from a Bayesian point of view as credibility intervals. In the latter case, a credibility interval may be derived by assuming a (conjugate) prior Dirichlet distribution over the model parameters. Let us first consider only $\tilde{P}(T)$ to have an imprecise formulation, e.g., using the point-wise restrictions

$$\begin{aligned}\tilde{P}^0_\perp \leq \tilde{P}(T = 0) \leq \tilde{P}^0_\top,\\ \tilde{P}^1_\perp \leq \tilde{P}(T = 1) \leq \tilde{P}^1_\top.\end{aligned}$$

It immediately follows that a solution is viable if and only if

$$\begin{aligned}\max\{\tilde{P}^0_\perp, 1 - \tilde{P}^1_\top\} \leq P(U_T = v_0) \leq \min\{1 - \tilde{P}^1_\perp, \tilde{P}^0_\top\},\\ P(U_T = v_1) = 1 - P(U_T = v_0).\end{aligned}$$

As long as the solution space is not degenerate, this is a partial SCM that therefore can be represented as a credal network.

A more interesting situation arises if $\tilde{P}(S|T)$ is imprecise. Consider the effect of relaxing the empirical distributions from the original example ($\tilde{P}(S = 0|T = 0) = .462$, $\tilde{P}(S = 0|T = 1) = .323$) with the following specification:

$$\begin{aligned}0.4 \leq \tilde{P}(S = 0|T = 0) \leq 0.5,\\ 0.3 \leq \tilde{P}(S = 0|T = 1) \leq 0.35.\end{aligned} \tag{3}$$

In this case the solution space is defined by the system of linear inequalities

$$\begin{aligned}0.4 \leq \sum_{u \in \Omega_{U_S}} P(u) \cdot P(S = 0|T = 0, u) \leq 0.5,\\ 0.3 \leq \sum_{u \in \Omega_{U_S}} P(u) \cdot P(S = 0|T = 1, u) \leq 0.35,\end{aligned} \tag{4}$$

together with the equality constraint $\sum_{u\in\Omega_{U_S}} P(u) = 1$. These inequalities define a convex set determined by the convex hull over the eight extreme points:

$$\begin{array}{ll} P(U_S = u_0): & \begin{bmatrix} 0 \\ 0.5 \\ 0.35 \\ 0.15 \end{bmatrix}, \begin{bmatrix} 0.35 \\ 0.15 \\ 0 \\ 0.5 \end{bmatrix}, \begin{bmatrix} 0 \\ 0.5 \\ 0.3 \\ 0.2 \end{bmatrix}, \begin{bmatrix} 0.3 \\ 0.2 \\ 0 \\ 0.5 \end{bmatrix}, \begin{bmatrix} 0 \\ 0.4 \\ 0.35 \\ 0.25 \end{bmatrix}, \begin{bmatrix} 0.35 \\ 0.05 \\ 0 \\ 0.6 \end{bmatrix}, \begin{bmatrix} 0 \\ 0.4 \\ 0.3 \\ 0.3 \end{bmatrix}, \begin{bmatrix} 0.3 \\ 0.1 \\ 0 \\ 0.6 \end{bmatrix}. \\ P(U_S = u_1): & \\ P(U_S = u_2): & \\ P(U_S = u_3): & \end{array}$$

More generally, for model $\mathcal{M}$, the linear inequality constraints

$$\tilde{P}_{\perp}(Y|\mathbf{X}) \leq \sum_{u\in\Omega_U} P(u)\cdot P(Y|\mathbf{X},u) \leq \tilde{P}_{\top}(Y|\mathbf{X}),$$

together with the equality constraint $\sum_{u\in\Omega_U} P(u) = 1$ define the solution space for $\mathcal{M}$:

$$\mathcal{K}(U) := \left\{ P(U) : \tilde{P}_{\perp}(Y|\mathbf{X}) \leq \sum_{u\in\Omega_U} P(u)\cdot P(Y|\mathbf{X},u) \leq \tilde{P}_{\top}(Y|\mathbf{X}) \right\}, \tag{5}$$

which can in principle be solved using any standard LP solver. Similar to the case where we have a fixed empirical distribution over the endogeneous variables, any interventional or counterfactual query can now be bounded by evaluating the query over all the extreme points defining the convex solution set $\mathcal{K}(U)$ and aggregating the results. The complexity of this brute force approach is directly determined by the LP solver after deploying the Charnes-Cooper substitution (Charnes and Cooper, 1962). We rely on a variant of the simplex algorithm , see, e.g., Murty (1983). The algorithm's complexity per pivot is polynomial in the number of inequalities and variables, while the number of pivots has worst-case complexity exponential in the number of variables. To address this complexity issue, we will in the next section consider an alternative solution method, where we adapt the divide-and-conquer algorithm for bounding counterfactual probabilities (Bjøru et al., 2025) to the setting with imprecise empirical distributions.

4. D3CID: DCCC with Imprecise Data distributions

The Divide and Conquer for Causal Computation (DCCC) method (Bjøru et al., 2025) reduces computational complexity by transforming the model into a collection of simpler SCMs by systematically removing exogeneous states through the reduction operation (see Definition 1). This operation is complete in the sense that the full solution set is defined by the solution sets produced by the reductions. A similar result can be derived for the case with imprecise empirical distributions.

Theorem 2 *Let $\mathcal{M}$ be a partially-specified Markovian SCM with binary endogeneous variables. Consider an endogenous child variable Y with n endogeneous parent variables and an exogeneous variable U. If $(\tilde{P}_{\perp}^{\mathcal{M}}, \tilde{P}_{\top}^{\mathcal{M}})$ define an imprecise empirical distribution over Y, then*

$$\mathcal{K}(U) = \texttt{conv_hull}\left(\bigcup_{U'\in\mathcal{R}(\mathcal{M})} \mathcal{K}(U') \right),$$

where $\mathcal{R}(\mathcal{M})$ is the set of all possible reductions over $\mathcal{M}$ s.t. $|\Omega_{U'}| \leq 2^n + 1$.

Proof If $P(U) \in \mathcal{K}(U)$, then there exists a $\tilde{P}$ that elementwise resides within the bounds $(\tilde{P}_{\perp}^{\mathcal{M}}, \tilde{P}_{\top}^{\mathcal{M}})$ s.t. $P(U) \in \mathcal{S}_{\mathcal{M},\tilde{P}}$. If $P(U)$ is extreme, then according to (Bjøru et al., 2025, Theorem 4), there exists a reduction $U' = R(\mathcal{M}, \mathcal{A}_U)$, so that $P(U)$ is a (single) model in $\mathcal{S}_{R(\mathcal{M},\mathcal{A}_U),\tilde{P}}$. Since $\tilde{P}$ is within the bounds $(\tilde{P}_{\perp}^{\mathcal{M}}, \tilde{P}_{\top}^{\mathcal{M}})$, we have that $P(U) \in \mathcal{S}_{R(\mathcal{M},\mathcal{A}_U),\tilde{P}} \subseteq \mathcal{S}_{R(\mathcal{M},\mathcal{A}_U),(\tilde{P}_{\perp}^{\mathcal{M}},\tilde{P}_{\top}^{\mathcal{M}})} = \mathcal{K}(U')$. According to Bjøru et al. (2025)[Theorem 5], $|\Omega_{U'}| \leq 2^n + 1$.

On the other hand, assume that $P(U)$ defines an internal model. Then (by definition) $P(U)$ is in the convex hull of a collection of fully-specified extreme models $\{M_i\} \subseteq \mathcal{S}_{\mathcal{M},\tilde{P}}$, and for each of these extreme models M_i, there exists a reduction $U_i' = R(\mathcal{M}, \mathcal{A}_U)$. Focusing on the convex hull of the solution set defined by each of the reductions, the remainder of the proof then follows from the arguments above.

The proof of the reverse direction is trivial as each $\mathcal{K}(U') \subseteq \mathcal{K}(U)$.

■

For illustration, consider again the model $\mathcal{M}$ in Figure 1, where the empirical distributions is relaxed for $\tilde{P}(S = 0|T = 0)$ and $\tilde{P}(S = 0|T = 1)$:

$$\begin{aligned} 0.4 \leq \tilde{P}(S = 0|T = 0) \leq 0.5, \\ 0.3 \leq \tilde{P}(S = 0|T = 1) \leq 0.35. \end{aligned} \tag{6}$$

There are four possible reductions $R(\mathcal{M}, u_i)$, $i \in \{0, 1, 2, 3\}$, for variable U in $\mathcal{M}$, but only the two reductions $R(\mathcal{M}, u_0)$ and $R(\mathcal{M}, u_2)$ define valid solution sets, $P_0(U_S)$ and $P_2(U_S)$; for the remaining two reductions, the solutions sets are empty. For example, $P_0(U_S)$ is defined by four vertices $P_0(U_S) = \texttt{conv_hull}\{[0, 0.4, 0.3, 0.3], [0, 0.5, 0.3, 0.2], [0, 0.4, 0.35, 0.25], [0, 0.5, 0.35, 0.15]\}$ as illustrated in Figure 2 (left).

Note that the (imprecise) empirical distribution may give rise to additional constraints $P(U = u_i) = 0$ beyond those identified by the reduction procedure. For example, if we relax the empirical data distribution further:

$$\begin{aligned} 0.3 \leq \tilde{P}(S = 0|T = 0) \leq 0.6 \\ 0.2 \leq \tilde{P}(S = 0|T = 1) \leq 0.45, \end{aligned} \tag{7}$$

we obtain the results for $P_0(U_S)$ shown in Figure 2 (right). Notice that the vertex-set now is of size 5. The reason is that the solution contains vertices from *two* reductions. We have already considered the reduction $P(U_S = u_0) = 0$, but two of the vertices in the vertex-set now also includes the extra constraint $P(U_S = u_3) = 0$. This extended reduction correspond to the following set of inequalities

$$\begin{aligned} 0.3 \leq \quad & P(U_S = u_1) & \leq 0.6 \\ 0.2 \leq \quad & P(U_S = u_2) & \leq 0.45 \\ 1 \leq \quad & P(U_S = u_1) + P(U_S = u_2) & \leq 1, \end{aligned}$$

resulting in the two vertices $\{[0, 0.6, 0.4, 0], [0, 0.55, 0.45, 0]\}$.

Based on Theorem 2, we can define a general algorithm (Algorithm 1) for bounding a counterfactual query through model reductions and by refining the bounds of the counterfactual query for each reduction considered. This general algorithmic formulation can be

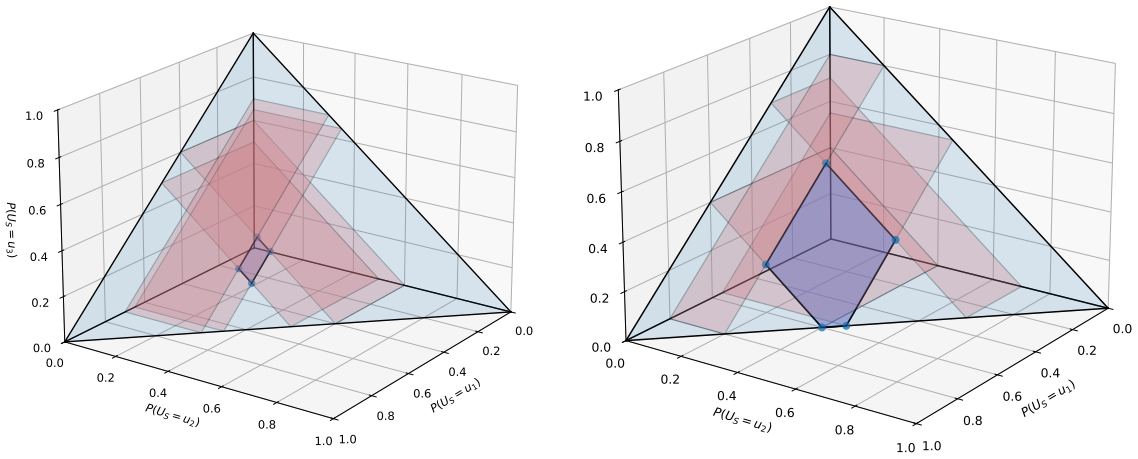


Figure 2: Left: The solution for $P_0(U_S)$. Note that since $P(U_S = u_0) = 0$ for all vertices it is omitted from the plot. Observe how the solution is defined by the half-spaces induced by the imprecise probabilities. Right: The same result, but with the empirical distribution constrained by Equation (7).

instantiated with different reduction strategies. First, Algorithm 1 can, without modification, be seen as an anytime algorithm. After each iteration of the outer loop, $\mathcal{I}$ holds an *inner approximation* that is always contained within the true interval as the bounds on the query are never tightened. Secondly, we can use heuristics to prioritize apparently informative reduction candidates. One possibility in that respect is to follow Bjøru et al. (2025), who used a heuristic to favor low-probability clauses and sparse extreme configurations early in their loop, with the goal of finding promising models early.

As noted in Section 3, the worst-case complexity of finding the solution space $\mathcal{K}(U)$ for an exogenous parent U of a binary variable with n endogenous parents is exponential in the number of variables. Hence, solving the reduced problem (with the number of variables exponential in n, not doubly exponential in n) is a considerable computational simplification. We note that the number of possible reductions is still $\binom{2^{2^n}}{2^n+1}$, but the contribution from this part of the total computational complexity can be controlled by an anytime instantiation of the algorithm as briefly mentioned above.

5. Experiments

For the validation study[1], we use a benchmark consisting of 80 randomly generated SCMs based on the two inverted-tree topologies depicted in Figure 3. All the considered models are Markovian, implying a one-to-one relationship between exogenous and endogenous variables. The endogenous variables are binary. In particular, Y is the child variable, while $\mathbf{X} := \{X_1, X_2\}$ denotes the set of endogenous parent variables. Concerning the exogenous variables, each model is initially specified in canonical form. To obtain models with different levels of complexity, some structural equations are transformed into non-canonical ones by randomly eliminating selected states of the exogenous variables. Once the exogenous

1. Code: https://github.com/PGM-Lab/2026-PGM-data-uncertainty

Input: Counterfactual query Q
Output: $\mathcal{I}$: smallest interval such that $P(Q) \in \mathcal{I}$ is guaranteed

$\mathcal{I} \leftarrow \emptyset$

repeat

Get next reduction candidate $\mathcal{A}_U$

Bound Q for the reduced problem $R(\mathcal{M}, \mathcal{A}_U)$ via the LP solver:

$$Q_{R(\mathcal{M}, \mathcal{A}_U), (\tilde{P}_\perp, \tilde{P}_\top)} = \left[\min_{M \in \mathcal{S}_{R(\mathcal{M}, \mathcal{A}_U), (\tilde{P}_\perp, \tilde{P}_\top)}} Q_M, \max_{M \in \mathcal{S}_{R(\mathcal{M}, \mathcal{A}_U), (\tilde{P}_\perp, \tilde{P}_\top)}} Q_M \right]$$

Extend $\mathcal{I}$ with new results:

$$\mathcal{I} \leftarrow \mathcal{I} \cup Q_{R(\mathcal{M}, \mathcal{A}_U), (\tilde{P}_\perp, \tilde{P}_\top)}$$

until candidate list exhausted **or** $\mathcal{I} = [0, 1]$;

return $\mathcal{I}$

Algorithm 1: The D3CID Algorithm. Different reduction strategies can be employed, allowing for both an exhaustive search of the solutions space as well as anytime solutions.

distributions are randomly initialized, a dataset of 1000 instances is generated from each model. These models were selected for two reasons: learning is performed independently for each exogenous variable, which avoids an increase in computational complexity for larger models with the same maximum in-degree, and they allow exact computation of the bounds for all queries. Note that scalability is limited by the size of the state space of the exogenous variables, which grows exponentially on the cardinality of the joint state space of the endogenous parents. Uncertainty in the data is simulated by introducing a perturbation in

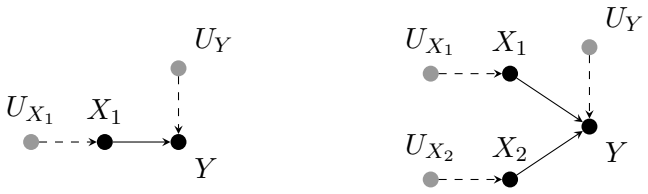


Figure 3: Graphs of the SCMs considered in the experimentation.

the empirical distribution associated with Y, controlled by a parameter ϵ. Specifically, for each value $(y, \mathbf{x}) \in \Omega_Y \times \Omega_{\mathbf{X}}$, the lower bound is defined as $\tilde{P}_\perp(y \mid \mathbf{x}) = \tilde{P}(y \mid \mathbf{x}) - \epsilon$, while the upper bound is given by $\tilde{P}_\top(y \mid \mathbf{x}) = \tilde{P}(y \mid \mathbf{x}) + \epsilon$. After applying the perturbation, the intervals are adjusted, if necessary, to satisfy the reachability condition (De Campos et al., 1994).

In this experimentation, two algorithms are considered. First, the LPID approach applies linear programming without reductions and under imprecise descriptions of the empirical distribution, as described in Section 3. Second, DC3ID corresponds to the method

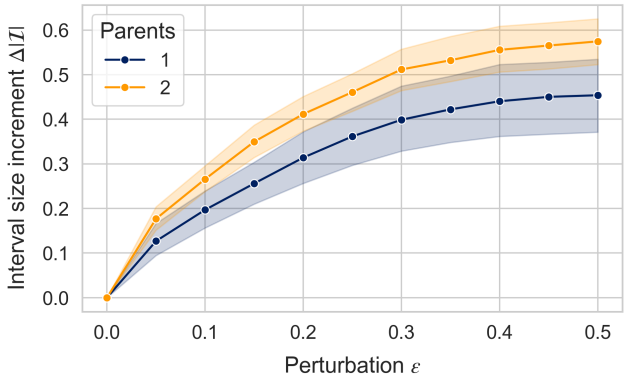


Figure 4: Interval size increment $\Delta|\mathcal{I}|$ against an increasing perturbation value ϵ for models with one and two parent nodes. The shadow represents the 95% confidence interval.

proposed in Section 4, which adapts the divide-and-conquer approach to handle imprecise descriptions of the empirical distribution. For each model and method, the *probability of necessity* and the *probability of sufficiency* (Pearl, 2009) of X_1 on Y, as well as those of X_2 on Y, are computed.

The first objective is to analyze how the level of uncertainty affects inference. To do this, we compare the resulting interval sizes of the queries using LPID at various perturbation levels against the baseline case without uncertainty (i.e., LPID with $\epsilon = 0$). If we let $|\mathcal{I}^{LPID_\epsilon}|$ and $|\mathcal{I}^{LPID_0}|$ denote the interval sizes obtained with a perturbation of ϵ and $\epsilon = 0$, respectively, we can define the size increment as $\Delta|\mathcal{I}| = |\mathcal{I}^{LPID_\epsilon}| - |\mathcal{I}^{LPID_0}|$. Figure 4 shows how this increment varies as the perturbation ϵ increases. It can be observed that the increment is approximately double the perturbation; this is consistent with the fact that the perturbation is applied to both bounds. Moreover, the increment is higher in the model with two parents than in the model with only one.

The second objective is to analyze the behavior of DC3ID as an approximate method for bounding counterfactual queries with imprecise descriptions of the empirical distributions. To do this, we analyze the root mean squared error (RMSE) of DC3ID with respect to LPID, defined as:

$$\text{RMSE} = \sqrt{\frac{(a - a^*)^2 + (b - b^*)^2}{2}}$$

where $[a^*, b^*]$ denotes the exact interval and $[a, b]$ is the approximation. Figure 5 (left) displays the RMSE as a function of the number of solutions generated by DC3ID. Here, a generated solution refers to a fully-specified SCM that can be used to calculate a precise value for the query, which is then aggregated to form the final intervals. Recall that DC3ID is an anytime algorithm, meaning that it does not require calculating all solutions to approximate the inference. Furthermore, it can be observed that with only a small number of solutions, the error converges to zero. Secondly, Figure 5 (right) shows the speed-up of DC3ID with respect to LPID for the topology with one parent node (in blue) and two parents (in orange). The dashed red line depicts a speed-up equal to one, which corresponds to the threshold beyond which LPID becomes more efficient. It can be observed that for the model with two parents, there is a clear improvement in efficiency. Conversely, for the

one-parent case, the difference is less pronounced, though DC3ID remains more efficient than LPID. Because LPID already performs efficiently in one-parent topologies, the margin for improvement is smaller

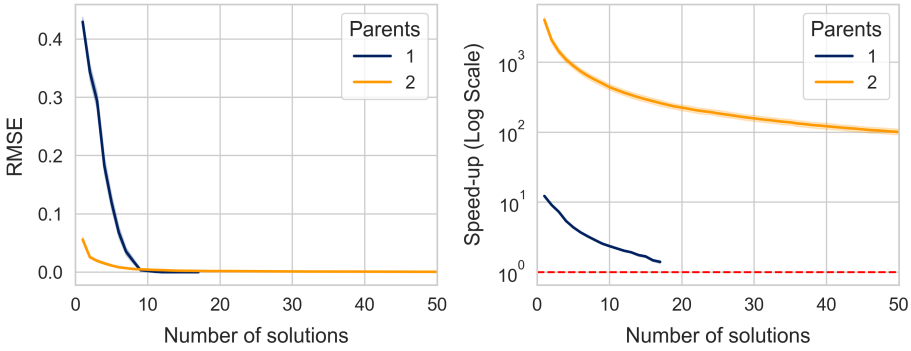


Figure 5: RMSE (left) and speed-up (right) of DC3ID with respect to LPID for models with one and two parent nodes.

6. Conclusion

In this paper we propose a method for efficient calculation of bounds for counterfactual probabilities, that is able to handle interval-valued empirical distributions and thus accounts for data uncertainty. We formulated a linear programming problem under this setting, that utilized the divide-and-conquer algorithm for bounding counterfactual probabilities proposed by Bjøru et al. (2025). Initial experiments show substantial speed-up over brute-force methods with only minor loss of precision.

There are several avenues of future work that we want to pursue. First, the experiments should be extended to more realistic settings. Secondly, we want to improve the ordering of the potential reductions presented in Algorithm 1. Bjøru et al. (2025) ordered reductions heuristically based on empirical distributions, and similar ideas may prove beneficial also in our setting. Next, we would like to investigate how expert knowledge can be used to shape the model's behavior. As an example, consider the model in Figure 1, where $U_S = u_1$ maps to the structural equation "$S = T$" and $U_S = u_2$ corresponds to "$S = \neg T$". Therefore, a domain expert wishing to enforce that the treatment is not harmful on average can express this as $P(U_S = u_2) \leq P(U_S = u_1)$. While it is straightforward to extend D3CID with such constraints, we want to investigate their effect and how they can be combined with the heuristics mentioned above. Finally, we note that the approximate intervals produced by the anytime aspect of the proposed method are *inner approximations*, as they are always contained within the true interval. In a sense, they can be seen as "optimistic" approximations of the true result. Outer approximations are, on the other hand, conservative: they can be used to make claims like that "the counterfactual probability is at most α". It would be very valuable to devise an anytime algorithm for outer approximation of counterfactual queries.

Acknowledgments

Authors are supported by projects PID2022-139293NB-C31 and PID2025-172960NB-C33 funded by MICIU/AEI/10.13039/501100011033 and by ERDF "A way of making Europe" and by "AURA - Avansert Utvikling av KI for Regional Akselerasjon".

References

Chen Avin, Ilya Shpitser, and Judea Pearl. Identifiability of path-specific effects. In *International Joint Conference on Artificial Intelligence*, pages 357–363, 2005.

Elias Bareinboim, Juan D. Correa, Duligur Ibeling, and Thomas Icard. On Pearl's hierarchy and the foundations of causal inference. In *Probabilistic and causal inference: The works of Judea Pearl*, pages 507–556. ACM, 2022.

Anna Rodum Bjøru, Rafael Cabañas, Helge Langseth, and Antonio Salmerón. A divide and conquer approach for solving structural causal models. In Johan Kwisthout and Silja Renooij, editors, *Proceedings of The 12th International Conference on Probabilistic Graphical Models*, volume 246 of *Proceedings of Machine Learning Research*, pages 348–360, 2024.

Anna Rodum Bjøru, Rafael Cabañas, Helge Langseth, and Antonio Salmerón. Divide and conquer for causal computation. *International Journal of Approximate Reasoning*, 186, 2025. Paper no. 109520.

Abraham Charnes and William W. Cooper. Programming with linear fractional functionals. *Naval Research Logistics Quarterly*, 9(3-4):181–186, 1962.

Fabio G. Cozman. Credal networks. *Artificial intelligence*, 120(2):199–233, 2000.

Luis M. De Campos, Juan F. Huete, and Serafin Moral. Probability intervals: a tool for uncertain reasoning. *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, 2(02):167–196, 1994.

Scott Mueller, Ang Li, and Judea Pearl. Causes of effects: Learning individual responses from population data. In *Proceedings of the Thirty-First International Joint Conference on Artificial Intelligence*, pages 2712–2718, 2022.

Katta G Murty. *Linear programming.* Chichester, 1983. ISBN 047109725X.

Judea Pearl. *Causality. Models, inference and reasoning. Second edition.* Cambridge University Press, New York, 2009.

Judea Pearl and Dana Mackenzie. *The Book of Why: The New Science of Cause and Effect.* Basic Books, Inc., 2018.

Marco Zaffalon, Alessandro Antonucci, and Rafael Cabañas. Structural causal models are (solvable by) credal networks. In *International Conference on Probabilistic Graphical Models*, pages 581–592. PMLR, 2020.

Marco Zaffalon, Alessandro Antonucci, Rafael Cabañas, David Huber, and Dario Azzimonti. Efficient computation of counterfactual bounds. *International Journal of Approximate Reasoning*, 2024. Paper no. 109111.

Junzhe Zhang, Jin Tian, and Elias Bareinboim. Partial counterfactual identification from observational and experimental data. In *International Conference on Machine Learning*, pages 26548–26558. PMLR, 2022.