Deriving Bounds and Inequality Constraints Using Logical Relations Among Counterfactuals

Noam Finkelstein
Department of Computer Science
Johns Hopkins University
Baltimore, MD

Ilya Shpitser
Department Computer Science
Johns Hopkins University
Baltimore, MD

Abstract

Causal parameters may not be point identified in the presence of unobserved confounding. However, information about non-identified parameters, in the form of bounds, may still be recovered from the observed data in some cases. We develop a new general method for obtaining bounds on causal parameters using rules of probability and restrictions on counterfactuals implied by causal graphical models. We additionally provide inequality constraints on functionals of the observed data law implied by such causal models. Our approach is motivated by the observation that logical relations between identified and non-identified counterfactual events often yield information about non-identified events. We show that this approach is powerful enough to recover known sharp bounds and tight inequality constraints, and to derive novel bounds and constraints.

1 INTRODUCTION

Directed acyclic graphs (DAGs) are commonly used to represent causal relationships between random variables, with a directed edge from A to Y ($A \rightarrow Y$) representing that A "directly causes" Y. Under the interventionist view of causality, this relationship is taken to mean that Y may change if any set of variables $\mathbf{S}$ that includes A is set, possibly contrary to fact, to values $\mathbf{s}$. The operation that counterfactually sets values of variables is known as an *intervention* and has been denoted by the $do(\mathbf{s})$ operator in [11].

The variable Y after an intervention $do(a)$ is performed is denoted $Y(a)$, and is referred to as a potential outcome, or a counterfactual random variable [9]. Distributions over counterfactuals such as $P(Y(a))$ may be used to quantify cause-effect relationships by means of a hypothetical randomized controlled trial (RCT). For example, the *average causal effect (ACE)* is defined as $\mathbb{E}[Y(a)] - \mathbb{E}[Y(a')]$ and is a comparison of means in two arms of a hypothetical RCT, with the arms defined by $do(a)$ and $do(a')$ operations.

Since counterfactuals are not observed directly in the data, assumptions are needed to link counterfactual parameters with the observed data distribution. These assumptions are provided by causal models (often represented by DAGs), which are substantively justified using background knowledge or learned directly from data [18].

Under some causal models, counterfactual distributions $P(Y(a))$ may be identified exactly (expressed as functionals of the observed data distribution) [19, 17, 5]. However, when causally relevant variables are not observed, counterfactual distributions may not be identified.

The ideal approach for dealing with non-identified parameters is additional data collection that would "expand" the observed data distribution by rendering previously unobserved variables observable, and thus a previously non-identified parameter identifiable.

If additional data collection is not possible, the alternative is to impose additional parametric assumptions on the causal model which would imply identification, or retreat to a weaker notion of identification, where the observed data distribution is used to obtain *bounds* on the non-identified parameter of interest. The existence of such bounds may yield substantively significant conclusions, for instance by indicating that the causal effect is present (if the corresponding parameter is bounded away from 0).

A well known example of a causal model with a non-identified causal parameter with non-trivial bounds is the instrumental variable (IV) model [1, 8, 16]. The original sharp bounds for the causal parameter in the IV model were derived using a computationally intensive, and difficult to interpret, convex polytope vertex enumeration

approach [1]. Subsequent work [12] has extended this approach to other scenarios where a counterfactual objective can be expressed as a linear function of the observed data law. These approaches are limited both by computational complexity and by the required linear form of the objective.

Bounds on causal parameters are related to inequality constraints on the observed data law implied by hidden variable DAGs, as demonstrated by the derivation of the original IV inequalities [1], and subsequent work on inequality constraints [3, 6]. The approach developed in [21] for deriving such inequality constraints is very general and is conjectured to be able to recover all constraints implied by a hidden variable DAG, but is computationally challenging to evaluate, and has no bounded running time.

In this paper, we present a new approach for deriving bounds on non-identified causal parameters that directly uses restrictions implied by a causal model, rules of probability theory, and logical relations between identified and non-identified counterfactual events. We then build on this approach to present a new class of inequality constraints on the observed data law.

The paper is organized as follows. We introduce notation and relevant concepts in Section 2. We provide an intuitive introduction to our method by re-deriving known sharp bounds in the binary IV model [1] in Section 3. In Section 4, we present results important to our approach, and provide a general algorithm for obtaining bounds on causal parameters. Section 5 demonstrates how our approach may be used to derive generalized instrumental variable inequalities, of which the original IV inequalities and Bonet's inequalities [2] are special cases. Finally, in Section 6 we make use of these results to provide novel bounds and inequality constraints for two sample models.

2 PRELIMINARIES

We let $\mathcal{G}$ denote a DAG with a vertex set $\mathbf{V}$ such that each element of $\mathbf{V}$ corresponds to a random variable. The *statistical* model of $\mathcal{G}$ is the set of joint distributions $P(\mathbf{V})$ that are Markov relative to the DAG $\mathcal{G}$. Specifically, it's the set $\{P(\mathbf{V}) : P(\mathbf{V}) = \prod_{V \in \mathbf{V}} P(V | \mathrm{pa}_{\mathcal{G}}(V))\}$, where $\mathrm{pa}_{\mathcal{G}}(V)$ is the set of parents of V in $\mathcal{G}$.

The *causal* model of a DAG is also a set of joint distributions, but over counterfactual random variables. A counterfactual $Y(a)$ denotes the random variable Y in a counterfactual world, where A is exogenously set to the value a.

Such a causal model can be described by a set of structural equations $\{f_V(\mathrm{pa}_{\mathcal{G}}(V), \epsilon_V) \mid V \in \mathbf{V}\}$, where each f_V can be thought of as a causal mechanism that maps values

of $\mathrm{pa}_{\mathcal{G}}(V)$ (parents of V) and the exogenous noise term ϵ_V to a value of V. For a given set of values $\mathbf{a}$ of $\mathrm{pa}_{\mathcal{G}}(V)$, variation of ϵ_V yields the counterfactual random variable $V(\mathbf{a})$ as the output of $f_V(\mathbf{a}, \epsilon_V)$.

Other counterfactuals can be defined through recursive substitution [14], as follows:

$$Y(\mathbf{a}) = \begin{cases} \mathbf{a}_Y & \text{if } Y \in \mathbf{A} \\ f_Y(\{V(\mathbf{a}) \mid V \in \mathrm{pa}_{\mathcal{G}}(Y)\}, \epsilon_Y) & \text{otherwise} \end{cases} \quad (1)$$

This definition, following from the structural equation model view of the causal model of a DAG, allows the effects of exogenous intervention to propagate downstream to the outcome of interest. Under this view, only the noise variables in the set $\{\epsilon_V : V \in \mathbf{V}\}$ are random. The distributions of the observed data and of counterfactual random variables can be thought of as the distributions of different functions of $\{\epsilon_V : V \in \mathbf{V}\}$ as described in (1).

As a notational convention, given any set $\{Y_1, \ldots, Y_k\} \equiv \mathbf{Y}$, and a set of treatments $\mathbf{A}$ set to $\mathbf{a}$, we will denote a set of counterfactuals $\{Y_1(\mathbf{a}), \ldots, Y_k(\mathbf{a})\}$ defined by (1) by the shorthand $\mathbf{Y}(\mathbf{a})$. We will sometimes denote single variable events $Y(\mathbf{a}) = y$ via the shorthand $y(\mathbf{a})$ for conciseness, similarly multivariable events $\mathbf{Y}(\mathbf{a}) = \mathbf{y}$ will sometimes be denoted as $\mathbf{y}(\mathbf{a})$.

One consequence of (1) is that some counterfactuals $Y(\mathbf{a})$ only depend on a subset of values in $\mathbf{a}$, specifically those values that make an appearance in one of the base cases of the definition. Restrictions of this sort are sometimes called *exclusion restrictions*.

The recursive substitution definition above implies the following *generalized consistency property*, which states that for any disjoint subsets $\mathbf{A}, \mathbf{B}, \mathbf{Y}$ of $\mathbf{V}$,

$$\mathbf{B}(\mathbf{a}) = \mathbf{b} \text{ implies } \mathbf{Y}(\mathbf{a}, \mathbf{b}) = \mathbf{Y}(\mathbf{a}). \quad (2)$$

If all variables $\mathbf{V}$ in a causal model represented by a DAG $\mathcal{G}$ are observed, every interventional distribution $P(\mathbf{Y}(\mathbf{a}))$, where $\mathbf{A} \subseteq \mathbf{V}$, $\mathbf{Y} \subseteq \mathbf{V} \setminus \mathbf{A}$, is identified from $P(\mathbf{V})$ via the following functional: $\sum_{\mathbf{V} \setminus (\mathbf{Y} \cup \mathbf{A})} \prod_{V \in \mathbf{V} \setminus (\mathbf{Y} \cup \mathbf{A})} P(V | \mathrm{pa}_{\mathcal{G}}(V))|_{\mathbf{A}=\mathbf{a}}$, known as the g-formula [15].

In practice, not all variables in a causal model may be observed. In a hidden variable causal model, represented by a DAG $\mathcal{G}(\mathbf{V} \cup \mathbf{H})$, where no data is available on variables in $\mathbf{H}$, not every counterfactual distribution is identified.

Reasoning about parameter identification is often performed via an acyclic directed mixed graph (ADMG) summary of $\mathcal{G}(\mathbf{V} \cup \mathbf{H})$ called a *latent projection* $\mathcal{G}(\mathbf{V})$ [20]. The latent projection keeps vertices corresponding to $\mathbf{V}$, and adds two kinds of edges between these vertices. A directed edge ($\rightarrow$) between any $V_i, V_j \in \mathbf{V}$ is added if
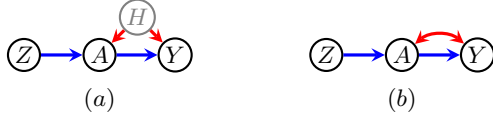
Figure 1: (a) The classic instrumental variable model, and (b) its latent projection.

there exists a directed path from V_i to V_j in $\mathcal{G}(\mathbf{V} \cup \mathbf{H})$ and all intermediate vertices on the path are in $\mathbf{H}$. A bidirected edge ($\leftrightarrow$) between any $V_i, V_j \in \mathbf{V}$ is added if there exists a path from V_i to V_j which starts with an edge into V_i, ends with an edge into V_j, has no two adjacent edges pointing into the same vertex on the path, and has all intermediate elements in $\mathbf{H}$. See Fig. 1 (a) and (b) for a simple example of this construction for the IV model.

If $P(\mathbf{Y}(\mathbf{a}))$ is identified from $P(\mathbf{V})$ in a causal model represented by a hidden variable DAG $\mathcal{G}(\mathbf{V} \cup \mathbf{H})$, then its identifying functional may be expressed using $\mathcal{G}(\mathbf{V})$ via the ID algorithm [19]. See [13] for details.

If $P(\mathbf{y}(\mathbf{a}))$ is not identified from $P(\mathbf{V})$ in a causal model given by $\mathcal{G}(\mathbf{V} \cup \mathbf{H})$, bounds may nevertheless be placed on this distribution. Before describing our approach for obtaining bounds in full generality, we illustrate how it may be used to obtain known sharp bounds for a non-identified counterfactual probability in the binary IV model.

3 BOUNDS IN THE BINARY INSTRUMENTAL VARIABLE MODEL

The instrumental variable model is represented graphically in Fig. 1 (a), with its latent projection shown in Fig. 1 (b). We are interested in the counterfactual probability $P(y(a))$, which is known not to be identified without parametric assumptions. In this section, we demonstrate that known sharp bounds on $P(y(a))$ can be recovered by reasoning causally about the structure of this graph and the associated counterfactual distributions.

The Instrumental Variable Thought Experiment

We consider the binary IV model in Fig. 1 (a) in the setting of clinical trials with non-compliance. We interpret Z to indicate treatment assignment, A to indicate treatment actually taken, and Y to indicate a clinical outcome of interest. Unobserved factors, such as personality traits, may influence both treatment decision and outcome, and thus act as confounders.

Suppose we intervene to assign some subject to treatment arm z. We observe that after this intervention, our subject takes treatment a and has outcome y. Thus, in this subject,

we observe the event $A(z) = a \wedge Y(z) = y$.

Now suppose we are interested in intervening directly on treatment for the same subject. We would like to set $A = a$, leaving Z (arm assignment) to the physician's choice. Under the model, the outcome Y depends only on A and the noise term ϵ_Y, representing in this case subject-specific personality traits. Because these traits are unchanged by intervention on A or Z, and A is set to the same value it took under our first intervention, we must conclude that under this second intervention we would observe the same outcome as under the first, denoted by the event $Y(a) = y$.

This thought experiment demonstrates that an event in one hypothetical world, under one intervention, can imply an event under another intervention. We call this phenomenon "cross-world implication," and it is formalized in proposition 2. We now develop the intuition further to recover sharp bounds for the IV model with binary random variables.

Sharp Bounds in the Binary IV Model

In the following derivation of sharp bounds for $P(y(a))$ in the binary IV model, we will denote values 1 for all variables by lower case (e.g. a), and values 0 by a lower case with a bar (e.g. $\bar{a}$).

Our strategy will be to *partition* the event $Y(a) = y$ into smaller, more manageable events:

$$A(\bar{z}) = a \wedge Y(\bar{z}) = y \tag{3}$$
$$A(\bar{z}) = \bar{a} \wedge A(z) = a \wedge Y(z) = y \tag{4}$$
$$A(\bar{z}) = \bar{a} \wedge A(z) = \bar{a} \wedge Y(a) = y. \tag{5}$$

Note that events in (4) and (5) are related to the *compliers* and *never-takers* principal strata [4].

To see that these events form a partition (i.e. are mutually exclusive and exhaustive) of the event $Y(a) = y$, we first observe that by the exclusion restriction in the model, and generalized consistency, $Y(\bar{z})$ and $Y(z)$ in equations (3) and (4) respectively will be equal to $Y(a)$. Then we can see that (3) covers the portion of $Y(a) = y$ where $A(\bar{z}) = a$, (4) covers the portion where $A(\bar{z}) = \bar{a} \wedge A(z) = a$, and (5) covers the portion where $A(\bar{z}) = A(z) = \bar{a}$.

Because these events partition $Y(a) = y$, the sum of their probabilities will be equal to $P(Y(a) = y)$, and a sum of lower-bounds on their probabilities will yield a lower bound on $P(Y(a) = y)$. A general form of partitions of this sort for counterfactual events under models with an exclusion restriction will be given in Proposition 5.

The event (3) represents a single world event with an identified probability, that therefore does not need to be bounded. We will see that we can recover sharp bounds

without bounding the probability of (5).

We therefore turn our attention to event (4). This event can be understood as a conjunction of events in two worlds. First, $A(\bar{z}) = \bar{a}$ in the world in which Z is set to $\bar{z}$. Then, $A(z) = a \wedge Y(z) = y$ in the world in which Z is set to z. As a cross world event, (4) does not have an identified density. Our goal will be to provide lower bounds for this event that are identified.

First, we find some event E_1 under the intervention $Z = \bar{z}$ that entails $A(\bar{z}) = \bar{a}$. We can then identify outcomes under the (conflicting) intervention $Z = z$ that are *compatible* with E_1, i.e. that would not be ruled out by observing E_1 under intervention $Z = \bar{z}$. We denote such events by $\psi_z(E_1)$. Now, by definition:

$$P(E_1) - P(E_1, \neg(A(z) = a \wedge Y(z) = y)) \qquad (6)$$
$$= P(E_1, (A(z) = a \wedge Y(z) = y)).$$

We note that $E_1 \implies \psi_z(E_1)$ by construction, as E_1 rules out all outcomes in the sample space not in $\psi_z(E_1)$, so $P(E_1, \neg(A(z) = a \wedge Y(z) = y))$ is bounded from above by $P(\psi_z(E_1), \neg(A(z) = a \wedge Y(z) = y))$.

Substituting this bound into equation (6) yields:

$$P(E_1) - P(\psi_z(E_1), \neg(A(z) = a \wedge Y(z) = y))$$
$$\leq P(E_1, (A(z) = a \wedge Y(z) = y)). \qquad (7)$$

Because E_1 was chosen to entail $A(\bar{z}) = \bar{a}$, the probability of event (4) is bounded from below by $P(E_1, (A(z) = a \wedge Y(z) = y))$. Therefore by equation (7), the probability of event (4) is also bounded from below by:

$$P(E_1) - P(\psi_z(E_1), \neg(A(z) = a \wedge Y(z) = y)). \qquad (8)$$

Through exactly analogous reasoning, we can obtain another lower bound on the probability of event (4) by starting with some event E_2 under $Z = z$ that entails $A(z) = a \wedge Y(z) = y$:

$$P(E_2) - P(\psi_{\bar{z}}(E_2), \neg(A(\bar{z}) = \bar{a})). \qquad (9)$$

To apply these bounds, we must select events that satisfy the criteria for E_1 and E_2. We start by examining potential events E_1. We note that there are only three options: $A(\bar{z}) = \bar{a}$, $A(\bar{z}) = \bar{a} \wedge Y(\bar{z}) = \bar{y}$, and $A(\bar{z}) = \bar{a} \wedge Y(\bar{z}) = y$. It turns out that we need only consider the latter two of these (see Proposition 10 in Appendix C).

First, we take E_1 to be $A(\bar{z}) = \bar{a} \wedge Y(\bar{z}) = \bar{y}$. Then we note $\psi_z(A(\bar{z}) = \bar{a} \wedge Y(\bar{z}) = \bar{y})$ is:

$$(A(z) = \bar{a} \wedge Y(z) = \bar{y})$$
$$\vee (A(z) = a \wedge Y(z) = \bar{y})$$
$$\vee (A(z) = a \wedge Y(z) = y).$$

The only outcome under the intervention $Z = z$ excluded from this event is $A(z) = \bar{a} \wedge Y(z) = y$. Any subject

who experienced this event could not have experienced $A(\bar{z}) = \bar{a} \wedge Y(\bar{z}) = \bar{y}$, due to the exclusion restriction in the IV model.

According to (8), to obtain a bound we will need to subtract from the mass of E_1 the mass of the portion of $\psi_z(E_1)$ where $A(z) = a \wedge Y(z) = y$ does not hold. This will be the mass of the first two events in the disjunction above.

Using this value of E_1, we therefore obtain the following lower bound on the probability of (4):

$$P(A(\bar{z}) = \bar{a}, Y(\bar{z}) = \bar{y}) - \qquad (10)$$
$$P((A(z) = \bar{a}, Y(z) = \bar{y}) \vee (A(z) = a, Y(z) = \bar{y})).$$

We now consider the bound induced by using $A(\bar{z}) = \bar{a} \wedge Y(\bar{z}) = y$ as the event E_1. Following an analogous procedure, we produce the lower bound:

$$P(A(\bar{z}) = \bar{a}, Y(\bar{z}) = y) - \qquad (11)$$
$$P((A(z) = \bar{a}, Y(z) = y) \vee (A(z) = a_1, Y(z) = \bar{y})).$$

Next, we consider possible values of E_2. In this simple case, there is only one such possibility, $A(z) = a \wedge Y(z) = y$, which of course entails itself. We observe that $\psi_{\bar{z}}(E_2) \wedge \neg(A(\bar{z}) = \bar{a})$ is equivalent to $A(\bar{z}) = a \wedge Y(\bar{z}) = y$, yielding the following lower bound by expression (9):

$$P(A(z) = a, Y(z) = y) - P(A(\bar{z}) = a, Y(\bar{z}) = y).$$

We now have all the pieces we need to obtain a sharp lower bound on $P(y(a))$. We make use of the fact that distributions of potential outcomes after interventions on Z are identified as the distribution of the corresponding observed random variables conditioned on Z (since Z is randomized in the IV model). Noting that the density of the event (4) is also bounded from below by 0, we add the identified density of the event (3) to the best of the lower bounds we have obtained for (4). Then

$$P(y(a)) \geq P(a, y \mid \bar{z}) +$$
$$\max \begin{cases} 0 \\ P(\bar{a}, \bar{y} \mid \bar{z}) - P(\bar{a}, \bar{y} \mid z) - P(a, \bar{y} \mid z) \\ P(\bar{a}, y \mid \bar{z}) - P(\bar{a}, y \mid z) - P(a, \bar{y} \mid z) \\ P(a, y \mid z) - P(a, y \mid \bar{z}). \end{cases}$$

This is the sharp lower bound obtained by Balke [1]. $P(y(a))$ may be bounded from above by 1 less the lower bound on $P(\bar{y}(a))$. In the binary case, bounds on the ACE may simply be represented as differences between appropriate bounds on $P(y(a))$ and $P(y(\bar{a}))$. Each of these bounds bounds is sharp for the binary IV model. However, characterizing models for which bounds derived by the procedure we propose, described in the next section, are sharp is an open problem.

4 BOUNDS ON COUNTERFACTUAL EVENTS

In this section we provide a graphical criterion for the presence of an implicative relationship between counterfactual events, which we call *cross-world implications*, and demonstrate its use in bounding non-identified probabilities of counterfactual events. We then show how these bounds can be aggregated to bound non-identified counterfactual events of primary interest. Proofs of all claims are found in Appendix A.

Causal Irrelevance, Event Implication and Event Contradiction

In deriving bounds on a counterfactual event under the IV model, we made use of the exclusion restriction $Y(z, a) = Y(a)$. We begin this section by providing a general graphical criterion for when such restrictions appear in causal models.

Proposition 1 (Causal Irrelevance). *If all directed paths from $\mathbf{Z}$ to $\mathbf{Y}$ contain members of $\mathbf{A}$, then*

$$\mathbf{Y}(\mathbf{Z} = \mathbf{z}, \mathbf{A} = \mathbf{a}) = \mathbf{Y}(\mathbf{A} = \mathbf{a}).$$

In such cases, we say $\mathbf{Z}$ is *causally irrelevant* to $\mathbf{Y}$ given $\mathbf{A}$, because after intervening on $\mathbf{A}$, intervening on $\mathbf{Z}$ will not affect $\mathbf{Y}$. If in addition the joint distribution $P(\mathbf{Y}(\mathbf{z}), \mathbf{A}(\mathbf{z}))$ is identified, $\mathbf{Z}$ is said to be a *generalized instrument* for $\mathbf{A}$ with respect to $\mathbf{Y}$. If the set $\mathbf{A}$ can be partitioned into $\mathbf{A_1}$ and $\mathbf{A_2}$ such that $\mathbf{A_1}$ is causally irrelevant to $\mathbf{Y}$ given $\mathbf{A_2}$, then any such $\mathbf{A_1}$ is said to be causally irrelevant to $\mathbf{Y}$ in $\mathbf{A}$. See also rule 3^* in [7], and the discussion of minimal labeling of counterfactuals in [14]. As noted earlier, constraints in a causal model corresponding to the existence of causally irrelevant variables are sometimes called *exclusion restrictions*.

In the following proposition, we observe that whenever an exclusion restriction appears in the graph, there exists a logical implication connecting counterfactual events across interventional worlds.

Proposition 2 (Cross-world Implication). *Let $\mathbf{Z}$ be causally irrelevant to $\mathbf{Y}$ given $\mathbf{A}$. Then*

$$\mathbf{A}(\mathbf{z}) = \mathbf{a} \wedge \mathbf{Y}(\mathbf{z}) = \mathbf{y} \implies \mathbf{Y}(\mathbf{a}) = \mathbf{y}.$$

We define a collection of events to be *compatible* if none of them implies the negation of any other event in the collection. We define a collection of events to be *contradictory* if it is not compatible. Conceptually, events in different hypothetical worlds are contradictory if, under the model, no single subject can experience all of the events under their corresponding interventions.

For example, in the IV model though experiment, we saw that no single subject can experience both the event $A(z) = a \wedge Y(z) = y$ and the event $Y(a) \neq y$, rendering them contradictory.

It will be of use to be able to determine whether events are contradictory through reference to the graphical model. To that end, we provide a recursive graphical criterion that is sufficient to establish that events are contradictory.

Proposition 3 (Contradictory Events). *Two events $\mathbf{X}(\mathbf{a}) = \mathbf{x}$ and $\mathbf{Y}(\mathbf{b}) = \mathbf{y}$ are contradictory if there exists $Z \in \mathbf{X} \cup \mathbf{Y}$ such that $Z(\mathbf{a}) \neq Z(\mathbf{b})$, and all of the following hold:*

(i) Variables in the subsets of both $\mathbf{X} \cup \mathbf{A}$ and $\mathbf{Y} \cup \mathbf{B}$ causally relevant for Z are set to the same values in $\mathbf{x}, \mathbf{a}$, and $\mathbf{y}, \mathbf{b}$.

(ii) Let $C \in \{\mathbf{X} \cup \mathbf{A}\} \setminus \{\mathbf{Y} \cup \mathbf{B}\}$ be any variable that is causally relevant to Z in $\mathbf{X} \cup \mathbf{A}$ and causally relevant to Z given $\mathbf{Y} \cup \mathbf{B}$, with C set to c in $\mathbf{x}, \mathbf{a}$. Then $\mathbf{X}(\mathbf{a}) = \mathbf{x}$ and $\mathbf{Y}(\mathbf{b}) = \mathbf{y} \wedge C(\mathbf{b}) = c'$ are known to be contradictory by this proposition if $c \neq c'$.

(iii) Let $C \in \{\mathbf{Y} \cup \mathbf{B}\} \setminus \{\mathbf{X} \cup \mathbf{A}\}$ be any variable that is causally relevant to Z in $\mathbf{Y} \cup \mathbf{B}$ and causally relevant to Z given $\mathbf{X} \cup \mathbf{A}$, with C set to c in $\mathbf{y}, \mathbf{b}$. Then $\mathbf{Y}(\mathbf{b}) = \mathbf{y}$ and $\mathbf{X}(\mathbf{a}) = \mathbf{x} \wedge C(\mathbf{a}) = c'$ are known to be contradictory by this proposition if $c \neq c'$.

Propositions 2 and 3 provide graphical criteria for implication and contradiction, based on paths in the causal diagram. Both criteria are stated in terms of exclusions restrictions in the graph. It should be noted that not all exclusion restrictions can be represented graphically; for example, some exclusions may obtain only for certain levels of the variables in the graph, and not universally. If such context-specific exclusion restrictions arise, they may lead to implications or contradictions not captured by these criteria. However, in the absence of exclusion restrictions not represented by the graphical model, the graphical criteria provided by these propositions are necessary and sufficient. See Appendix B for details.

Bounds Via a Single Cross-World Implication

In this section, we describe a lower bound on $P(\mathbf{y}(\mathbf{a}))$ induced by a single cross-world implication, of the sort described by Proposition 2. We will demonstrate that this line of reasoning can be used to recover the bounds for the IV model in [8, 16], and produce a new class of bounds on densities of counterfactual events where the density is identified under intervention on a subset of the treatment variables.
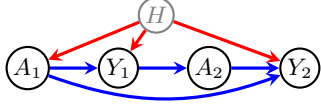
Figure 2: A Sequential Treatment Scenario

We begin with a simple result from probability theory, which can broadly be viewed as stating that supersets will always have weakly larger measure than their subsets.

Proposition 4. *Let E_1, E_2 be any events in a causal model such that $E_1 \implies E_2$. Then $P(E_1) \leq P(E_2)$.*

In the case of the IV model, we have noted the exclusion restriction between Z and Y given A. Due to the implication established by the IV thought experiment and formalized in Proposition 2, Proposition 4 then yields $P(Y(a) = y) \geq P(A(z) = a, Y(z) = y)$ for any value of z. Noting that Z has no parents in the model, and that therefore the interventional distribution is identified as the conditional, we can write this as $\max_z P(A = a, Y = y \mid Z = z)$, which is equivalent to the binary IV bounds in [8, 16].

We now present new bounds on causal parameters, based on the observation that the empty set may act as a generalized instrument for any treatment set $\mathbf{A}$ with respect to any outcome $\mathbf{Y}$. This observation allows us to combine Propositions 2 and 4 to obtain the following Corollary.

Corollary 1. *For any sets of variables $\mathbf{Y}, \mathbf{A}$,*

$$P(\mathbf{y}(\mathbf{a})) \in [P(\mathbf{Y} = \mathbf{y}, \mathbf{A} = \mathbf{a}), 1 - P(\mathbf{Y} \neq \mathbf{y}, \mathbf{A} = \mathbf{a})].$$

A consequence of this Corollary is that for discrete variables, densities of counterfactual events can be non-trivially bounded for *any* causal model, though we do not expect these bounds to be informative in general.

Finally, we present bounds on densities of counterfactuals when a subset of the treatment set can act as a generalized instrument for the remainder.

Corollary 2. *Let $\tilde{\mathbf{A}}$ and $\hat{\mathbf{A}}$ partition $\mathbf{A}$, such that the density $P(\mathbf{Y}(\tilde{\mathbf{a}}), \hat{\mathbf{A}}(\tilde{\mathbf{a}}))$ is identified, where $\tilde{\mathbf{a}}$ is the subset of $\mathbf{a}$ corresponding to $\tilde{\mathbf{A}}$. Then*

$$P(\mathbf{Y}(\tilde{\mathbf{a}}) = \mathbf{y}, \hat{\mathbf{A}}(\tilde{\mathbf{a}}) = \hat{\mathbf{a}}) \leq P(\mathbf{Y}(\mathbf{a}) = \mathbf{y})$$

$$1 - P(\mathbf{Y}(\tilde{\mathbf{a}}) \neq \mathbf{y}, \hat{\mathbf{A}}(\tilde{\mathbf{a}}) = \hat{\mathbf{a}}) \geq P(\mathbf{Y}(\mathbf{a}) = \mathbf{y}),$$

where $\hat{\mathbf{a}}$ is the subset of $\mathbf{a}$ corresponding to $\hat{\mathbf{A}}$.

Example 1 (Sequential Treatment Scenario). *In the model depicted in Fig. 2, the A variables represent treatments and the Y variables represent outcomes. This model may be applicable if the initial treatment A_1 is selected by the subject – and is therefore confounded with*

the outcomes through the subject's unobserved traits H – but the second treatment A_2 is selected solely on the basis of the first-stage outcome, Y_1.

We may be interested in $P(Y_2(a_1, a_2))$ – the distribution of the second stage outcome under intervention on both treatments. This distribution is not identified. However, since $P(A_1(a_2) = a_1, Y_2(a_2) = y_2)$ is identified as $\sum_{y_1} \frac{P(a_1, y_1, a_2, y_2)}{P(a_2 \mid y_1)}$, Corollary 2 yields the bounds:

$$\sum_{y_1} \frac{P(a_1, y_1, a_2, y_2)}{P(a_2 \mid y_1)} \leq P(Y(a_1, a_2) = y_2)$$

$$1 - \sum_{y_1, \tilde{y}_2 \neq y_2} \frac{P(a_1, y_1, a_2, \tilde{y}_2)}{P(a_2 \mid y_1)} \geq P(Y(a_1, a_2) = y_2).$$

Bounds Via Multiple Cross-World Implications

In this section, we show how information from multiple cross-world implications may be used to obtain bounds. We begin by describing a partition of the event of interest, where the partition is defined by cross-world potential outcomes. We then develop a method for bounding the density of the cross-world events in the partition, and aggregate the bounds. This section generalizes the procedure used to obtain sharp bounds for the binary IV model in Section 3.

Partitioning the Event of Interest

We denote our event of interest as $\mathbf{Y}(\mathbf{a_1}) = \mathbf{y}$. We are interested in providing a lower bound for the non-identified probability $P(\mathbf{Y}(\mathbf{a_1}) = \mathbf{y})$. We begin by defining a partition of the event to interest. We will work with these partition sets for the remainder of the section.

Proposition 5 (Partition Sets). *Let $\mathbf{Z}$ be causally irrelevant to $\mathbf{Y}$ given $\mathbf{A}$. Assume $\mathbf{Z}, \mathbf{A}$ are discrete, and take levels $\mathbf{z_1}, \cdots, \mathbf{z_M}$ and $\mathbf{a_1}, \cdots, \mathbf{a_N}$ respectively.*

Then the following events are a partition of $\mathbf{Y}(\mathbf{a_1}) = \mathbf{y}$:

$$\mathbf{Y}(\mathbf{a_1}) = \mathbf{y} \wedge \forall \mathbf{z}(\mathbf{A}(\mathbf{z}) \neq \mathbf{a_1}) \tag{12}$$

$$\mathbf{A}(\mathbf{z_1}) = \mathbf{a_1} \wedge \mathbf{Y}(\mathbf{z_1}) = \mathbf{y} \tag{13}$$

and, for $\mathbf{k} = 2, \cdots, N$,

$$\mathbf{A}(\mathbf{z_1}) = \mathbf{a_k} \wedge \exists \mathbf{z}(\mathbf{A}(\mathbf{z}) = \mathbf{a_1} \wedge \mathbf{Y}(\mathbf{z}) = \mathbf{y}). \tag{14}$$

We now develop lower bounds on the density of each of the partition events, which we can then use to lower bound the density of the target event $\mathbf{Y}(\mathbf{a_1}) = \mathbf{y}$.

Bounding Partition-Set Densities

Event (13) is a single world event with identified density, so there is no need to find a lower bound. Subjects that experience event (12) will never experience

$\mathbf{Y}(\mathbf{z}) = \mathbf{y} \wedge \mathbf{A}(\mathbf{z}) = \mathbf{a_1}$ under any intervention $\mathbf{Z} = \mathbf{z}$ with an identified distribution. Because our strategy uses information from identified distributions to bound unidentified densities, we cannot provide bounds on the density of this partition event.

We now turn our attention to events of the form of (14). We let E^k denote the event of the form of (14) where the first term is $\mathbf{A}(\mathbf{z_1}) = \mathbf{a_k}$. Then we note E^k can be represented as the disjunction $\bigvee_{j=1}^{M} \gamma_j^k$, where γ_j^k denotes the event:

$$\gamma_j^k \triangleq \mathbf{A}(\mathbf{z_1}) = \mathbf{a_k} \wedge \mathbf{A}(\mathbf{z_j}) = \mathbf{a_1} \wedge \mathbf{Y}(\mathbf{z_j}) = \mathbf{y}. \quad (15)$$

Because $\gamma_j^k \subseteq E^k$, we know $\gamma_j^k \implies E^k$. It follows from Proposition 4 that

$$P(E^k) \geq \max_j P(\gamma_j^k). \quad (16)$$

Unfortunately $P(\gamma_j^k)$ is also not point-identified, and must be bounded from below itself.

The event γ_j^k conjoins statements about potential outcomes under two different interventions. Its density is the portion of the population who would experience $\mathbf{A}(\mathbf{z_1}) = \mathbf{a_k}$ under $\mathbf{Z} = \mathbf{z_1}$, and $\mathbf{A}(\mathbf{z_j}) = \mathbf{a_1} \wedge \mathbf{Y}(\mathbf{z_j}) = \mathbf{y}$ under $\mathbf{Z} = \mathbf{z_j}$. We know the exact proportion of the population who would experience either, because $\mathbf{Z}$ is a generalized instrument, but we do not know the exact portion of the population that would experience both.

To address this problem, we first consider the problem of bounding the density of an event conjoining potential outcomes under two different interventions in general terms, in the following Proposition. This result can then be directly applied to lower bound the density of γ_j^k.

Proposition 6 (Cross-World Lower Bounds). *Let $\psi_\mathbf{c}(E)$ represent the disjunction of all outcomes in the sample space under intervention $\mathbf{C} = \mathbf{c}$ that do not contradict the event E, such that $E \implies \psi_\mathbf{c}(E)$.*

Let $E_\mathbf{x}$ be any event that implies $\mathbf{X}(\mathbf{a}) = \mathbf{x}$, and $E_\mathbf{y}$ be any event that implies $\mathbf{Y}(\mathbf{b}) = \mathbf{y}$. Then $P(\mathbf{X}(\mathbf{a}) = \mathbf{x}, \mathbf{Y}(\mathbf{b}) = \mathbf{y})$ is bounded from below by each of:

$$P(E_\mathbf{x}) - P(\psi_\mathbf{b}(E_\mathbf{x}), \mathbf{Y}(\mathbf{b}) \neq \mathbf{y})$$

$$P(E_\mathbf{y}) - P(\psi_\mathbf{a}(E_\mathbf{y}), \mathbf{X}(\mathbf{a}) \neq \mathbf{x}).$$

Proposition 6 is useful because, in each of the bounds provided, each of the densities involved are in terms of events under a single intervention. If densities under those interventions are identified, the bounds can be calculated exactly.

We return to our goal of bounding $P(\gamma_j^k)$ from below. In the case of γ_j^k, the two interventions we are interested in are on the same set of variables, $\mathbf{Z}$. As described

above, we are interested in the proportion of patients who experience $\mathbf{A}(\mathbf{z_1}) = \mathbf{a_k}$ under the intervention $\mathbf{Z} = \mathbf{z_1}$, and $\mathbf{A}(\mathbf{z_j}) = \mathbf{a_1} \wedge \mathbf{Y}(\mathbf{z_j}) = \mathbf{y}$ under the intervention $\mathbf{Z} = \mathbf{z_j}$. Substituting these values for $\mathbf{X}(\mathbf{a}) = \mathbf{x}$ and $\mathbf{Y}(\mathbf{b}) = \mathbf{y}$ into Proposition 6 immediately yields the following Corollary.

Corollary 3 (Lower bounds on $P(\gamma_j^k)$). *Let E_1 be an event under intervention $\mathbf{Z} = \mathbf{z_1}$ that entails $\mathbf{A}(\mathbf{z_1}) = \mathbf{a_k}$, and E_2 be event under intervention $\mathbf{Z} = \mathbf{z_j}$ that entails $\mathbf{A}(\mathbf{z_j}) = \mathbf{a_1} \wedge \mathbf{Y}(\mathbf{z_j}) = \mathbf{y}$.*

Then $P(\mathbf{A}(\mathbf{z_1}) = \mathbf{a_k} \wedge \mathbf{A}(\mathbf{z_j}) = \mathbf{a_1} \wedge \mathbf{Y}(\mathbf{z_j}) = \mathbf{y})$ is bounded from below by each of:

$$P(E_1) - P(\psi_{\mathbf{z_j}}(E_1) \wedge \neg(\mathbf{A}(\mathbf{z_j}) = \mathbf{a_1} \wedge \mathbf{Y}(\mathbf{z_j}) = \mathbf{y}))$$

$$P(E_2) - P(\psi_{\mathbf{z_1}}(E_2) \wedge \mathbf{A}(\mathbf{z_1}) \neq \mathbf{a_k}).$$

With this result in hand, we can we can modify (16) to obtain the following bound on $P(E^k)$ in terms of the observed data law. Let $\xi(\cdot)$ represent the set of lower bounds on the density $P(\cdot)$ obtained through Corollary 3 for all possible values of E_1 and E_2. Then

$$P(E^k) \geq \max_j \left(\max \xi(\gamma_j^k) \right). \quad (17)$$

Recalling that $P(\mathbf{Y}(\mathbf{a_1}) = \mathbf{y})$ can be bounded from below by the sum of lower bounds on densities of its partition sets, we obtain the lower bound

$$P(\mathbf{A}(\mathbf{z_1}) = \mathbf{a_1}, \mathbf{Y}(\mathbf{z_1}) = \mathbf{y}) + \sum_{k=2}^{N} \max_j \left(\max \xi(\gamma_j^k) \right),$$

where the first term corresponds to the density of event (13) and the second term is a sum is over lower bounds on the densities of the events of the form of (14).

Algorithm 1 Lower Bounds on $P(\mathbf{Y}(\mathbf{a_1}) = \mathbf{y})$

Input: event $\mathbf{Y}(\mathbf{a_1}) = \mathbf{y}$
       generalized instrument $\mathbf{Z}$
Output: bounds on $P(\mathbf{Y}(\mathbf{a_1}) = \mathbf{y})$
1: Bounds = {}
2: **For** $k = 2, \ldots, N$:
3:     KBounds = {}
4:     **For** $j = 1, \ldots, M$:
5:       **For** E_1 in $\{E_1 \implies \mathbf{Y}(\mathbf{z_1}) = \mathbf{a_k}\}$
6:         induce bound by E_1 through Corollary 3
7:         KBounds.add(bound)
8:       **For** E_2 in $\{E_2 \implies \mathbf{Y}(\mathbf{z_j}) = \mathbf{y} \wedge \mathbf{A}(\mathbf{z_j}) = \mathbf{a_1}\}$
9:         induce bound by E_2 through Corollary 3
10:        KBounds.add(bound)
11:     Bounds.add($\max$(KBounds))
12: $\mathbf{P}(\mathbf{Y}(\mathbf{z_1}) = \mathbf{y_1}, \mathbf{A}(\mathbf{z_1}) = \mathbf{a_1})$ + sum(Bounds)

Algorithm 1 summarizes how the results described in this section can be used to calculate these bounds.

5 GENERALIZED INSTRUMENTAL INEQUALITIES

In this section, we develop generalized instrumental inequalities and show that the IV inequalities, and Bonet's inequalities [2], are special cases. To begin, we bound the sum of probabilities of events in terms of the size of the largest subset thereof that is made up of compatible events.

Proposition 7. *Let $E_1, \cdots, E_N$ be events under arbitrary interventions such that at most k of the events are compatible. Then*

$$\sum_{i=1}^{N} P(E_i) \leq k.$$

This result is a consequence of the fact that by construction, no value of $\epsilon_{\mathbf{V}}$ can lead to contradictory events. If such a value did exist, observing one of the events leaves open the possibility that $\epsilon_{\mathbf{V}}$ takes that value, in which case we would observe the other event under the appropriate intervention, and the events would not be contradictory. It follows that if we are adding the densities of events of which at most k are compatible, no set in the domain of $\epsilon_{\mathbf{V}}$ may have its measure counted more than k times.

We make use of this result, in combination with our existing results about causal irrelevance, to obtain the following class of inequality constraints.

Corollary 4 (Generalized Instrumental Inequalities). *Let $\mathbf{Z}$ be causally irrelevant to $\mathbf{Y}$ given $\mathbf{A}$, and let $\mathcal{S}$ be any set of triples $(\mathbf{z}, \mathbf{a}, \mathbf{y})$ which represent levels of $\mathbf{Z}, \mathbf{A}, \mathbf{Y}$. Then*

$$\sum_{(\mathbf{z},\mathbf{a},\mathbf{y}) \in \mathcal{S}} P(\mathbf{A}(\mathbf{z}) = \mathbf{a}, \mathbf{Y}(\mathbf{z}) = \mathbf{y}) \leq \Phi(\mathcal{S})$$

where

$$\Phi(\mathcal{S}) = \max \big\{ |\mathcal{Q}| \mid \mathcal{Q} \subseteq S \wedge \forall (\mathbf{z}, \mathbf{a}, \mathbf{y}), (\mathbf{z}', \mathbf{a}', \mathbf{y}') \in \mathcal{Q}$$
$$\neg((\mathbf{z} = \mathbf{z}' \wedge \mathbf{a} \neq \mathbf{a}') \vee (\mathbf{a} = \mathbf{a}' \wedge \mathbf{y} \neq \mathbf{y})) \big\}.$$

This result makes use of the fact that by Proposition 3, if $\mathbf{Z}$ is causally irrelevant to $\mathbf{Y}$ given $\mathbf{A}$ and $(\mathbf{z} = \mathbf{z}' \wedge \mathbf{a} \neq \mathbf{a}') \vee (\mathbf{a} = \mathbf{a}' \wedge \mathbf{y} \neq \mathbf{y}')$, then $\mathbf{A}(\mathbf{z}) = \mathbf{a} \wedge \mathbf{Y}(\mathbf{z}) = \mathbf{y}$ and $\mathbf{A}(\mathbf{z}') = \mathbf{a}' \wedge \mathbf{Y}(\mathbf{z}') = \mathbf{y}'$ are contradictory. $\Phi(\mathcal{S})$ can therefore be interpreted as the size of largest compatible subset of $\mathcal{S}$.

The IV inequalities derived in [10], which can be written $\forall a \sum_y \max_z P(Y(z), A(z)) \leq 1$, are a special case of the generalized instrumental inequalities with $k = 1$. For each selection of a, the sum is over densities of events with different values of y, rendering them pairwise contradictory. We now review the inequality derived in [2].

Example 2 (Bonet's Inequalities). *Bonet [2] presents the following constraint for the IV model, where treatment and outcome are binary and the instrument is ternary:*

$$P(a_1, y_2 \mid z_2) + P(a_1, y_1 \mid z_3) + P(a_1, y_2 \mid z_1)$$
$$+ P(a_2, y_2 \mid z_2) + P(a_2, y_1 \mid z_1) \leq 2$$

These densities are respectively equal to the densities of the following events, through the fact that densities under intervention on variables with no parent are identified as the conditional distribution:

$$A(z_2) = a_1 \wedge Y(z_2) = y_2 \qquad (18)$$
$$A(z_3) = a_1 \wedge Y(z_3) = y_1 \qquad (19)$$
$$A(z_1) = a_1 \wedge Y(z_1) = y_2 \qquad (20)$$
$$A(z_2) = a_2 \wedge Y(z_2) = y_2 \qquad (21)$$
$$A(z_1) = a_2 \wedge Y(z_1) = y_1. \qquad (22)$$

It can easily be confirmed that no subset of size 3 or greater is mutually compatible. For example, event (18) is compatible with events (20) and (22), but these are incompatible with each other, due to Z taking the same value in both but A taking a different value in each. The same pattern follows for all events; each event is compatible with two others which in turn are not compatible with each other.

It follows from Corollary 4 that the sum of the densities of these events must be bounded from above by 2.

6 EXAMPLE APPLICATIONS

In this section, we derive bounds and inequality constraints for the ADMGs presented in Fig. 3 using the results presented in Sections 4 and 5. We are not aware of any existing methods that can obtain the bounds presented below. Code used to obtain these results, as well as a general implementation of the methods described in this paper, is publicly available [1].

Due to space constraints, we denote the identified distribution under intervention on $Z = z$ as $P_z(\cdot)$. In addition, we do not consider more complicated scenarios, e.g. involving multiple instruments $\mathbf{Z}$ and treatments $\mathbf{A}$, instruments with challenging identifying functionals, or non-binary variables. However, bounds and constraints in such scenarios may be obtained using our software.

The IV Model With Covariates

We first consider the model represented by Fig. 3 (a). In the traditional IV model, the instrument must be randomized with respect to the treatment and outcome. In practice, it can be difficult to find such instruments. The

---
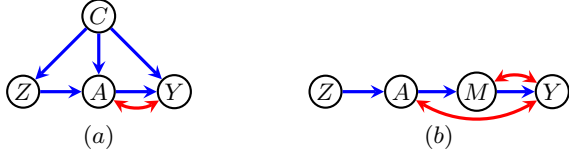
[1] https://noamfinkelste.in/partial-id

Figure 3: (a) The IV model with covariates, and (b) the confounded frontdoor IV model.

IV model with covariates allows for the instrument to be *conditionally* randomized.

In the social sciences, exogenous shocks are often used as instrumental variables. For example, suppose an earthquake damages a number of school buildings, increasing class size at nearby schools. An economist studying the effect of class size on test scores might use school closure due to the earthquake as an instrument for class size.

This instrument may not be entirely plausible. Families with more resources may be able to avoid living in areas at risk of earthquake damage, and wealthier school districts may be better able to build robust school buildings. In this case, the instrument would be confounded with the treatment and outcome. Observed baseline covariates for the school districts, including information on tax revenue, may be sufficient to account for this kind of confounding. In settings of this kind, the IV model with covariates is appropriate, whereas the traditional IV model is not.

We present the following lower bound on $P(Y(\bar{a}) = \bar{y})$ under this model when variables are binary:

$$P_{\bar{z}}(\bar{a}, \bar{y}) +$$

$$\max \begin{cases} 0 \\ P_z(\bar{a}, \bar{y}, c) - P_{\bar{z}}(c, \bar{a}, \bar{y}) \\ P_{\bar{z}}(a, c, \bar{y}) - \big(P_z(c, \bar{a}, y) + P_z(c, a, \bar{y})\big) \\ P_{\bar{z}}(a, \bar{c}, y) - P_z(\bar{c}, y) \\ P_z(\bar{a}, \bar{y}) - P_{\bar{z}}(\bar{a}, \bar{y}) \\ P_{\bar{z}}(a, y) - P_z(y) \\ P_{\bar{z}}(a, \bar{y}) - \big(P_z(a, \bar{y}) + P_z(\bar{a}, y)\big) \\ P_{\bar{z}}(a, c, y) - P_z(c, y) \\ P_z(\bar{a}, \bar{y}, \bar{c}) - P_{\bar{z}}(\bar{c}, \bar{a}, \bar{y}) \\ P_{\bar{z}}(a, c) - \big(P_z(c, y) + P_z(c, a, \bar{y})\big) \\ P_{\bar{z}}(a, \bar{c}) - \big(P_z(\bar{c}, a, \bar{y}) + P_z(\bar{c}, y)\big) \\ P_{\bar{z}}(a, \bar{c}, \bar{y}) - \big(P_z(\bar{c}, \bar{a}, y) + P_z(\bar{c}, a, \bar{y})\big). \end{cases}$$

A derivation of these bounds is provided in Appendix E.

We can use Corollary 4 to obtain inequality constraints on the observed data law implied by the model. Such constraints cannot easily be concisely expressed. Two representative expressions, each bounded from above by 1, are as follows:

$$\max_{(c,a) \neq (c',a')} \left(\sum_{y \neq z} \frac{P(a, c, y, z)}{P(z \mid c)} + \sum_{y = z} \frac{P(a', c', y, z)}{P(z \mid c')} \right)$$

$$\max_{c \neq c', z \neq z'} \sum_{a, y} \left(\frac{P(a, c, y, z)}{P(z \mid c)} + \frac{P(a, c', y, z')}{P(z' \mid c')} \right).$$

Front-Door IV Model With Confounding

This model, illustrated in Fig. 3 (b), is appropriate when the effect of treatment is only through an observed mediator, which is itself confounded with the outcome. In such cases, the traditional IV model can be applied by ignoring data on M, but tighter bounds can be obtained when the mediator is considered. When all variables are binary, our method yields the following lower bound on $P(Y(\bar{a}) = \bar{y})$:

$$P_{\bar{z}}(\bar{a}, \bar{y}) +$$

$$\max \begin{cases} 0 \\ P_z(\bar{a}, \bar{y}, \bar{m}) - P_{\bar{z}}(\bar{a}, \bar{m}, \bar{y}) \\ P_z(\bar{a}, \bar{y}) - P_{\bar{z}}(\bar{a}, \bar{y}) \\ P_{\bar{z}}(a, \bar{m}, y) - \big(P_z(\bar{m}, y) + P_z(\bar{a}, m, y)\big) \\ P_{\bar{z}}(a, m, y) - \big(P_z(m, y) + P_z(\bar{a}, \bar{m}, y)\big) \\ P_{\bar{z}}(a, \bar{y}) - \big(P_z(a, \bar{y}) + P_z(\bar{a}, y)\big) \\ P_{\bar{z}}(a, m) - \big(P_z(a, m, \bar{y}) + P_z(m, y) + P_z(\bar{a}, \bar{m}, y)\big) \\ P_z(\bar{a}, \bar{y}, m) - P_{\bar{z}}(\bar{a}, m, \bar{y}) \\ P_{\bar{z}}(a, m, \bar{y}) - \big(P_z(\bar{a}, \bar{m}, y) + P_z(a, m, \bar{y})\big) \\ P_{\bar{z}}(a, \bar{m}, \bar{y}) - \big(P_z(\bar{a}, m, y) + P_z(a, \bar{m}, \bar{y})\big) \\ P_{\bar{z}}(a, y) - P_z(y) \\ P_{\bar{z}}(a, \bar{m}) - \big(P_z(\bar{m}, y) + P_z(\bar{a}, m, y) + P_z(a, \bar{m}, \bar{y})\big). \end{cases}$$

Finally, we present two functionals of the observed data law that, under the model, are bounded from above by 1. Each is representative of a class of constraints that does not have a concise general formula.

$$P(\bar{a}, \bar{m}, y \mid \bar{z}) + P(a, m, y \mid \bar{z}) + P(\bar{a}, m, \bar{y} \mid z) + P(a, \bar{m}, \bar{y} \mid z)$$

$$\sum_{a'} P(a', \bar{m}, \bar{y} \mid \bar{z}) + \sum_{y'} P(a, m, y' \mid \bar{z}) + P(a, \bar{m}, y \mid z).$$

7   CONCLUSION

The methods pursued in this work take advantage of identified counterfactual distributions to bound causal parameters that are not identified, and provide inequality constraints on functionals of the observed data law. These bounds expand the class of causal models under which counterfactual random variables may be meaningfully analyzed, and the inequality constraints facilitate falsification of causal models by observed data. Characterizing the conditions under which these bounds and inequalities are sharp remains an open question. We also leave open application of these ideas to other areas of study interested in counterfactual parameters, such as missing data, dependent data, and policy learning.

References

[1] Alexander Balke and Judea Pearl. Bounds on treatment effects from studies with imperfect compliance. *Journal of the American Statistical Association*, 92(439):1171–1176, 1997.

[2] Blai Bonet. Instrumentality tests revisited. In *Proceedings of the Seventeenth Conference on Uncertainty and Artificial Intelligence*, pages 48–55, 2001.

[3] Robin J. Evans. Graphical methods for inequality constraints in marginalized dags. In *Proceedings of the 2012 IEEE International Workshop on Machine Learning for Signal Processing*, 2012.

[4] Constantine Frangakis and Donald B. Rubin. Principal stratification in causal inference. *Biometrics*, 58(1):21–29, 2002.

[5] Yimin Huang and Marco Valtorta. Pearl's calculus of interventions is complete. In *Twenty Second Conference On Uncertainty in Artificial Intelligence*, 2006.

[6] 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. AUAI Press, 2006.

[7] Daniel Malinsky, Ilya Shpitser, and Thomas Richardson. A potential outcomes calculus for identifying conditional path-specific effects. *Proceedings of machine learning research*, 89:3080, 2019.

[8] C.F. Manski. Nonparametric bounds on treatment effects. *The American Economic Review*, 80:319–323, 1990.

[9] Jerzy Neyman. Sur les applications de la thar des probabilities aux experiences agaricales: Essay des principle. excerpts reprinted (1990) in English. *Statistical Science*, 5:463–472, 1923.

[10] Judea Pearl. Causal inference from indirect experiments. *Artificial intelligence in medicine*, 7(6):561–582, 1995.

[11] Judea Pearl. *Causality: Models, Reasoning, and Inference*. Cambridge University Press, 2 edition, 2009.

[12] Roland R. Ramsahai. Causal bounds and observable constraints for non-deterministic models. *Journal of Machine Learning Research*, 2012.

[13] Thomas S. Richardson, Robin J. Evans, James M. Robins, and Ilya Shpitser. Nested Markov properties for acyclic directed mixed graphs, 2017. Working paper.

[14] Thomas S. Richardson and Jamie M. Robins. Single world intervention graphs (SWIGs): A unification of the counterfactual and graphical approaches to causality. *preprint:* `http://www.csss.washington.edu/Papers/wp128.pdf`, 2013.

[15] James M. Robins. A new approach to causal inference in mortality studies with sustained exposure periods – application to control of the healthy worker survivor effect. *Mathematical Modeling*, 7:1393–1512, 1986.

[16] James M. Robins. The analysis of randomized and non-randomized aids treatment trials using a new approach to causal inference in longitudinal studies. In L. Sechrest, H. Freeman, and A. Mulley, editors, *Health Service Research Methodology: A Focus on AIDS*, pages 113–159. NCHSR, U.S. Public Health Service, 1989.

[17] Ilya Shpitser and Judea Pearl. Identification of joint interventional distributions in recursive semi-Markovian causal models. In *Proceedings of the Twenty-First National Conference on Artificial Intelligence (AAAI-06)*. AAAI Press, Palo Alto, 2006.

[18] Peter Spirtes, Clark Glymour, and Richard Scheines. *Causation, Prediction, and Search*. Springer Verlag, New York, 2 edition, 2001.

[19] 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 (UAI-02)*, volume 18, pages 519–527. AUAI Press, Corvallis, Oregon, 2002.

[20] Thomas S. Verma and Judea Pearl. Equivalence and synthesis of causal models. Technical Report R-150, Department of Computer Science, University of California, Los Angeles, 1990.

[21] Elie Wolfe, Robert AW. Spekkens, and Tobias Fritz. The inflation technique for causal inference with latent variables. `https://arxiv.org/abs/1609.00672`, 2016.