s-ID: Causal Effect Identification in a Sub-Population

Amir Mohammad Abouei*[1], Ehsan Mokhtarian*[1], Negar Kiyavash[1,2]

[1]School of Computer and Communication Sciences, EPFL, Lausanne, Switzerland
[2]College of Management of Technology, EPFL, Lausanne, Switzerland
{amir.abouei, ehsan.mokhtarian, negar.kiyavash}@epfl.ch

Abstract

Causal inference in a sub-population involves identifying the causal effect of an intervention on a specific subgroup, which is distinguished from the whole population through the influence of systematic biases in the sampling process. However, ignoring the subtleties introduced by sub-populations can either lead to erroneous inference or limit the applicability of existing methods. We introduce and advocate for a causal inference problem in sub-populations (henceforth called S-ID), in which we merely have access to observational data of the targeted sub-population (as opposed to the entire population). Existing inference problems in sub-populations operate on the premise that the given data distributions originate from the entire population, thus, cannot tackle the S-ID problem. To address this gap, we provide necessary and sufficient conditions that must hold in the causal graph for a causal effect in a sub-population to be identifiable from the observational distribution of that sub-population. Given these conditions, we present a sound and complete algorithm for the S-ID problem.

Introduction

In machine learning, variable(s) $\mathbf{Y}$ are commonly predicted from observed variable(s) $\mathbf{X}$ by estimating the conditional probability distribution $P(\mathbf{Y}|\mathbf{X})$ (Bishop and Nasrabadi 2006). This approach is effective for understanding correlations or associations in the data, but it falls short when we seek to understand how changes in $\mathbf{X}$ would affect $\mathbf{Y}$. Such an understanding requires a different methodology, known as causal inference (in population), which involves estimating the *interventional distributions (or causal effect)*, denoted by $P_{\mathbf{X}}(\mathbf{Y})$. $P_{\mathbf{X}}(\mathbf{Y})$ represents the probability of an outcome $\mathbf{Y}$ if we were to intervene or change the values of the input variable(s) $\mathbf{X}$ (Pearl 2000, 2009; Hernán and Robins 2010).[1]

The gold standard for estimating a causal effect is to perform experiments/interventions in the environment, for instance, by using techniques such as randomized controlled trials (RCTs) (Fisher 1936). However, these methods often require real-world experiments, which can be prohibitively expensive, unethical, or simply infeasible in many scenarios. Alternatively, researchers can turn to observational methods, utilizing the *causal graph* of the environment and available data to estimate interventional distributions (Pearl 2009; Spirtes et al. 2000). The causal graph, a graphical representation that depicts the causal relationships between variables, plays a central role in this methodology. This observational approach avoids the need for costly or impractical experiments but comes with its own challenges. In particular, computing interventional distributions uniquely may not always be feasible.

Identifiablity in population. Identifiability refers to the ability to uniquely compute a distribution from the available data. When all variables in the system are observable and the causal graph is known, all interventional distributions are identifiable using the so-called back-door adjustment sets, meaning all causal effects are identifiable (Pearl 1993). However, only a subset of causal effects can be identified in the presence of unobserved variables or hidden confounders (Pearl 1995). Selection bias can also make some causal effects unidentifiable (Shpitser and Pearl 2006a). This bias, which is similar to distribution mismatch in learning theory (Masiha et al. 2021), often arises from conditioning on selection variables. The problem of causal effect identification in population pertains to whether, given the causal graph, an interventional distribution can be uniquely computed from the available data. Various forms of available data lead to different problems in causal inference in population, the most well-known of which is the ID problem (Pearl 1995; Tian and Pearl 2003). This problem arises when the available data is from the joint distribution of the observed variables. A summary of these problems is provided in Table 1, and a more comprehensive discussion can be found in the Related Work section.

Conditional causal effects represent the conditional distributions that capture the impact of a treatment on the outcome within specific contexts or sub-populations. This concept allows for targeted interventions and tailored policies, offering valuable insights for practical applications (Qian and Murphy 2011). Shpitser and Pearl (2006a) considered the c-ID problem, which pertains to identifying a conditional interventional distribution $P_{\mathbf{X}}(\mathbf{Y}|\mathbf{Z})$ from the joint distribution

*Equal contribution.

[1]We utilize Judea Pearl's framework for causal inference to present our findings. Within this framework, alternative notations for interventional distributions include $P(\mathbf{Y}|\mathrm{do}(\mathbf{X}))$ and $P_{\mathrm{do}(\mathbf{X})}(\mathbf{Y})$, which employ the $do()$ operator to denote an intervention. Nevertheless, for the sake of simplicity in notation, we adopt the latter representation and drop the $do()$.

Causal inference problem		Given distribution(s)	Target interventional distribution
On Population	ID	$P(\mathbf{V})$	$P_{\mathbf{X}}(\mathbf{Y})$
	S-Recoverability	$P(\mathbf{V}\|S=1)$	$P_{\mathbf{X}}(\mathbf{Y})$
	gID	$\{P_{\mathbf{Z}_i}(\mathbf{V}\setminus\mathbf{Z}_i)\}_{i=0}^{m}$	$P_{\mathbf{X}}(\mathbf{Y})$
On Sub-Population	c-ID	$P(\mathbf{V})$	$P_{\mathbf{X}}(\mathbf{Y}\|\mathbf{Z})$
	c-gID	$\{P_{\mathbf{Z}_i}(\mathbf{V}\setminus\mathbf{Z}_i)\}_{i=0}^{m}$	$P_{\mathbf{X}}(\mathbf{Y}\|\mathbf{Z})$
	S-ID	$P(\mathbf{V}\|S=1)$	$P_{\mathbf{X}}(\mathbf{Y}\|S=1)$

Table 1: Various causal inference problems based on given and target distributions. Herein, $\mathbf{V}$ is the set of observed variables, $\mathbf{X}$ is the set of intervened variables, $\mathbf{Y}$ is the set of outcome variables, and $S=1$ corresponds to a sub-population. In this paper, we introduce the S-ID problem. Note that in all of these problems, the causal graph is given.

of observed variables.[2] An important practical limitation of the c-ID formulation is that it assumes access to samples from the observational distribution of the *entire population* rather than just the target sub-population. Unfortunately, the c-ID identification result cannot be directly extended to the setting where the available samples are from the target sub-population, which is often the prevailing scenario in practical applications. The recent extension of c-ID, known as c-gID problem (Correa, Lee, and Bareinboim 2021; Kivva, Etesami, and Kiyavash 2023), which we will discuss in Related Work, also suffers from the same practical limitation.

Identifiablity in sub-populations. As mentioned earlier, a sub-population is a specific subset of individuals within a larger population distinguished by certain characteristics or traits.[3] We utilize an auxiliary binary variable S to model a sub-population akin to Bareinboim and Tian (2015): S is added as a child variable representing the specific traits that distinguish the sub-population of the population (S can have several parents), and $S=1$ corresponds to the target sub-population. We will formally introduce the auxiliary variable S in Equation (1). In this paper, we address the problem of causal inference in a sub-population, where the objective is to identify $P_{\mathbf{X}}(\mathbf{Y}|S=1)$, which is the causal effect of a treatment or intervention on a specific subgroup of individuals within a larger population. Specifically, we introduce the S-*ID* problem, an identification problem on sub-population when we merely have access to observational data of the target sub-population. That is, given the causal graph, we seek to determine when $P_{\mathbf{X}}(\mathbf{Y}|S=1)$ can be uniquely computed from $P(\mathbf{V}|S=1)$, where $\mathbf{V}$ is the set of observed variables.

A real-world example. Consider the causal graphs depicted in Figure 1, where we analyze a hypothetical scenario in a random country. Here:

- The treatment variable X denotes whether smoking is banned in public areas.
- The mediator variable Z indicates the percentage of the population that smokes.

[2]In the notation $P_{\mathbf{X}}(\mathbf{Y}|\mathbf{Z})$, it is important to note the sequence of operations. The notation signifies that we first intervene on the set $\mathbf{X}$ and then, within the resulting distribution, condition on $\mathbf{Z}$.

[3]A sample in a sub-population is generated from a conditional distribution that determines the characteristics of the sub-population. This often introduces selection bias, as the sampling process might not be representative of the entire population.

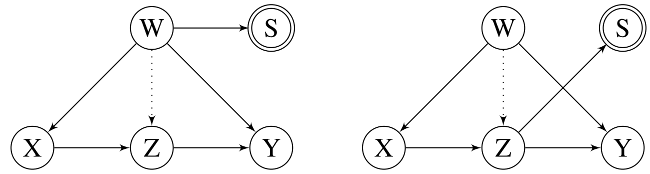


Figure 1: X: whether the public health policy bans smoking in public areas. Y: rate of lung cancer. Z: percentage of people who smoke. W: the average age of people. In the left causal graph, $P_X(Y|S=1)$ is S-ID, i.e., can be computed from $P(X,Y,Z,W|S=1)$, while it is not S-ID in the right causal graph.

- The outcome variable Y measures the rate of lung cancer.
- The confounder variable W captures the average age of the population.

Clearly X influences Z, and both Z and W affect Y. The relationship between W and X can be explained by the possibility that in countries with older populations, there may be greater awareness and concern about the health risks of smoking, potentially leading to stricter health policies such as public smoking bans. Additionally, one could argue that W may also have an impact on Z. Nevertheless, our subsequent analysis remains valid whether or not we consider a causal link between W and Z. Now, consider the scenario where the data from X, Y, Z, W is available from a subset of countries (sub-population) with younger populations than the world average. This scenario is illustrated in the left graph in Figure 1. The S-ID problem aims to identify the causal effect of a new policy X on the outcome variable Y for this target sub-population, given only observational data from this group. As we will demonstrate, this causal effect is *identifiable* and can be calculated using Algorithm 1.

In contrast, in the setting of the S-Recoverability problem, a causal inference problem in population (refer to the second row of Table 1), the task is to compute the causal effect of X on Y for the entire population using only data from this sub-population. The limitation of data coming only from the sub-population renders the inference for the whole population particularly challenging. Accordingly, Bareinboim and Tian (2015) showed that in this example, the causal effect of X on Y (in population) is unidentifiable. In the c-ID setting, the conditional causal effect of X on Y in sub-population is identifiable, but it requires observational data from the entire population, i.e., from all the countries in the world. Lastly, consider another scenario where the sub-population is based on a condition on the mediator variable Z rather than the confounder W (the right graph in Figure 1). An example of this scenario might involve a sub-population of countries that have had high smoking rates in recent years. Applying our Theorem 1, we can show that in this case, $P_X(Y|S=1)$ is *not identifiable* from $P(\mathbf{V}|S=1)$. Note that in the ID setting, $P_X(Y)$ is identifiable from $P(\mathbf{V})$. This shows that simply ignoring the sub-population and applying any algorithms in the ID setting leads to an erroneous inference.

The purpose of this example is to (i) demonstrate the critical role of causal graphs in whether a causal effect in a sub-population is identifiable or not and (ii) show that previous identification results in the literature do not suffice to answer the S-ID problem. An additional example is provided in Appendix A.

Our main contributions are as follows.

- We formally introduce the S-ID problem, a practical scenario for causal inference in a sub-population. This problem asks whether, given a causal graph, a causal effect in a sub-population can be uniquely computed from the observational distribution pertaining to that sub-population.
- We provide necessary and sufficient conditions on the causal graph for when a causal effect in a sub-population can be uniquely computed from the observational distribution of the same sub-population (Theorems 1 and 2).
- We propose a sound and complete algorithm for the S-ID problem (Algorithm 1).

Preliminaries

Throughout the paper, we denote random variables by capital letters and sets of variables by bold letters. We use $\sum_{\mathbf{X}}$ to denote marginalization, i.e., summation (or integration for continuous variables) over all the realizations of the variables in a set $\mathbf{X}$.

Let $\mathcal{G}$ be a directed acyclic graph (DAG) over a set of variables $\mathbf{V}$. We denote by $Pa_{\mathcal{G}}(X)$, $Ch_{\mathcal{G}}(X)$, and $Anc_{\mathcal{G}}(X)$ the set of parents, children, and ancestors of X (including X) in $\mathcal{G}$, respectively. We further define $Anc_{\mathcal{G}}(\mathbf{X}) = \bigcup_{X\in\mathbf{X}} Anc_{\mathcal{G}}(X)$ for a set $\mathbf{X} \subseteq \mathbf{V}$. A structural equation model (SEM) describes the dynamics of a system using a collection of equations

$$X = f_X(Pa_{\mathcal{G}}(X), \varepsilon_X), \quad \forall X \in \mathbf{V},$$

where $\mathcal{G}$ is the causal graph, f_X is a deterministic function and ε_X is the exogenous noise of variable X, which is independent of all the other exogenous noises. A SEM $\mathcal{M}$ with causal DAG $\mathcal{G}$ induces a unique joint distribution $P^{\mathcal{M}}(\mathbf{V})$ that satisfies Markov property with respect to $\mathcal{G}$. That is, $P^{\mathcal{M}}(\mathbf{V})$ can be factorized according to $\mathcal{G}$ as

$$P^{\mathcal{M}}(\mathbf{V}) = \prod_{X\in\mathbf{V}} P^{\mathcal{M}}(X|Pa_{\mathcal{G}}(X)).$$

We drop $\mathcal{M}$ from $P^{\mathcal{M}}(\cdot)$ when it is clear from the context. In this paper, an intervention on a set $\mathbf{X}$ converts $\mathcal{M}$ to a new SEM where the equations of the variables in $\mathbf{X}$ are replaced by some constants.[4] We denote the corresponding post-interventional distribution by $P_{\mathbf{X}}(\mathbf{V}\setminus\mathbf{X})$. The goal of causal inference in population is to compute an interventional distribution $P_{\mathbf{X}}(\mathbf{Y})$ for two disjoint subsets $\mathbf{X}$ and $\mathbf{Y}$ of $\mathbf{V}$.

Let $\mathbf{X}, \mathbf{Y}, \mathbf{W}$ be three disjoint subsets of $\mathbf{V}$. A path $\mathcal{P} = (X, Z_1, \ldots, Z_k, Y)$ between $X \in \mathbf{X}$ and $Y \in \mathbf{Y}$ in $\mathcal{G}$ is called *blocked* by $\mathbf{W}$ if there exists $1 \le i \le k$ such that

- Z_i is a collider[5] on $\mathcal{P}$ and $Z_i \notin Anc_{\mathcal{G}}(\mathbf{W})$, or
- Z_i is not a collider on $\mathcal{P}$ and $Z_i \in \mathbf{W}$.

Denoted by $(\mathbf{X} \perp\!\!\!\perp \mathbf{Y}|\mathbf{W})_{\mathcal{G}}$, we say $\mathbf{W}$ d-separates $\mathbf{X}$ and $\mathbf{Y}$ if for any $X \in \mathbf{X}$ and $Y \in \mathbf{Y}$, $\mathbf{W}$ blocks all the paths in $\mathcal{G}$ between X and Y. Conversely, $(\mathbf{X} \not\perp\!\!\!\perp \mathbf{Y}|\mathbf{W})_{\mathcal{G}}$ if there exists at least one *active path* between a variable in $\mathbf{X}$ and a variable in $\mathbf{Y}$ that is not blocked by $\mathbf{W}$.

The following three rules, commonly referred to as Pearl's do-calculus rules (Pearl 2000), provide a tool for calculating interventional distributions using the causal graph.

- **Rule 1**: If $(\mathbf{Y} \perp\!\!\!\perp \mathbf{Z}|\mathbf{X},\mathbf{W})_{\mathcal{G}_{\overline{X}}}$, then
$$P_{\mathbf{X}}(\mathbf{Y}|\mathbf{Z},\mathbf{W}) = P_{\mathbf{X}}(\mathbf{Y}|\mathbf{W}).$$
- **Rule 2**: If $(\mathbf{Y} \perp\!\!\!\perp \mathbf{Z}|\mathbf{X},\mathbf{W})_{\mathcal{G}_{\overline{\mathbf{X}},\underline{\mathbf{Z}}}}$, then
$$P_{\mathbf{X},\mathbf{Z}}(\mathbf{Y}|\mathbf{W}) = P_{\mathbf{X}}(\mathbf{Y}|\mathbf{Z},\mathbf{W}).$$
- **Rule 3**: If $(\mathbf{Y} \perp\!\!\!\perp \mathbf{Z}|\mathbf{X},\mathbf{W})_{\mathcal{G}_{\overline{\mathbf{X}},\overline{\mathbf{Z}(W)}}}$, where $\mathbf{Z}(\mathbf{W}) := \mathbf{Z}\setminus Anc_{\mathcal{G}_{\overline{\mathbf{X}}}}(\mathbf{W})$, then
$$P_{\mathbf{X},\mathbf{Z}}(\mathbf{Y}|\mathbf{W}) = P_{\mathbf{X}}(\mathbf{Y}|\mathbf{W}).$$

In these rules, $\mathcal{G}_{\overline{\mathbf{X}}\underline{\mathbf{Z}}}$ denotes the graph obtained by removing the incoming edges to $\mathbf{X}$ and outgoing edges from $\mathbf{Z}$.

The S-ID Problem

In this section, we start by discussing the integration of an auxiliary variable S into a SEM to model a sub-population and introduce the S-ID problem and formulate our objective. Subsequently, we provide necessary and sufficient graph conditions for the S-ID problem. Finally, given the causal DAG, we provide a sound and complete algorithm for computing $P_{\mathbf{X}}(\mathbf{Y}|S=1)$ from $P(\mathbf{V}|S=1)$, when this conditional causal effect is S-ID.

Modeling a Sub-Population: Auxiliary Variable S

Let $\mathcal{M}$ be a SEM with the set of variables $\mathbf{V}$, causal DAG $\mathcal{G}$, and observational distribution $P(\mathbf{V})$, representing the distribution of the entire *population*. That is, a sample is from the population if it is generated from $P(\mathbf{V})$. A *sub-population*, on the other hand, refers to a biased sampling mechanism. Formally, a sample is from a sub-population if it is generated from a conditional distribution $P(\mathbf{V}|S=1)$, in which

$$S := f_S(\mathbf{V}_S, \varepsilon_S), \tag{1}$$

[4]There are other types of interventions, such as soft-interventions, which are not considered herein.

[5]A non-endpoint vertex on a path is called a collider if both of the edges incident to it on the path point to it.

where f_S is a binary function that determines the characteristics or traits of the sub-population, $\mathbf{V}_S \subseteq \mathbf{V}$, and ε_S is the exogenous noise variable independent of the other exogenous variables. Under this modeling approach, $S = 1$ signifies that the sample is generated from a specific sub-population.

We denote by $\mathcal{G}^{\mathrm{s}}$ the augmented DAG obtained by adding S to $\mathcal{G}$, such that $Pa_{\mathcal{G}^{\mathrm{s}}}(S) = \mathbf{V}_S$, and S does not have any children. As a result, $\mathcal{G}^{\mathrm{s}}$ is the causal graph of the SEM obtained by adding S to the set of variables. Moreover, we define $P^{\mathrm{s}}(\mathbf{V}) := P(\mathbf{V}|S = 1)$, which is the observational distribution of the target sub-population. We often omit the graph subscript in $Pa_{\mathcal{G}}()$ and $Anc_{\mathcal{G}}()$ notations as parents and ancestor sets are identical in $\mathcal{G}$ and $\mathcal{G}^{\mathrm{s}}$.

Problem Formulation: Definition of S-ID

As mentioned earlier, $P(\mathbf{V}|S = 1)$ (or $P^{\mathrm{s}}(\mathbf{V})$) is the observational distribution of a sub-population. Furthermore, for two disjoint subsets $\mathbf{X}$ and $\mathbf{Y}$ of $\mathbf{V}$, $P_{\mathbf{X}}(\mathbf{Y}|S = 1)$ (or $P^{\mathrm{s}}_{\mathbf{X}}(\mathbf{Y})$) corresponds to the causal effect of $\mathbf{X}$ on $\mathbf{Y}$ in that sub-population. The problem of S-ID, formally defined in the following, considers the identifiability of $P_{\mathbf{X}}(\mathbf{Y}|S = 1)$ from $P(\mathbf{V}|S = 1)$.

Definition 1 (S-ID). *Suppose $\mathbf{X}$ and $\mathbf{Y}$ are disjoint subsets of a set $\mathbf{V}$, and let $\mathcal{G}^{\mathrm{s}}$ be the augmented causal graph of a SEM over $\mathbf{V} \cup \{S\}$. Conditional causal effect $P_{\mathbf{X}}(\mathbf{Y}|S = 1)$ is S-ID in $\mathcal{G}^{\mathrm{s}}$ if for any two SEMs $\mathcal{M}_1$ and $\mathcal{M}_2$ with causal graph $\mathcal{G}^{\mathrm{s}}$ such that $P^{\mathcal{M}_1}(\mathbf{V}|S = 1) = P^{\mathcal{M}_2}(\mathbf{V}|S = 1) > 0$, then $P^{\mathcal{M}_1}_{\mathbf{X}}(\mathbf{Y}|S = 1) = P^{\mathcal{M}_2}_{\mathbf{X}}(\mathbf{Y}|S = 1)$.*

In other words, this definition states that $P^{\mathrm{s}}_{\mathbf{X}}(\mathbf{Y})$ is S-ID when it can be uniquely computed from $P^{\mathrm{s}}(\mathbf{V})$.

In the rest of the paper, we address the following questions. Given an augmented causal DAG $\mathcal{G}^{\mathrm{s}}$ over a set $\mathbf{V} \cup \{S\}$ and for two disjoint subsets $\mathbf{X}$ and $\mathbf{Y}$ of $\mathbf{V}$,

- What are the necessary and sufficient conditions on $\mathcal{G}^{\mathrm{s}}$ such that $P^{\mathrm{s}}_{\mathbf{X}}(\mathbf{Y})$ is S-ID in $\mathcal{G}^{\mathrm{s}}$?
- When $P^{s}_{\mathbf{X}}(\mathbf{Y})$ is S-ID in $\mathcal{G}^{\mathrm{s}}$, how can we compute it from $P^{\mathrm{s}}(\mathbf{V})$?

To address the first question, for pedagogical reasons, we first consider the case where $\mathbf{X}$ and $\mathbf{Y}$ each contain only one variable. We subsequently extend our findings to the multivariate scenario. In the last subsection, we address the second question and propose a sound and complete algorithm for the S-ID problem.

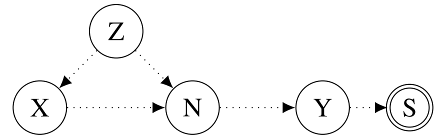
Conditions for s-Identifiability: Singleton Case

Suppose $\mathbf{X}$ and $\mathbf{Y}$ are singleton, where $\mathbf{X} = \{X\}$ and $\mathbf{Y} = \{Y\}$. The following theorem provides a necessary and sufficient condition for $P^{\mathrm{s}}_X(Y)$ to be S-ID in $\mathcal{G}^{\mathrm{s}}$.

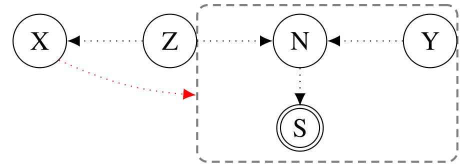
Theorem 1. *For two variables X and Y, conditional causal effect $P^{\mathrm{s}}_X(Y)$ is S-ID in DAG $\mathcal{G}^{\mathrm{s}}$ if and only if*

$$X \notin Anc(S) \quad or \quad (X \perp\!\!\!\perp Y|S)_{\mathcal{G}^{\mathrm{s}}_{\underline{X}}}. \tag{2}$$

Detailed proofs of our results appear in Appendices B and C. In the main text, we provide concise proof sketches to emphasize the key steps of our proofs.



(a) Type 1. N can coincide with Y.



(b) Type 2. N can coincide with S.

Figure 2: Two types of DAGs used in the proof of Theorem 1. The dotted edges indicate the presence of a directed path.

Sketch of proof. Sufficiency. Suppose Equation (2) holds. Applying do-calculus rules allows us to show the following cases.

- If $(X \perp\!\!\!\perp Y|S)_{\mathcal{G}^{\mathrm{s}}_{\underline{X}}}$, then $P^{\mathrm{s}}_X(Y) = P^{\mathrm{s}}(Y|X)$.
- If $X \notin Anc(S)$ and $Y \in Pa(X)$, then $P^{\mathrm{s}}_X(Y) = P^{\mathrm{s}}(Y)$.
- If $X \notin Anc(S)$ and $Y \notin Pa(X)$, then

$$P^{\mathrm{s}}_X(Y) = \sum_{Pa(X)} P^{\mathrm{s}}(Y|X, Pa(X))P^{\mathrm{s}}(Pa(X)).$$

Necessity. For the necessary part, which is the challenging part of the proof, we need to show that when $X \in Anc(S)$ and $(X \not\perp\!\!\!\perp Y|S)_{\mathcal{G}^{\mathrm{s}}_{\underline{X}}}$, then $P^{\mathrm{s}}_X(Y)$ is not S-ID in $\mathcal{G}^{\mathrm{s}}$. We first consider a special case where $Y \in Anc(X)$ and prove the following in the appendix.

Claim 1. *If $Y \in Anc(X)$ and $X \in Anc(S)$, then $P^{\mathrm{s}}_X(Y)$ is not S-ID in $\mathcal{G}^{\mathrm{s}}$.*

Accordingly, to complete the proof, suppose $Y \notin Anc(X)$.

Claim 2. *If $P^{\mathrm{s}}_X(Y)$ is not S-ID in a subgraph of $\mathcal{G}^{\mathrm{s}}$, then $P^{\mathrm{s}}_X(Y)$ is not S-ID in $\mathcal{G}^{\mathrm{s}}$.*

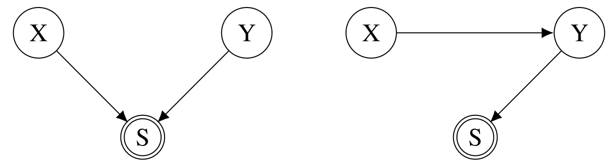
To prove the theorem using Claim 2, we first introduce a subgraph of $\mathcal{G}^{\mathrm{s}}$ and then show that $P^{\mathrm{s}}_X(Y)$ is not S-ID in that subgraph.

Claim 3. *There exists a path between X and Y in $\mathcal{G}^{\mathrm{s}}_{\underline{X}}$, which is not blocked by S, and it contains at most one collider.*

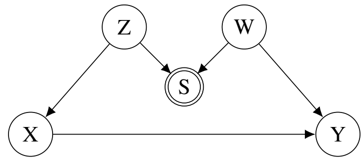
Denote by $\mathcal{P}$, a path between X and Y in $\mathcal{G}^{\mathrm{s}}_{\underline{X}}$ with the minimum number of colliders such that S does not block $\mathcal{P}$. Due to Claim 3, path $\mathcal{P}$ exists and has at most one collider.

Let $\mathcal{G}'$ be a minimal (in terms of edges) subgraph of $\mathcal{G}^{\mathrm{s}}$ such that (i) $\mathcal{G}'$ contains $\mathcal{P}$, (ii) $X \in Anc_{\mathcal{G}'}(S)$, and (iii) if $\mathcal{P}$ has exactly one collider, then the collider is an ancestor of S in $\mathcal{G}'$. Note that if $\mathcal{P}$ has a collider, then it is an ancestor of S in $\mathcal{G}^{\mathrm{s}}$ since S does not block $\mathcal{P}$. Thus, graph $\mathcal{G}'$ with these properties exists.

Figure 2 illustrates two types of DAGs, where the dotted edges indicate the presence of a directed path, and the directed paths do not share any edges. Variable N can coincide



(a) $P^{\mathrm{s}}_X(Y) = P^{\mathrm{s}}(Y|X)$



(b) $P^{\mathrm{s}}_X(Y) = \sum_Z P^{\mathrm{s}}(Y|X,Z)P^{\mathrm{s}}(Z)$

Figure 3: Three DAGs in which $P^{\mathrm{s}}_X(Y)$ is S-ID.

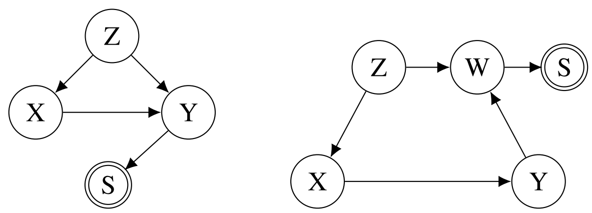


Figure 4: Two DAGs in which $P^{\mathrm{s}}_X(Y)$ is not S-ID.

with Y in Figure 2a and with S in Figure 2b. Furthermore, in Figure 2b, the directed path in red is towards a variable inside the box, i.e., the variables in the directed paths from Z to N (except Z itself), from N to S, and from Y to N.

In the appendix, we introduce a series of transformations to simplify $\mathcal{G}'$ and convert it to one of the two forms depicted in Figure 2. Denote by $\mathcal{G}''$, the DAG obtained by this conversion. This conversion ensures that if $P^{\mathrm{s}}_X(Y)$ is not S-ID in $\mathcal{G}''$, then it is not S-ID in $\mathcal{G}'$. Therefore, it suffices to show that $P^{\mathrm{s}}_X(Y)$ is not S-ID in $\mathcal{G}''$. To this end, in the appendix, we introduce two SEMs with causal graph $\mathcal{G}''$, denoted by $\mathcal{M}_1$ and $\mathcal{M}_2$, and show that $P^{\mathcal{M}_1}_X(Y|S=1) \neq P^{\mathcal{M}_2}_X(Y|S=1)$, while $P^{\mathcal{M}_1}(\mathbf{V}|S=1) = P^{\mathcal{M}_2}(\mathbf{V}|S=1) > 0$. This proves that $P^{\mathrm{s}}_X(Y)$ is not S-ID in $\mathcal{G}''$ and completes the proof. □

Figure 3 depicts three example graphs in which $P^{\mathrm{s}}_{\mathbf{X}}(\mathbf{Y})$ is S-ID. In both DAGs in Figure 3a, $(X \perp\!\!\!\perp Y|S)_{\mathcal{G}^{\mathrm{s}}_{\underline{X}}}$. As mentioned in the sketch of proof of Theorem 1, this implies that $P^{\mathrm{s}}_X(Y) = P^{\mathrm{s}}(Y|X)$. We note that in the left graph, X does not have any causal effect on Y in the population (i.e., $P_X(Y) = P(Y)$) since Y is not a descendent of X. However, X has causal effect on Y in the sub-population (i.e., $P^{\mathrm{s}}_X(Y) \neq P^{\mathrm{s}}(Y)$) due to the dependency of X and Y in P^{s}. In Figure 3b, since $X \notin Anc(S) = \{Z, W, S\}$ and $Y \notin Pa(X) = \{Z\}$, we have $P^{\mathrm{s}}_X(Y) = \sum_Z P^{\mathrm{s}}(Y|Z,X)P^{\mathrm{s}}(Z)$. Note that while $P_X(Y) = P(Y|X)$, $P^{\mathrm{s}}_X(Y) \neq P^{\mathrm{s}}(Y|X)$.

Remark 1. *These examples show that ignoring S and assuming that our available samples are generated from P (as opposed to P^{s}) might lead to erroneous inferences.*

Figure 4 depicts two DAGs in which $X \in Anc(S)$ and $(X \not\perp\!\!\!\perp Y|S)_{\mathcal{G}^{\mathrm{s}}_{\underline{X}}}$ (in the left graph $X \leftarrow Z \rightarrow Y$ and in the right graph $X \leftarrow Z \rightarrow W \leftarrow Y$ is an active path in $\mathcal{G}^{\mathrm{s}}_{\underline{X}}$). Hence, Equation (2) does not hold, and Theorem 1 implies that $P^{\mathrm{s}}_X(Y)$ is not S-ID.

Conditions for s-Identifiability: Multivariate Case

We present a necessary and sufficient condition for $P^{\mathrm{s}}_{\mathbf{X}}(\mathbf{Y})$ to be S-ID in DAG $\mathcal{G}^{\mathrm{s}}$ in the multivariate case. To do so, we decompose $\mathbf{X}$ into two parts: ancestors and non-ancestors of S. The following proposition demonstrates that the conditional causal effect of the latter portion of $\mathbf{X}$ on any other subset is always S-ID.

Proposition 1. *Suppose $\mathcal{G}^{\mathrm{s}}$ is an augmented DAG over $\mathbf{V} \cup \{S\}$, and let $\mathbf{X} \subsetneq \mathbf{V}$. For $\mathbf{X}_2 := \mathbf{X} \setminus Anc(S)$, conditional causal effect $P^{\mathrm{s}}_{\mathbf{X}_2}(\mathbf{V} \setminus \mathbf{X}_2)$ is S-ID in $\mathcal{G}^{\mathrm{s}}$ and can be computed from $P^{\mathrm{s}}(\mathbf{V})$ by*

$$P^{\mathrm{s}}(Anc(S) \setminus S) \prod_{W \in \mathbf{W}} P^{\mathrm{s}}(W|Pa(W)), \tag{3}$$

where $\mathbf{W} = \mathbf{V} \setminus (\mathbf{X}_2 \cup Anc(S))$.

Corollary 1. *For $\mathbf{Y} \subseteq \mathbf{V} \setminus \mathbf{X}_2$, conditional causal effect $P^{\mathrm{s}}_{\mathbf{X}_2}(\mathbf{Y})$ is S-ID in $\mathcal{G}^{\mathrm{s}}$ since*

$$P^{\mathrm{s}}_{\mathbf{X}_2}(\mathbf{Y}) = \sum_{\mathbf{V} \setminus (\mathbf{X}_2 \cup \mathbf{Y})} P^{\mathrm{s}}_{\mathbf{X}_2}(\mathbf{V} \setminus \mathbf{X}_2). \tag{4}$$

So far, we have shown that $P^{\mathrm{s}}_{\mathbf{X}_2}(\mathbf{V} \setminus \mathbf{X}_2)$ is always S-ID in $\mathcal{G}^{\mathrm{s}}$, where $\mathbf{X}_2 = \mathbf{X} \setminus Anc(S)$. The following theorem provides a necessary and sufficient condition for $P^{\mathrm{s}}_{\mathbf{X}}(\mathbf{Y})$ to be S-ID in $\mathcal{G}^{\mathrm{s}}$. When this condition holds, the following theorem presents a formula to compute $P^{\mathrm{s}}_{\mathbf{X}}(\mathbf{Y})$ in terms of $P^{\mathrm{s}}_{\mathbf{X}_2}(\mathbf{V} \setminus \mathbf{X}_2)$, which is always S-ID as established in Corollary 1.

Theorem 2. *For disjoint subsets $\mathbf{X}$ and $\mathbf{Y}$ of $\mathbf{V}$, let $\mathbf{X}_1 := \mathbf{X} \cap Anc(S)$ and $\mathbf{X}_2 := \mathbf{X} \setminus Anc(S)$.*

- *If $\mathbf{X}_1 = \varnothing$: Conditional causal effect $P^{\mathrm{s}}_{\mathbf{X}}(\mathbf{Y})$ is S-ID and can be computed from Equation (4).*
- *If $\mathbf{X}_1 \neq \varnothing$: Conditional causal effect $P^{\mathrm{s}}_{\mathbf{X}}(\mathbf{Y})$ is S-ID if and only if*

$$(\mathbf{X}_1 \perp\!\!\!\perp \mathbf{Y}|\mathbf{X}_2, S)_{\mathcal{G}^{\mathrm{s}}_{\underline{\mathbf{X}_1}\overline{\mathbf{X}_2}}}. \tag{5}$$

Moreover, when (5) holds, we have

$$P^{\mathrm{s}}_{\mathbf{X}}(\mathbf{Y}) = P^{\mathrm{s}}(\mathbf{X}_1)^{-1} \sum_{\mathbf{V} \setminus (\mathbf{X} \cup \mathbf{Y})} P^{\mathrm{s}}_{\mathbf{X}_2}(\mathbf{V} \setminus \mathbf{X}_2), \tag{6}$$

where $P^{\mathrm{s}}_{\mathbf{X}_2}(\mathbf{V} \setminus \mathbf{X}_2)$ can be computed from $P^{\mathrm{s}}(\mathbf{V})$ using Equation (3).

Corollary 2. *Conditional causal effect $P^{\mathrm{s}}_{\mathbf{X}}(\mathbf{Y})$ is S-ID in $\mathcal{G}^{\mathrm{s}}$ if and only if*

$$\mathbf{X}_1 = \varnothing \quad or \quad (\mathbf{X}_1 \perp\!\!\!\perp \mathbf{Y}|\mathbf{X}_2, S)_{\mathcal{G}^{\mathrm{s}}_{\underline{\mathbf{X}_1}\overline{\mathbf{X}_2}}}. \tag{7}$$

Furthermore, if $\mathbf{X} = \{X\}$ is singleton, then $\mathbf{X}_1 = \varnothing$, which is equivalent to $X \notin Anc(S)$ and Theorem 2 reduces to Theorem 1 for the singleton case.

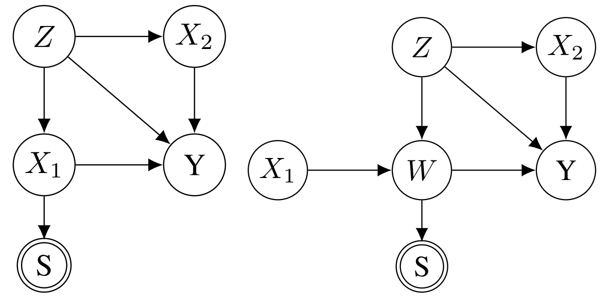


Figure 5: An example for the multivariate case where conditional causal effect $P^{\mathrm{s}}_{\{X_1,X_2\}}(Y)$ is not S-ID in the left graph while it is S-ID in the right graph and is equal to $\sum_{Z,W} P^{\mathrm{s}}(Z,W|X_1)P^{\mathrm{s}}(Y|X_2,Z,W)$.

Sketch of proof. The first part of the theorem (if $\mathbf{X}_1 = \varnothing$) is a direct consequence of Proposition 1. To show the second part, we assume $\mathbf{X}_1 \neq \varnothing$.

Sufficiency. Suppose Equation (5) holds. We need to show that Equation (6) holds. By applying Rules 2 and 3 of do-calculus, it can be shown that

$$P^{\mathrm{s}}_{\mathbf{X}}(\mathbf{Y}) = P^{\mathrm{s}}_{\mathbf{X}_2}(\mathbf{Y}|\mathbf{X}_1) = \frac{P^{\mathrm{s}}_{\mathbf{X}_2}(\mathbf{X}_1,\mathbf{Y})}{P^{\mathrm{s}}_{\mathbf{X}_2}(\mathbf{X}_1)} = \frac{P^{\mathrm{s}}_{\mathbf{X}_2}(\mathbf{X}_1,\mathbf{Y})}{P^{\mathrm{s}}(\mathbf{X}_1)}.$$

Moreover, Corollary 1 for $\mathbf{X}_1 \cup \mathbf{Y}$ implies that

$$P^{\mathrm{s}}_{\mathbf{X}_2}(\mathbf{X}_1,\mathbf{Y}) = \sum_{\mathbf{V}\setminus(\mathbf{X}\cup\mathbf{Y})} P^{\mathrm{s}}_{\mathbf{X}_2}(\mathbf{V}\setminus\mathbf{X}_2).$$

Equation (6) can be obtained by merging the above equations.

Necessity. Suppose $(\mathbf{X}_1 \not\perp\!\!\!\perp \mathbf{Y}|\mathbf{X}_2,S)_{\mathcal{G}^{\mathrm{s}}_{\underline{\mathbf{X}_1}\overline{\mathbf{X}_2}}}$. We need to show that $P^{\mathrm{s}}_{\mathbf{X}}(\mathbf{Y})$ is not S-ID in $\mathcal{G}^{\mathrm{s}}$.

Claim 4. *There exists $X^* \in \mathbf{X}_1$, $Y^* \in \mathbf{Y}$, and a subgraph $\mathcal{G}^*$ of $\mathcal{G}^{\mathrm{s}}$ such that*

- $\mathbf{X} \cap \mathit{Anc}_{\mathcal{G}^*_{\overline{X^*}}}(S) = \{X^*\}$,
- $(X^* \not\perp\!\!\!\perp Y^*|S)_{\mathcal{G}^*_{\underline{X^*}}}$, *and*
- $(\mathbf{X}\setminus\{X^*\} \perp\!\!\!\perp Y^*|X^*,S)_{\mathcal{G}^*_{\overline{\mathbf{X}}}}$.

The first property implies that $X^* \in \mathit{Anc}_{\mathcal{G}^*}(S)$. Hence, Equation (2) holds for X^* and Y^* in $\mathcal{G}^*$ and Theorem 1 implies that $P^{\mathrm{s}}_{X^*}(Y^*)$ is not S-ID in $\mathcal{G}^*$. To conclude the proof, similar to the sketch of proof of Theorem 1, it suffices to show that $P^{\mathrm{s}}_{\mathbf{X}}(\mathbf{Y})$ is not S-ID in $\mathcal{G}^*$.

For any SEM with causal graph $\mathcal{G}^*$, due to the first and third properties in Claim 4, Rule 3 of do-calculus implies that $P^{s}_{X^*}(Y^*) = P^{s}_{\mathbf{X}}(Y^*)$. Therefore, in $\mathcal{G}^*$, the s-identifiability of $P^{s}_{X^*}(Y^*)$ is equivalent to s-identifiability of $P^{s}_{\mathbf{X}}(Y^*)$, thus $P^{s}_{\mathbf{X}}(Y^*)$ is not S-ID in $\mathcal{G}^*$. This shows that $P^{s}_{\mathbf{X}}(\mathbf{Y})$ is also not S-ID in $\mathcal{G}^*$ since $Y^* \in \mathbf{Y}$, which concludes our proof. □

Consider the two DAGs in Figure 5, where we are interested in computing $P^{\mathrm{s}}_{\mathbf{X}}(\mathbf{Y})$ for $\mathbf{X} = \{X_1,X_2\}$ and $\mathbf{Y} = \{Y\}$. Accordingly, we have $\mathbf{X}_1 = \mathbf{X} \cap \mathit{Anc}(S) = \{X_1\}$ and $\mathbf{X}_2 = \mathbf{X} \setminus \mathit{Anc}(S) = \{X_2\}$ for both DAGs. In the DAG on the left, $(X_1 \not\perp\!\!\!\perp Y|X_2,S)_{\mathcal{G}^{\mathrm{s}}_{\underline{X_1}\overline{X_2}}}$ since $X_1 \leftarrow Z \rightarrow Y$ is an active path. Hence, Theorem 2 implies that $P^{\mathrm{s}}_{\{X_1,X_2\}}(Y)$ is not S-ID. On the other hand, $P^{\mathrm{s}}_{\{X_1,X_2\}}(Y)$ is S-ID in the DAG on the right since $(X_1 \perp\!\!\!\perp Y|X_2,S)_{\mathcal{G}^{\mathrm{s}}_{\underline{X_1}\overline{X_2}}}$. Moreover, $P^{\mathrm{s}}_{X_2}(Y) = P^{\mathrm{s}}(X_1,Z,W)P^{\mathrm{s}}(Y|X_2,Z,W)$ due to Proposition 1, thus, $P^{\mathrm{s}}_{\{X_1,X_2\}}(Y) = \sum_{Z,W} P^{\mathrm{s}}(Z,W|X_1)P^{\mathrm{s}}(Y|X_2,Z,W)$. Note that $P^{\mathrm{s}}(X_1)^{-1}P^{\mathrm{s}}(X_1,Z,W) = P^{\mathrm{s}}(Z,W|X_1)$.

Algorithm 1: A sound and complete algorithm for S-ID

1: **Input:** $\mathbf{X},\mathbf{Y},\mathcal{G}^{\mathrm{s}},P^{\mathrm{s}}(\mathbf{V})$
2: **Output:** A formula for $P^{\mathrm{s}}_{\mathbf{X}}(\mathbf{Y})$ based on $P^{\mathrm{s}}(\mathbf{V})$ if it is S-ID, otherwise, FAIL
3: $\mathbf{X}_1 \leftarrow \mathbf{X} \cap \mathit{Anc}(S)$
4: $\mathbf{X}_2 \leftarrow \mathbf{X} \setminus \mathit{Anc}(S)$
5: $\mathbf{W} \leftarrow \mathbf{V} \setminus (\mathbf{X}_2 \cup \mathit{Anc}(S))$
6: **if** $\mathbf{X}_1 = \varnothing$ **then**
7:     **Return** $\sum_{\mathbf{V}\setminus(\mathbf{X}\cup\mathbf{Y})} P^{\mathrm{s}}(\mathit{Anc}(S)\setminus S) \prod_{W\in\mathbf{W}} P^{\mathrm{s}}(W|\mathit{Pa}(W))$
8: **else if** $(\mathbf{X}_1 \perp\!\!\!\perp \mathbf{Y}|\mathbf{X}_2,S)_{\mathcal{G}^{\mathrm{s}}_{\underline{\mathbf{X}_1}\overline{\mathbf{X}_2}}}$ **then**
9:     **Return** $\sum_{\mathbf{V}\setminus(\mathbf{X}\cup\mathbf{Y})} \frac{P^{\mathrm{s}}(\mathit{Anc}(S)\setminus S)}{P^{\mathrm{s}}(\mathbf{X}_1)} \prod_{W\in\mathbf{W}} P^{\mathrm{s}}(W|\mathit{Pa}(W))$
10: **else**
11:     **Return** FAIL
12: **end if**

A Sound And Complete Algorithm For S-ID

Equipped by Proposition 1 and Theorem 2, we present Algorithm 1 for the S-ID problem.[6] The inputs are the set of intervened variables $\mathbf{X}$, the set of outcome variables $\mathbf{Y}$, augmented causal DAG $\mathcal{G}^{\mathrm{s}}$, and the observational distribution of the target sub-population $P^{\mathrm{s}}(\mathbf{V})$. The algorithm returns a formula for conditional causal effect $P^{\mathrm{s}}_{\mathbf{X}}(\mathbf{Y})$ based on $P^{\mathrm{s}}(\mathbf{V})$ when it is S-ID in $\mathcal{G}^{\mathrm{s}}$. Otherwise, it returns FAIL which indicates that $P^{\mathrm{s}}_{\mathbf{X}}(\mathbf{Y})$ is not S-ID in $\mathcal{G}^{\mathrm{s}}$.

The algorithm starts by decomposing $\mathbf{X}$ to ancestors ($\mathbf{X}_1$) and non-ancestors ($\mathbf{X}_2$) of S. If $\mathbf{X}_1 = \varnothing$, due to the first part of Theorem 2, $P^{\mathrm{s}}_{\mathbf{X}}(\mathbf{Y})$ is S-ID and the algorithm returns Equation (4) by replacing $P^{\mathrm{s}}_{\mathbf{X}_2}(\mathbf{V}\setminus\mathbf{X}_2)$ with Equation (3). Otherwise (i.e., when $\mathbf{X}_1 \neq \varnothing$), the algorithm checks the s-identifiability condition of Equation (5) in line 8. If the condition holds, it returns Equation (6), again, by replacing $P^{\mathrm{s}}_{\mathbf{X}_2}(\mathbf{V}\setminus\mathbf{X}_2)$ with Equation (3). If the condition does not hold, Theorem 2 implies that $P^{\mathrm{s}}_{\mathbf{X}}(\mathbf{Y})$ is not S-ID, and the algorithm returns FAIL.

Corollary 3. *Algorithm 1 is sound and complete for the S-ID problem. That is, when $P^{\mathrm{s}}_{\mathbf{X}}(\mathbf{Y})$ is S-ID in $\mathcal{G}^{\mathrm{s}}$, it returns a sound formula for it based on $P^{\mathrm{s}}(\mathbf{V})$ (soundness), and otherwise, it returns* FAIL *(completeness).*

We can use efficient methods such as the one presented Darwiche (2009) to verify the d-separation in Line 8 of Algorithm 1. Accordingly, the time complexity of the algorithm

[6]Our implementation is at https://github.com/amabouei/s-ID.

is $\mathcal{O}(n+m)$, where n and m represent the number of nodes and edges in the graph, respectively.

Related Work

In this section, we review related problems in the causal inference literature. A summary of the settings for these problems can be found in Table 1.

Causal Inference in Population

The goal of causal inference in the entire population is to compute a causal effect $P_{\mathbf{X}}(\mathbf{Y})$. The seminal ID problem (Pearl 1995), proposed by Judea Pearl, is concerned with calculating $P_{\mathbf{X}}(\mathbf{Y})$ based on observational distribution $P(\mathbf{V})$ when the causal graph is known. Pearl proposed three fundamental rules known as do-calculus, which, along with probabilistic manipulations, can be used to compute interventional distributions. Applying these rules, Tian and Pearl (2003) proposed an algorithm for the ID problem, and later, Shpitser and Pearl (2006b) and Huang and Valtorta (2006) concurrently and with two different approaches showed that the proposed algorithm is sound and complete for the ID problem. The former introduced a graph structure called *Hedge* and showed that the existence of a hedge is equivalent to the non-identifiability of an interventional distribution in the setting of the ID problem. The latter showed that the identifiability of a causal effect $P_{\mathbf{X}}(\mathbf{Y})$ is equivalent to the identifiability of $Q[\mathbf{Z}] := P_{\mathbf{V}\setminus\mathbf{Z}}(\mathbf{Z})$ for $\mathbf{Z} = Anc_{\mathcal{G}_{\underline{\mathbf{X}}}}(\mathbf{Y})$. They then showed that the recursive algorithm by Tian and Pearl (2003) is sound and complete for the identifiability of Q distributions.

A more generalized formulation of the ID problem is known as gID or general identifiability (Lee, Correa, and Bareinboim 2019; Kivva et al. 2022). Similar to the ID problem, the goal in gID is to compute a causal effect $P_{\mathbf{X}}(\mathbf{Y})$ but from $\{P_{\mathbf{Z}_i}(\mathbf{V} \setminus \mathbf{Z}_i)\}_{i=0}^m$ for some subsets $\{\mathbf{Z}_i\}_{i=0}^m$ of observed variables. Hence, ID is a special case of gID when $m = 0$ and $\mathbf{Z}_0 = \varnothing$. Kivva et al. (2022) extended the approach of Huang and Valtorta (2006) for the ID problem to gID and proposed a sound and complete algorithm for gID.

Another problem in causal inference on population is the so-called S-Recoverability (Bareinboim, Tian, and Pearl 2014; Bareinboim and Tian 2015; Correa, Tian, and Bareinboim 2019). In contrast to ID and gID, the given distribution in S-Recoverability originates from a sub-population, yet the aim remains to calculate a causal effect for the entire population. The constraint of having data from merely a sub-population makes the inference task for the whole population particularly challenging. Consequently, it is plausible to anticipate that a majority of causal effects would be unidentifiable, a fact that inherently restricts the practical applicability of the S-Recoverability problem.

Causal Inference in a Sub-Population

Shpitser and Pearl (2006a) tackled the c-ID problem by proposing a sound and complete algorithm for computing a conditional causal effect $P_{\mathbf{X}}(\mathbf{Y}|\mathbf{Z})$ from observational distribution $P(\mathbf{V})$. They showed that $\mathbf{Z}$ can be decomposed into two parts, namely $\mathbf{Z}_1$ and $\mathbf{Z}_2$, such that $P_{\mathbf{X}}(\mathbf{Y}|\mathbf{Z})$ is c-ID if and only if $P_{\mathbf{X}_1,\mathbf{Z}_1}(\mathbf{Y},\mathbf{Z}_2)$ is ID. Hence, solving c-ID can be reduced to solving an ID problem. Similar to gID that generalizes the ID problem, Correa, Lee, and Bareinboim (2021) and Kivva, Etesami, and Kiyavash (2023) generalized c-ID to c-gID. The objective in c-gID is again the computation of a conditional causal effect $P_{\mathbf{X}}(\mathbf{Y}|\mathbf{Z})$, but from a set of interventional distributions of form $\{P_{\mathbf{Z}_i}(\mathbf{V} \setminus \mathbf{Z}_i)\}_{i=0}^m$ instead of merely the observational distribution.

Both the c-ID and c-gID settings operate on the premise that the given distributions originate from the entire population. Thus, to make an inference for a target sub-population, they require samples from the whole population (observational distribution in the case of c-ID and interventional distributions for c-gID). By contrast, the S-ID problem uses the observational distribution merely from the target sub-population.

Causal Graph Variations

In all the aforementioned causal inference problems, the causal graph is assumed to be given. Although many causal discovery algorithms, such as the ones proposed by Colombo et al. (2012); Claassen, Mooij, and Heskes (2013); Bernstein et al. (2020); Akbari et al. (2021); Huang et al. (2022); Mokhtarian et al. (2021, 2023), aim to learn the causal graph using observational distribution, the causal graph is only identifiable up to the so-called Markov equivalence class (Spirtes et al. 2000; Pearl 2009). Addressing this gap, Jaber, Zhang, and Bareinboim (2019) and Jaber et al. (2022) provided algorithms for the ID and c-ID problems, respectively, where instead of the causal graph, a partial ancestral graph (PAG) that represents the equivalence class of the causal graph is known. Akbari et al. (2023) consider the ID problem when the underlying graph is probabilistically defined. Tikka, Hyttinen, and Karvanen (2019) and Mokhtarian et al. (2022) consider a scenario for the ID problem where additional information about the causal graph is available in the form of context-specific independence (CSI) relations. They show that this side information renders more causal effects identifiable.

Conclusion and Future Work

We introduced S-ID, a practical scenario for causal inference in a sub-population. The S-ID problem asks whether, given the causal graph, a causal effect in a sub-population can be identified from the observational distribution pertaining to the same sub-population. We provided a sound and complete algorithm for the S-ID problem. While previous work, such as the c-ID and S-Recoverability problems, provide considerable insights, they cannot solve the S-ID problem. Indeed through various examples, we demonstrated that ignoring the subtleties introduced by sub-populations in causal modeling can lead to erroneous inferences in the S-ID problem.

Our current framework assumes that all variables in the sub-population are observable. We acknowledge the potential practical situations where this may not be the case. Investigating the S-ID problem in the presence of latent variables is an important future direction. Furthermore, to numerically estimate a causal effect, three key phases are involved: identification, estimation, and sensitivity analysis. This paper has addressed the identification problem, establishing a foundation for further research in the other two critical phases.

Acknowledgments

This research was in part supported by the Swiss National Science Foundation under NCCR Automation, grant agreement 51NF40_180545 and Swiss SNF project 200021_204355 /1.

References

Akbari, S.; Jamshidi, F.; Mokhtarian, E.; Vowels, M. J.; Etesami, J.; and Kiyavash, N. 2023. Causal Effect Identification in Uncertain Causal Networks. In *Thirty-seventh Conference on Neural Information Processing Systems*.

Akbari, S.; Mokhtarian, E.; Ghassami, A.; and Kiyavash, N. 2021. Recursive causal structure learning in the presence of latent variables and selection bias. *Advances in Neural Information Processing Systems*, 34: 10119–10130.

Bareinboim, E.; and Tian, J. 2015. Recovering Causal Effects from Selection Bias. *Proceedings of the AAAI Conference on Artificial Intelligence*, 29(1).

Bareinboim, E.; Tian, J.; and Pearl, J. 2014. Recovering from Selection Bias in Causal and Statistical Inference. *Proceedings of the AAAI Conference on Artificial Intelligence*, 28(1).

Bernstein, D.; Saeed, B.; Squires, C.; and Uhler, C. 2020. Ordering-based causal structure learning in the presence of latent variables. In *International Conference on Artificial Intelligence and Statistics*, 4098–4108. PMLR.

Bishop, C. M.; and Nasrabadi, N. M. 2006. *Pattern recognition and machine learning*, volume 4. Springer.

Claassen, T.; Mooij, J. M.; and Heskes, T. 2013. Learning Sparse Causal Models is not NP-hard. In *In Proceedings of the 29th Annual Conference on Uncertainty in Artificial Intelligence (UAI)*, 172–181.

Colombo, D.; Maathuis, M. H.; Kalisch, M.; and Richardson, T. S. 2012. Learning high-dimensional directed acyclic graphs with latent and selection variables. *The Annals of Statistics*, 294–321.

Correa, J.; Lee, S.; and Bareinboim, E. 2021. Nested counterfactual identification from arbitrary surrogate experiments. *Advances in Neural Information Processing Systems*, 34: 6856–6867.

Correa, J. D.; Tian, J.; and Bareinboim, E. 2019. Identification of Causal Effects in the Presence of Selection Bias. *Proceedings of the AAAI Conference on Artificial Intelligence*, 33(01): 2744–2751.

Darwiche, A. 2009. *Modeling and reasoning with Bayesian networks*. Cambridge university press.

Fisher, R. A. 1936. Design of experiments. *British Medical Journal*, 1(3923): 554.

Hernán, M. A.; and Robins, J. M. 2010. Causal inference.

Huang, B.; Low, C. J. H.; Xie, F.; Glymour, C.; and Zhang, K. 2022. Latent hierarchical causal structure discovery with rank constraints. *Advances in Neural Information Processing Systems*, 35: 5549–5561.

Huang, Y.; and Valtorta, M. 2006. Identifiability in causal bayesian networks: A sound and complete algorithm. In *AAAI*, 1149–1154.

Jaber, A.; Ribeiro, A.; Zhang, J.; and Bareinboim, E. 2022. Causal Identification under Markov equivalence: Calculus, Algorithm, and Completeness. In Koyejo, S.; Mohamed, S.; Agarwal, A.; Belgrave, D.; Cho, K.; and Oh, A., eds., *Advances in Neural Information Processing Systems*, volume 35, 3679–3690. Curran Associates, Inc.

Jaber, A.; Zhang, J.; and Bareinboim, E. 2019. Causal Identification under Markov Equivalence: Completeness Results. In Chaudhuri, K.; and Salakhutdinov, R., eds., *Proceedings of the 36th International Conference on Machine Learning*, volume 97 of *Proceedings of Machine Learning Research*, 2981–2989. PMLR.

Kivva, Y.; Etesami, J.; and Kiyavash, N. 2023. On Identifiability of Conditional Causal Effects. *The 39th Conference on Uncertainty in Artificial Intelligence*.

Kivva, Y.; Mokhtarian, E.; Etesami, J.; and Kiyavash, N. 2022. Revisiting the general identifiability problem. In *Uncertainty in Artificial Intelligence*, 1022–1030. PMLR.

Lee, S.; Correa, J. D.; and Bareinboim, E. 2019. General Identifiability with Arbitrary Surrogate Experiments. In *Conference on Uncertainty in Artificial Intelligence*.

Masiha, M. S.; Gohari, A.; Yassaee, M. H.; and Aref, M. R. 2021. Learning under distribution mismatch and model misspecification. In *2021 IEEE International Symposium on Information Theory (ISIT)*, 2912–2917. IEEE.

Mokhtarian, E.; Akbari, S.; Ghassami, A.; and Kiyavash, N. 2021. A recursive markov boundary-based approach to causal structure learning. In *The KDD'21 Workshop on Causal Discovery*, 26–54. PMLR.

Mokhtarian, E.; Jamshidi, F.; Etesami, J.; and Kiyavash, N. 2022. Causal effect identification with context-specific independence relations of control variables. In *International Conference on Artificial Intelligence and Statistics*, 11237–11246. PMLR.

Mokhtarian, E.; Khorasani, M.; Etesami, J.; and Kiyavash, N. 2023. Novel ordering-based approaches for causal structure learning in the presence of unobserved variables. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 37, 12260–12268.

Pearl, J. 1993. [Bayesian analysis in expert systems]: comment: graphical models, causality and intervention. *Statistical Science*, 8(3): 266–269.

Pearl, J. 1995. Causal diagrams for empirical research. *Biometrika*, 82(4): 669–688.

Pearl, J. 2000. Causality: Models, reasoning and inference. *Cambridge, UK: CambridgeUniversityPress*, 19(2): 3.

Pearl, J. 2009. *Causality*. Cambridge university press.

Qian, M.; and Murphy, S. A. 2011. Performance guarantees for individualized treatment rules. *Annals of statistics*, 39(2): 1180.

Shpitser, I.; and Pearl, J. 2006a. Identification of conditional interventional distributions. *Proceedings of the 22nd Conference on Uncertainty in Artificial Intelligence*.

Shpitser, I.; and Pearl, J. 2006b. Identification of joint interventional distributions in recursive semi-Markovian causal models. In *Proceedings of the National Conference on Artificial Intelligence*, volume 21, 1219. Menlo Park, CA; Cambridge, MA; London; AAAI Press; MIT Press; 1999.

Spirtes, P.; Glymour, C. N.; Scheines, R.; and Heckerman, D. 2000. *Causation, prediction, and search*. MIT press.

Tian, J.; and Pearl, J. 2003. On the identification of causal effects. Technical report, Department of Computer Science, University of California.

Tikka, S.; Hyttinen, A.; and Karvanen, J. 2019. Identifying causal effects via context-specific independence relations. *Advances in neural information processing systems*, 32.