Identification and Estimation of Direct Causal Effects

Emily R. Flanagan

Advisor: Emilija Perković

June 2020

Abstract

Scientific research is often concerned with questions of cause and effect. For example, does smoking cause lung cancer? Ideally, such questions are answered by randomized controlled experiments. Often such experiments are costly or impossible to conduct; it would be unethical to force a group of people to smoke cigarettes if it possible that cigarettes are carcinogenic. In many cases, the only data available is observational. We use graphical models to represent causal relationships between variables. Under the assumption of no hidden variables, a directed acyclic graph (DAG) fully characterizes the causal system. When the causal DAG is known, the direct causal effect of variable X on variable Y can be defined as a controlled, or natural direct effect. We introduce a new type of direct causal effect: the average natural direct effect. Furthermore, we explore parametric assumptions that make some of these effects equivalent.

Using observational data alone, one cannot alway learn the causal DAG. We commonly obtain more ambiguous graphical structures in the form of a completed partially directed acyclic graph (CPDAG), or the maximally oriented partially directed acyclic graph (MPDAG). Both CPDAGs and MPDAGs represent equivalence classes of DAGs that contain the true causal DAG. We give identifiability conditions for the controlled direct effect given a CPDAG or MPDAG and compare these conditions to the conditions of identifiability of a total causal effect. Under linearity we also give an identifiability condition for the average natural direct effect.

Contents

1 Introduction

During introductory statistics we learn that "correlation does not imply causation." Using classical statistical tools, often the strongest conclusion one can come to is a general statement about association between variables. Yet, frequently the questions posed by researchers are not ones of affiliation, but causality.

The gold standard for determining a causal relationship is the randomized experiment. Yet, large-scale experiments can be prohibitively expensive or impossible to conduct. For instance, it would be unethical to randomly assign 10,000 people to smoke cigarettes and 10,000 people to abstain from smoking if it is possible that cigarettes are carcinogenic. In these scenarios, we can only obtain *observational data* which may contain a slough of *confounding variables* that can mislead analyses if not recognized. Causal effects, in general, cannot be identified from observational data alone. Instead, they require a set of specific assumptions. Numerous approaches to this problem have emerged over the last century, but with interest in causality only peaking in the past few decades, there is more territory left to be explored.

This paper focuses on identifying direct causal effects on the basis of causal graphical models. In Section 2, we introduce background definitions that make up the foundation of causal graphical models. In Section 3, we turn our attention to the calculus of intervention which serves as a tool to indicate when causal effects are identifiable. In Section 4, we define direct causal effects and explore their identifiability when the causal direction of all edges in the graph are known. Section 5 delves into identifiability when the causal directions of some edges in the causal graph are unknown. Finally, in Section 6, we enumerate further directions for research into identifying and estimating direct causal effects.

2 A Crash Course in Causal Models

In this section, we follow Kalisch and Bühlmann (2014) with additional background from Pearl (2009) and Perković et al. (2017).

2.1 Structural Causal Models

The relationships we seek to investigate in causal inference can be represented using *structural causal models*. Structural causal models (SCMs) define the distribution of a single variable as a function of the other variables that are its direct causes, and some random noise. For example, consider the set of three structural equations describing the distributions of variables X, Y and Z described in (1).

$$\begin{aligned} X &\leftarrow f_x(u_x) \\ Z &\leftarrow f_z(X,\, u_z) \\ Y &\leftarrow f_y(Z,\, X,\, u_y), \end{aligned} \tag{1}$$

The variables u_y, u_x, u_z indicate noise variables from natural variation, or omitted factors. They are assumed to be mutually independent and with mean zero. In this SCM, X is generated as a function of the random noise variable u_x, whereas the Z is generated as a function of both X and u_z, and Y is generated as a function of Z, X and u_y.

We use $\leftarrow$ instead of $=$ in SCMs such as those in (1) to indicate that the causal relationship is asymmetric. For example, if we were given that $X \leftarrow 2Z + 4$, we would interpret this as "X is generated by multiplying Z by 2 and adding 4," but we could not claim that "Z is generated by dividing X half in hand subtracting 2." Moreover, structural equations are assumed to be autonomous: a change in the generating process of one variable has no effect on the generating process of another.

2.2 Graphs

Graphs are a useful tool for visualizing SCMs. A *graph* $\mathcal{G} = (\mathbf{V},\, \mathbf{E})$ consists of a set of nodes $\mathbf{V} = \{X_1, \ldots, X_p\}$, $p \geq 1$, and a set of edges $\mathbf{E}$. We use capital letters (e.g. X) to denote both the nodes in the graph as well as the random variables that these nodes represent. Bold notation (e.g. $\mathbf{X}$) denotes a set of nodes and their corresponding random vector. As per convention, we do not include the error terms in the causal graph. For example, the SCM given in (1) corresponds to the following graph:

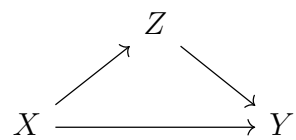

Figure 1: Causal DAG $\mathcal{G}$ representing representing the structure in (1)

We consider *simple* graphs that contain only one edge between each pair of nodes. The edges between the nodes are allowed to be directed ($\rightarrow$) and undirected ($-$). A directed edge $X \rightarrow Y$ indicates that X is a *direct cause* (see definitions in Section 4) whereas an undirected edge $X - Y$ indicates that either $X \rightarrow Y$, or $Y \rightarrow X$. While it would be optimal to always work with a *directed graph* that only contains directed edges, reality is often more challenging; the generating relationships of the variables frequently remain unknown from observational data, and we are left with with a *partially directed graph* that contains both directed and undirected edges.

2.3 Paths and Friends

The effects we investigate in this paper are indicated by *paths*. Paths tell us how variables are related to each other, either directly or indirectly. A path p from X to Y in $\mathcal{G}$ is a sequence of distinct nodes $\langle X = V_1, \ldots, V_k, Y \rangle$, $k \geq 1$ in which every pair of successive nodes is *adjacent*, meaning they have an edge between them. For example, in Figure 1, $\langle Z, X, Y \rangle$ and $\langle Z, Y \rangle$ are two paths from Z to Y. A path consisting of undirected edges is an *undirected* path. For two disjoint subsets $\mathbf{X}$ and $\mathbf{Y}$, we define a path from $\mathbf{X}$ to $\mathbf{Y}$ as a path from some $X \in \mathbf{X}$ to some $Y \in \mathbf{Y}$. A path from $\mathbf{X}$ to $\mathbf{Y}$ is called *proper* with respect to $\mathbf{X}$ if only its first node is in $\mathbf{X}$.

A special type of path is the *causal path*. A path $p = \langle X = V_0, \ldots, V_k = Y \rangle$, $k \geq 1$ is called a causal path from X to Y, that is, $X \rightarrow \cdots \rightarrow Y$. For instance, in Figure 1, both $X \rightarrow Y$ and $X \rightarrow Z \rightarrow Y$ are causal paths from X to Y. A path $p = \langle X = V_0, \ldots, V_k = Y \rangle$, $k \geq 1$ is called *possibly causal* if no edge $V_i \leftarrow V_j$, $1 \leq i < j \leq k$, is in $\mathcal{G}$. Otherwise, p is a *non-causal path*.

A causal path p from X to Y together with edge $Y \rightarrow X$ make up a *directed cycle*. A directed graph without directed cycles is a *directed acyclic graph* (DAG). For instance, graphs in Figure 1 and 2 are examples of DAGs. When a DAG is given, we assume that all variables are observed. This is also known as *causal sufficiency* (Pearl, 2009). For example, for DAG $X \rightarrow Y$,

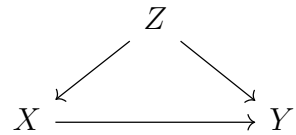

Figure 2: A causal DAG $\mathcal{G}$

we do not permit the possibility that there is some *latent variable* Z such that $X \to Z \to Y$. While approaches to causal inference that allow for latent variables exist, we do not address them in this paper (see Shpitser and Pearl, 2006; Richardson et al., 2017; Richardson and Spirtes, 2002; Zhang, 2008 and Jaber et al., 2019 for some examples of graphical causal research that allows for hidden variables).

Causal paths also give rise to important relationships between nodes. If $X \to Y$ in $\mathcal{G}$, then X is a *parent* of Y in $\mathcal{G}$. Intuitively, the parents of a node Y are its direct and most immediate influences. For example, in Figure 1, X and Z are parents of Y, and X is a parent of Z. We denote the parent set of a node Y by $\mathrm{Pa}(Y, \mathcal{G})$. If $X - Y$ or $X \to Y$ is in $\mathcal{G}$, then X is in the set of possible parents of Y, denoted $\mathrm{PossPa}(Y, \mathcal{G})$. For a set of nodes, $\mathbf{Y}$, we define $\mathrm{Pa}(\mathbf{Y}, \mathcal{G}) = (\cup_{Y \in \mathbf{Y}} \mathrm{Pa}(Y, \mathcal{G})) \setminus \mathbf{Y}$, and $\mathrm{PossPa}(\mathbf{Y}, \mathcal{G}) = \cup_{Y \in \mathbf{Y}} \mathrm{PossPa}(Y, \mathcal{G})$. Y is a *descendant* of X if there is any causal path from X to Y. We denote the set of descendants of X as $\mathrm{De}(X, \mathcal{G})$. Conversely, X is an *ancestor* of Y if there is any causal path from X to Y. We denote the set of ancestors of Y as $\mathrm{An}(Y, \mathcal{G})$.

Finally, if a path p contains $X_i \to X_j \leftarrow X_k$ as a subpath, then X_j is a *collider* on p. For example, in Figure 2, Y is a collider on the path $X \to Y \leftarrow Z$.

3 Do-Calculus

Any foray into causal inference requires a strong understanding of basic conditional probability, and *counterfactual logic*. Counterfactuals are concerned with what is and what could have been. This framework provides us the ability to ask what would be the value of some variable Y had we intervened and set variable X to value x. In line with Pearl (2009), we use $do(\mathbf{X} = \mathbf{x})$, or $do(\mathbf{x})$ to represent an outside *intervention* that sets $\mathbf{X}$ to $\mathbf{x}$. The counterfactual above can then be labeled as $Y_{X=x}$, Y_x, or $Y_{do(x)}$.

Imagine you toss a rock at your neighbors house, and the window breaks. Imagine if you had not thrown the rock - would the window still have broken? It may be reasonable to conclude that, all else remaining the same, if you had not thrown the rock, the window would not be shattered at present. In using counterfactuals, we want to know about what would have happened to the distribution of Y if we had stopped the usual process in charge of X and set $X = x$.

3.1 Probability Distributions and Graphs

In this section, we link probability models with graphs. We begin by establishing notions of independence between sets of nodes in a graph. Throughout the paper, we use notation relevant to continuous random variables, but these concepts hold analogously in the discrete case.

Definition 3.1 (D-separation, Pearl, 1986)**.** *Let X and Y be distinct nodes in a DAG $\mathcal{G}$ and let $\mathbf{Z}$ be a node set that does not contain X or Y in $\mathcal{G}$. A path p from X to Y in $\mathcal{G}$ is said to be* ***d-connecting*** *given $\mathbf{Z}$ if every non-collider on p is not in $\mathbf{Z}$, and every collider on p has a descendant in $\mathbf{Z}$. Otherwise, $\mathbf{Z}$* ***blocks*** *p.*

If $\mathbf{Z}$ blocks all paths between some disjoint node sets $\mathbf{X}$ and $\mathbf{Y}$ in $\mathcal{G}$, then $\mathbf{X}$ is ***d-separated*** *from $\mathbf{Y}$ given $\mathbf{Z}$ in $\mathcal{G}$ and we write $\mathbf{X} \perp_{\mathcal{G}} \mathbf{Y}|\mathbf{Z}$.*

Intuitively, if two variables X and Y are d-separated relative to another set of variables $\mathbf{Z}$ in a DAG; knowing anything about X tells us nothing more about Y if we already know Z. Hence, the concept of a d-separation is analogous to the concept of conditional independencies when certain conditions hold. The following properties make this connection more formal.

Definition 3.2 (Markov condition and faithfulness, Lauritzen, 1996)**.** *Let $\mathcal{G} = (\mathbf{V}, \mathbf{E})$ be a DAG, and f be a probability density over $\mathbf{V}$. Then f is said to be* ***Markov*** *with respect to $\mathcal{G}$ if every d-separation in $\mathcal{G}$ implies a conditional independence in f. Conversely, f is* ***faithful*** *with respect to $\mathcal{G}$ if every d-connection in $\mathcal{G}$ implies a conditional dependence in f.*

Together, the Markov property and and its converse, faithfulness, ensure that the probability distribution and the graph agree completely on what sets are and are not d-separated and conditionally independent. That is, conditional dependencies are identical in the graph and in the distribution.

In addition, distributions that are Markov with respect to a DAG $\mathcal{G}$ are also said to be *consistent* with $\mathcal{G}$ and have important factorization properties.

Definition 3.3 (Observational density, Pearl, 2009)**.** *A density f over $\mathbf{V}$ is* ***consistent*** *with a* DAG $\mathcal{G} = (\mathbf{V}, \mathbf{E})$ *if it factorizes as*

$$f(\mathbf{v}) = \prod_{V_i \in \mathbf{V}} f(v_i | \operatorname{pa}(v_i, \mathcal{G})),$$

where $\operatorname{pa}(v_i, \mathcal{G})$ *is used to denote the values of variables in* $\operatorname{Pa}(V_i, \mathcal{G})$*. A density f that is consistent with $\mathcal{G} = (\mathbf{V}, \mathbf{E})$ is also often called an* ***observational density*** *consistent with $\mathcal{G}$.*

There can be more than one distribution f over $\mathbf{V}$ that is both Markov and faithful with respect to $\mathcal{G} = (\mathbf{V}, \mathbf{E})$. Similarly, a distribution f over $\mathbf{V}$ can be Markov and faithful to more than one DAG $\mathcal{G}$ (see the notion of Markov equivalence class in Section 5).

Definition 3.4 (Bayesian network, Pearl, 2009)**.** *A DAG $\mathcal{G} = (\mathbf{V}, \mathbf{E})$ and a density f over $\mathbf{V}$ that is consistent with $\mathcal{G}$ together form a* ***Bayesian network*** *denoted as $(\mathcal{G}, f)$.*

3.2 The Rules of Intervention

The do-intervention describes the situation in a randomized, controlled experiment. One can think of the do-intervention on X as severing all edges into X in a DAG $\mathcal{G}$. The probability densities we obtain from the do-intervention on $\mathcal{G}$ form another class of probability densities and these together with a DAG form another type of Bayesian network.

Definition 3.5 (Interventional density, c.f. Pearl, 2009; Perković, 2020)**.** *Let $\mathbf{X}$ be a subset of $\mathbf{V}$ and $\mathbf{V}' = \mathbf{V} \setminus \mathbf{X}$ in a* DAG $\mathcal{G} = (\mathbf{V}, \mathbf{E})$*. A density over $\mathbf{V}'$ is denoted by $f(\mathbf{v}'|do(\mathbf{x}))$, or $f_{\mathbf{x}}(\mathbf{v}')$, and called an* ***interventional density*** *consistent with $\mathcal{G}$ if there is an observational density f consistent with $\mathcal{G}$ such that $f(\mathbf{v}'|do(\mathbf{x}))$ factorizes as*

$$f(\mathbf{v}'|do(\mathbf{x})) = \prod_{V_i \in \mathbf{V}'} f(v_i | \operatorname{pa}(v_i, \mathcal{G})), \tag{2}$$

for values $\operatorname{pa}(v_i, \mathcal{G})$ *of* $\operatorname{Pa}(V_i, \mathcal{G})$ *that are in agreement with $\mathbf{x}$. If $\mathbf{X} = \emptyset$, we define $f(\mathbf{v}|do(\emptyset)) = f(\mathbf{v})$.*

Definition 3.6 (Causal Bayesian Network, Pearl, 2009)**.** *Let $\mathcal{G} = (\mathbf{V}, \mathbf{E})$ be a DAG and let $\mathbf{F}$ be a family of all observational and interventional densities that are consistent with $\mathcal{G}$, then $(\mathcal{G}, \mathbf{F})$ form a* ***causal Bayesian network****.*

Equation (2) is known as the truncated factorization formula (Pearl, 2009), manipulated density formula (Spirtes et al., 2000) or the g-formula (Robins, 1986). Generally, if X and Y are two random variables, then $f(y|x) \neq f(y|do(x))$. That is, the observational and interventional densities of Y are not necessarily the same. In order to determine how a marginal interventional density (such as $f(y|do(x))$ relates to an observational density, we need to introduce the rules of the do-calculus.

Theorem 3.7 (Rules of the do-calculus, Pearl, 2009)**.** *Let $\mathbf{X}, \mathbf{Y}, \mathbf{Z}$ and $\mathbf{W}$ be pairwise disjoint (possibly empty) sets of nodes in a DAG $\mathcal{G} = (\mathbf{V}, \mathbf{E})$. Let $\mathcal{G}_{\overline{\mathbf{X}}}$ denote the graph obtained by deleting all edges into $\mathbf{X}$ from $\mathcal{G}$. Similarly, let $\mathcal{G}_{\underline{\mathbf{X}}}$ denote the graph obtained by deleting all edges out of $\mathbf{X}$ in $\mathcal{G}$ and let $\mathcal{G}_{\overline{\mathbf{X}}\underline{\mathbf{Z}}}$ denote the graph obtained by deleting all edges into $\mathbf{X}$ and all edges out of $\mathbf{Z}$ in $\mathcal{G}$. Let $(\mathcal{G}, \mathbf{F})$ be a causal graph, then the following rules hold for densities in $\mathbf{F}$.*

Rule 1. *(Insertion/deletion of observations) If $\mathbf{Y} \perp_{\mathcal{G}_{\overline{\mathbf{X}}}} \mathbf{Z}|\mathbf{X} \cup \mathbf{W}$, then*

$$f(\mathbf{y}|do(\mathbf{x}), \mathbf{w}) = f(\mathbf{y}|do(\mathbf{x}), \mathbf{z}, \mathbf{w}). \tag{3}$$

Rule 2. *(Action/observation exchange) If $\mathbf{Y} \perp_{\mathcal{G}_{\overline{\mathbf{X}}\underline{\mathbf{Z}}}} \mathbf{Z}|\mathbf{X} \cup \mathbf{W}$, then*

$$f(\mathbf{y}|do(\mathbf{x}), do(\mathbf{z}), \mathbf{w}) = f(\mathbf{y}|do(\mathbf{x}), \mathbf{z}, \mathbf{w}). \tag{4}$$

Rule 3. *(Insertion/deletion of actions) If $\mathbf{Y} \perp_{\mathcal{G}_{\overline{\mathbf{X}\mathbf{Z}(\mathbf{W})}}} \mathbf{Z}|\mathbf{X} \cup \mathbf{W}$, then*

$$f(\mathbf{y}|do(\mathbf{x}), \mathbf{w}) = f(\mathbf{y}|do(\mathbf{x}), do(\mathbf{z}), \mathbf{w}), \tag{5}$$

where $\mathbf{Z}(\mathbf{W}) = \mathbf{Z} \setminus \mathrm{An}(\mathbf{W}, \mathcal{G}_{\overline{\mathbf{X}}})$.

Do-calculus is an immensely powerful tool that will form the foundation of the remainder of the paper. The following lemma tells us why the rules of do-calculus are important. Moreover, it uncovers an important connection between interventional and observational distributions.

Lemma 3.8. *Let Y be a node in DAG $\mathcal{G} = (\mathbf{V}, \mathbf{E})$. Furthermore, let $\mathbf{P} = \mathrm{Pa}(Y, \mathcal{G})$ and $\mathbf{V}_y = \mathbf{V} \setminus Y$. Then $f(y|do(\mathbf{v}_y)) = f(y|do(\mathbf{p})) = f(y|\mathbf{p})$.*

Proof. Let $\mathbf{V}_p = \mathbf{V}_y \setminus \mathbf{P}$. Then, $\mathbf{V}_y = \mathbf{V}_p \cup \mathbf{P}$ and so $f(y|do(\mathbf{v}_y)) = f(y|do(\mathbf{v}_p), do(\mathbf{p}))$. Note that $\mathbf{V}_p \setminus \mathrm{An}(\emptyset, \mathcal{G}_{\overline{\mathbf{P}}}) = \mathbf{V}_p$ since $\emptyset$ has no ancestors in any graph. Then, the graph $\mathcal{G}_{\overline{\mathbf{V}_y}}$ is the mutilated graph $\mathcal{G}$ with all arrows into $\mathbf{V}_y$ removed. Hence, the only arrows that remain in $\mathcal{G}_{\overline{\mathbf{V}_y}}$ are those that point from $\mathbf{P}$ to Y. Therefore, there are no paths from Y to $\mathbf{V_p}$ in $\mathcal{G}_{\overline{\mathbf{V}_y}}$ and so $Y \perp_{\mathcal{G}_{\overline{\mathbf{V}_y}}} \mathbf{V}_p|\mathbf{P}$. It follows from Equation (5), the third rule of do-calculus, that $f(y|do(\mathbf{v}_p), do(\mathbf{p})) = f(y|do(\mathbf{p}))$.

Next, consider the graph $\mathcal{G}_{\underline{\mathbf{P}}}$, the mutilated graph $\mathcal{G}$ with all edges emerging from $\mathbf{P}$ removed. This implies that there are no edges pointing into Y in $\mathcal{G}_{\underline{\mathbf{P}}}$. As $\mathcal{G}$ is acyclic, there are no paths that emerge from Y that return to $\mathbf{P}$. Therefore, it follows that $Y \perp_{\mathcal{G}_{\underline{\mathbf{P}}}} \mathbf{P}$, and by Equation 4, the second rule of do-calculus, $f(y|do(\mathbf{p})) = f(y|\mathbf{p})$. □

Finally, the rules of do-calculus provide a complete characterization of the causal effects in a DAG, as seen below, further emphasizing their importance and utility.

Theorem 3.9 (Completeness of the do-calculus, Huang and Valtorta, 2006; Shpitser and Pearl, 2006)**.** *The three do-calculus rules, together with standard probability manipulations, are complete for determining identifiability of all causal effects in a DAG.*

4 Causal Effects in DAGs

Suppose that we want to know the effect of wearing a mask on the transmission rate of a disease. We could imagine that the mask may decrease infections since it physically blocks the airborne spread of germs. However, it could be the case that wearing a mask in public influences other people to move farther away from the mask wearer. These larger distances between people may also cause the disease rate to decrease. This relationship is represented by Figure (3).

This means that the *total causal effect* of mask wearing on disease rate is two fold: one direct through barring the spread of germs, and the other indirect through physical distancing. We may be interested in knowing how much of the reduction in disease rate is attributable to the physical barrier of the mask. This is appropriately called the *direct causal effect* of mask wearing on the disease rate. The other influence of mask wearing, through increased physical distancing, is called the *indirect causal effect* of the mask

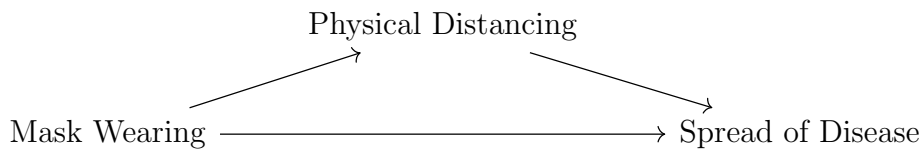

Figure 3: DAG $\mathcal{G}$ representing relationship between mask wearing, physical distance, and disease

on disease rate. Often times, we are mainly interested in the direct causal effect because it is more transportable. For instance, the influence of mask wearing on spread of disease through physical distancing in a grocery store would not translate to a hospital setting where masks are more common. Hence, we want to isolate the direct effect of the mask wearing on the spread of disease.

4.1 Total Causal Effects

We first introduce the definition of the average or total causal effect, or ACE.

Definition 4.1 (Average causal effect, Total causal effect, Pearl, 2009)**.** *Let X and Y be two distinct nodes in a DAG $\mathcal{G} = (\mathbf{V}, \mathbf{E})$. The **average** or **total causal effect** of X on Y when changing from $X = x$ to $X = x+1$ is defined as*

$$ACE(x, x+1, Y) := E[Y|do(x+1)] - E[Y|do(x)].$$

Generally speaking, it is not necessary to restrict ourselves to $X = x$ and $X = x + 1$. Much development in causal modeling has been through an epidemiological lens where $X = 0$ represents a control group, $X = 1$ represents a treatment group, and therefore $ACE(0, 1, Y)$ is used. However, it is equally as valid to evaluate the ACE by changing X from some chosen baseline, $X = x$, to some other value of $X = x^*$.

Under assumptions on the form of the SCM that underlies the DAG, we can make stronger statements about the ACE of X on Y.

Definition 4.2 (Linear SCM, Pearl, 2009)**.** *Random vector $\mathbf{V} = (V_1, \ldots, V_p)^T$, $p \geq 1$ is generated by a **linear SCM** compatible with DAG $\mathcal{G} = (\mathbf{V}, \mathbf{E})$ if*

$$V_i \leftarrow \sum_{V_j \in \mathrm{Pa}(V_i, \mathcal{G})} \alpha_{v_j v_i} V_j + u_{v_i}. \tag{6}$$

In Equation (6)*,* $i, j \in \{1, \ldots, p\}, i \neq j$*,* $\alpha_{v_j v_i} \in \mathbb{R}$*, is the* ***edge coefficient*** *corresponding to* $V_j \to V_i$*, and* $u_{v_1}, \ldots, u_{v_p}$ *are jointly independent random noise variables with mean 0 and finite variance.*

This definition leads us to the following results about the ACE in the linear setting.

Proposition 4.3 (Pearl, 2009)**.** *Let* X *and* Y *be distinct nodes in a DAG* $\mathcal{G} = (\mathbf{V}, \mathbf{E})$ *and suppose that* $\mathbf{V}$ *is generated by a linear SCM compatible with* $\mathcal{G}$*. Then*

$$E[Y|do(x)] = \tau_{xy} x,$$

where $\tau_{xy} \in \mathbb{R}$ *is the same for all* $X = x$*, and* $\tau_{xy} = ACE(x, x+1, Y)$*. Furthermore, for each node* $V_i \in \mathbf{V}$*,*

$$V_i \leftarrow \sum_{V_j \in \mathrm{An}(V_i, \mathcal{G})} \tau_{v_j v_i} u_{v_j} + u_{v_i}.$$

This result has been known for some time, albeit in a slightly different formulation, which we give below.

Proposition 4.4 (Wright, 1934)**.** *Let* X *and* Y *be distinct nodes in a DAG* $\mathcal{G} = (\mathbf{V}, \mathbf{E})$ *and suppose that* $\mathbf{V}$ *is generated by a linear SCM compatible with* $\mathcal{G}$*. The* ***ACE along a causal path*** p *from* X *to* Y *in* $\mathcal{G}$ *is the product of the edge coefficients along* p*. Then, the ACE of* X *on* Y *in* $\mathcal{G}$ *is equal to the sum of the total effects along all causal paths from* X *to* Y *in* $\mathcal{G}$*.*

Linear SCMs have many useful properties that make them convenient for working with DAGs, such as the following lemma.

Lemma 4.5. *Let* X *and* Y *be two distinct nodes in a DAG* $\mathcal{G} = (\mathbf{V}, \mathbf{E})$ *and suppose that* $X \to Y$ *is in* $\mathcal{G}$*. Furthermore, suppose that* $\mathbf{V}$ *is generated by a linear SCM compatible with* $\mathcal{G}$*. If* $\mathbf{P}' = \mathrm{Pa}(Y, \mathcal{G}) \setminus \{X\}$*, then*

$$E[Y|do(x)] = E[Y|do(x), do(E[\mathbf{P}'|do(x)])].$$

Proof. First, note that if $\mathbf{P}' = \emptyset$, the above holds immediately. Otherwise, let $\mathbf{P}' = (P_1, \ldots, P_k)^T$, $k \geq 1$. Let δ be the edge coefficient $X \to Y$, and let $\alpha_{p_i y}$ be the edge coefficient between $P_i \to Y$ (Definition 4.2). Then, by Proposition 4.3,

$$E[\mathbf{P}'|do(x)] = (E[P_1|do(x)], \ldots, E[P_k|do(x)])^T = (\tau_{xp_1} x, \ldots, \tau_{xp_k} x)^T.$$

Therefore,

$$E[Y|do(x), do(E[\mathbf{P}'|do(x)])] = \delta x + \sum_{i=1}^{k} \alpha_{p_i y} E[P_i|do(x)] + E[u_y|do(x)]$$
$$= (\delta + \sum_{i=1}^{k} \alpha_{p_i y} \tau_{x p_i}) x.$$

In the last equality, we use Proposition 4.3 again, and we use that u_y is independent of X, and additionally that $E[u_y] = 0$ by assumption. Note that u_y is independent of X, because $X = \sum_{V_i \in \mathrm{An}(X,\mathcal{G})} \tau_{v_i x} u_{v_i}$, $u_y \notin \mathrm{An}(X, \mathcal{G})$, and because random errors are mutually independent.

Now, by Theorem 4.4, $\tau_{xy} = (\delta + \sum_{i=1}^{k} \alpha_{p_i y} \tau_{x p_i})$, and hence

$$E[Y|do(x)] = (\delta + \sum_{i=1}^{k} \alpha_{p_i y} \tau_{x p_i}) x = E[Y|do(x), do(E[\mathbf{P}'|do(x)])].$$

□

4.2 Direct Causal Effects

In a DAG, the presence of a direct effect of variable X on variable Y is represented with a direct edge $X \to Y$. Naively, we could assume that we could estimate this effect by fixing X to some value (with a do-intervention) and measuring Y, then fixing X to some other value and measuring the outcome Y again. Unfortunately, things are not so simple; there are other variables in the graph to consider, and only changing X will give us the ACE. At what values should the other variables be held? Should they be held constant? In this section, we discuss two definitions of the direct effect: the controlled, and the average natural.

For ease of notation we let $\mathbf{V}' = \mathbf{V} \setminus \{X, Y\}$, $\mathbf{P} = \mathrm{Pa}(Y, \mathcal{G})$, $\mathbf{P}' = \mathrm{Pa}(Y, \mathcal{G}) \setminus X$, and $\mathbf{p}'_x = E[\mathbf{P}'|do(x)]$ for the rest of the paper.

Definition 4.6 (Controlled direct effect)**.** *Let X and Y be two distinct nodes in a DAG $\mathcal{G} = (\mathbf{V}, \mathbf{E})$. The* ***controlled direct effect*** *of X on Y when changing from $X = x$ to $X = x + 1$ and keeping $\mathbf{V}' = \mathbf{v}'$ is defined as*

$$CDE(x, \mathbf{v}', Y) \coloneqq E[Y|do(x+1), do(\mathbf{v}')] - E[Y|do(x), do(\mathbf{v}')].$$

To compute the controlled direct effect, or CDE, we set all other variables in the graph to some fixed values of our choosing. Unfortunately, performing a randomized controlled experiment with all non-exposure and non-response variables fixed to a particular value is a difficult, if not impossible undertaking. Fortunately, we have a way of computing this effect from observational data.

Corollary 4.7. *Let X and Y be two distinct nodes in a DAG $\mathcal{G} = (\mathbf{V}, \mathbf{E})$. Then,*

$$CDE(x, \mathbf{v}', Y) = CDE(x, \mathbf{p}', Y),$$

for values $\mathbf{P}' = \mathbf{p}'$ in agreement with $\mathbf{V}' = \mathbf{v}'$. Moreover,

$$CDE(x, \mathbf{p}', Y) = E[y|x+1, \mathbf{p}'] - E[y|x, \mathbf{p}'].$$

This result follows immediately from Lemma 3.8. Hence, we only need to control for a subset of all variables in the graph in order to estimate the CDE. Moreover, this result tells us that when we know the variables that represent the parents of Y in the DAG, observational data is sufficient to identify the CDE.

What if there is no edge $X \to Y$ in the DAG? Intuitively, we may assume that this implies there is no direct effect of X on Y. While this is intuitively true and generally accepted in the causal literature, to the best of our knowledge, there is no published formal proof; We provide one below.

Proposition 4.8. *Let X and Y be two distinct nodes in a DAG $\mathcal{G} = (\mathbf{V}, \mathbf{E})$. If $X \to Y$ is not in $\mathcal{G}$, then the CDE of X on Y is zero.*

Proof. Since there is no edge $X \to Y$ in $\mathcal{G}$, $X \notin \mathbf{P}$, and $\mathbf{P} = \mathbf{P}'$. Any causal path from X to Y in $\mathcal{G}$ must go through an element of $\mathbf{P}'$. Denote $\mathbf{W} = \mathbf{P} \cup \{X\}$. Note that $X \setminus \mathrm{An}(\emptyset, \mathcal{G}_{\overline{\mathbf{W}}}) = X$ since $\emptyset$ has no ancestors in any graph. The graph $\mathcal{G}_{\overline{\mathbf{W}}}$ is the mutilated graph $\mathcal{G}$ with all arrows pointing into X and $\mathbf{P}$ removed. Moreover, no paths between X and Y are d-connected by conditioning on $\mathbf{P}$ since no colliders between X and Y are ancestors of $\mathbf{P}$ in the graph $\mathcal{G}_{\overline{\mathbf{W}}}$. This is because there are no arrows pointing into $\mathbf{P}$ in the graph $\mathcal{G}_{\overline{\mathbf{W}}}$, and so $\mathrm{An}(\mathbf{P}, \mathcal{G}_{\overline{\mathbf{W}}}) = \emptyset$. Hence, there are no paths between X and Y in $\mathcal{G}_{\overline{\mathbf{W}}}$, and so $Y \perp_{\mathcal{G}_{\overline{\mathbf{W}}}} X|\mathbf{P}$. It then follows from the third rule of do-calculus that $f(y|do(x), do(\mathbf{p}')) = f(y|do(\mathbf{p}'))$ and therefore $CDE(x, Y) = 0$. $\square$

Using the CDE as an estimate of the "direct" causal effect is still unrealistic in many scenarios. For instance, it may be impossible to force all subjects to have the same values for certain variables. Consider parents of Y that are descendants of X; they are also affected by setting X to a certain value in a randomized controlled trial. Furthermore, not all possible values of these mediating variables are necessarily realistic under certain interventions on X. We propose another estimate of the direct effect that allows for a degree of natural variation among subjects.

Definition 4.9 (Average natural direct effect)**.** *Let X and Y be distinct nodes in a DAG $\mathcal{G} = (\mathbf{V}, \mathbf{E})$. If $X \to Y$ is in $\mathcal{G}$, the* ***average natural direct effect*** *of X on Y when changing from $X = x$ to $X = x + 1$ is*

$$ANDE(x, Y) := E[Y|do(x+1), do(\mathbf{p}'_x)] - E[Y|do(x)],$$

where $\mathbf{p}'_x = E[\mathbf{P}'|do(x)]$. If $X \to Y$ is not in $\mathcal{G}$, we set $ANDE(x, Y) = 0$.

Upon first inspection, the difference between the CDE and the ANDE may not be obvious. To illustrate, we provide the following example which is modified from (Petersen et al., 2006). Suppose that we are interested in learning about the impact of rescue medication on lung function. Consider the DAG in Figure 1. Let Y represent our outcome of interest, lung function, Z the amount of a rescue medication, and X the level of pollution.

In this case, the CDE is equal to the expected lung function if the entire population were exposed to an incremental increase in air pollution and every member of the population was to use a single fixed level of rescue medication both prior to and after this incremental intervention.

On the other hand, the ANDE is equal to the change in the expected lung function if the entire population were exposed to an incremental increase in air pollution, but every individual in the population behaved as they would have behaved *on average* if the air pollution had not increased.

One can imagine that in this scenario, the CDE may be an unrealistic measure of the direct effect, given that some individuals in the population have different baseline rates of rescue medication use. For instance, an individual with asthma may use more rescue medication than an individual without asthma, and forcing an individual to change their amount of rescue medication use could be ethically ambiguous.

Why do we require that the average natural direct effect, or $ANDE$, is zero when there is no directed edge $X \to Y$? While we proved in Lemma 4.5

that $E[Y|do(x)] = E[Y|do(x), do(\mathbf{p}'_x)]$ when the SCMs are linear, this is not the case in most scenarios, even when there is no edge $X \to Y$ in $\mathcal{G}$.

Example 4.10. *Consider the following SCM compatible with the DAG in Figure 1.*

$$\begin{aligned} X &\leftarrow u_x \\ Z &\leftarrow \sqrt{X} + u_z \\ Y &\leftarrow X^2 + Z^2 + u_y, \end{aligned}$$

where we choose $Var[u_x] = Var[u_z] = Var[u_y] = 1$. *We can now compute* $E[Y|do(x)]$.

$$\begin{aligned} E[Y|do(x)] &= E[X^2 + Z^2 + u_y|do(x)] \\ &= E[X^2 + (\sqrt{X} + u_z)^2 + u_y|do(x)] \\ &= x^2 + x + 2\sqrt{x}E[u_z|do(x)] + E[u_z^2|do(x)] + E[u_y|do(x)] \\ &= x^2 + x + 1, \end{aligned}$$

Above, we used the linearity of expectation and the fact that u_z, u_y *are independent of* $X = u_x$ *and have a mean of zero. Next, we compute* $z_x = E[Z|do(x)]$.

$$z_x = E[Z|do(x)] = E[\sqrt{X} + u_z|do(x)] = \sqrt{x}.$$

Above, we again used the linearity of expectation, the joint independence of errors and the fact that they are mean zero. Finally, we compute $E[Y|do(x), do(Z = z_x)]$.

$$\begin{aligned} E[Y|do(x),\ do(Z = z_x)] &= E[X^2 + Z^2 + u_y|do(x),\ do(Z = \sqrt{x})] \\ &= x^2 + (\sqrt{x})^2 + 0 \\ &= x^2 + x. \end{aligned}$$

Therefore, $E[Y|do(x)] = x^2 + x + 1 \neq x^2 + x = E[Y|do(x),\ do(Z = z_x)]$.

4.2.1 Identifiability of Direct Causal Effects

When can we identify the CDE and ANDE from observational data? From Definition 4.6 and Corollary 4.7, we see that the CDE of X on Y is identifiable

whenever $f(y|do(x), do(\mathbf{p}'))$ is identifiable. That is, whenever $f(y|do(x), do(\mathbf{p}'))$ is uniquely computable from the observational density f. Likewise, from Definition 4.9, we can deduce that the ANDE of X on Y is identifiable whenever $f(y|do(x))$, $f(\mathbf{p}'|do(x))$, and $f(y|do(x+1), do(\mathbf{P}' = E[\mathbf{P}'|do(x)]))$ are uniquely computable from the observational density. Under certain additional assumptions, say linearity, we may be able to identify ANDE with fewer assumptions.

Example 4.11. *Consider the DAG $\mathcal{G} = (\mathbf{V}, \mathbf{E})$ in Figure 1, and suppose that $\mathbf{V}$ is generated by the following linear SCM.*

$$\begin{aligned} X &\leftarrow u_x \\ Z &\leftarrow \beta X + u_z \\ Y &\leftarrow \gamma X + \delta Z + u_y, \end{aligned}$$

where u_x, u_z, and u_y are mutually independent errors with mean zero and constant variance. Then, the CDE of X on Y is given by

$$\begin{aligned} CDE(x, z, Y) &= E[Y|do(x+1), do(z)] - E[Y|do(x), do(z)] \\ &= \gamma(x+1) + \delta z - (\gamma x + \delta z) \\ &= \gamma. \end{aligned}$$

Let $z_x = E[Z|do(x)]$. Then the ANDE of X on Y is given by

$$\begin{aligned} ANDE(x, Y) &= E[Y|do(x+1), do(z_x)] - E[Y|do(x)] \\ &= E[Y|do(x+1), do(z_x)] - E[\gamma x + \delta Z + u_y|do(x)] \\ &= \gamma(x+1) + \delta z_x - (\gamma x + \delta E[Z|do(x)]) \\ &= \gamma(x+1) + \delta z_x - \gamma_x - \delta z_x \\ &= \gamma. \end{aligned}$$

Hence, in the case of a linear SCMs, the CDE and the ANDE are equal. In particular, both are fully specified by the edge coefficient of $X \to Y$. In this scenario, knowing the path specific coefficient is sufficient to estimate the causal effect; we do not need to know any conditional distribution of Y. This leads us to the following proposition.

Proposition 4.12. *Let X and Y be distinct nodes in DAG $\mathcal{G} = (\mathbf{V}, \mathbf{E})$ and suppose that $\mathbf{V}$ is generated by a linear SCM. Then, $CDE(x, \mathbf{p}', Y) = ANDE(x, Y) = \delta$, for all values x, where $\delta = \alpha_{xy}$ is the edge coefficient corresponding to $X \to Y$ in the linear SCM.*

Proof. Let $\mathbf{P}' = (V_1, \ldots, V_k)^T$. We can write $Y = \delta X + \sum_{i=1}^{k} \alpha_{p_i y} P_i + u_y$. It follows that the CDE is given by

$$\begin{aligned} CDE(x, \mathbf{p}', Y) &= E[Y|do(x+1), do(\mathbf{p}')] - E[Y|do(x), do(\mathbf{p}')] \\ &= E[Y|x+1, \mathbf{p}'] - E[Y|x, \mathbf{p}'] \\ &= \delta(x+1) + \sum_{i=1}^{k} \alpha_{p_i y} p_i - (\delta x + \sum_{i=1}^{k} \alpha_{p_i y} p_i) \\ &= \delta. \end{aligned}$$

Above we used Lemma 3.8, the linearity of expectation, joint independence of errors, and the fact that the errors have mean zero. Similarly, if $\mathbf{p}'_x = (p_1^{'}, \ldots, p_k^{'})^T$ then the ANDE is given by

$$\begin{aligned} ANDE(x, Y) &= E[Y|do(x+1), do(\mathbf{p}'_x)] - E[Y|do(x)] \\ &= E[Y|do(x+1), do(\mathbf{p}'_x)] - E[Y|do(x), do(\mathbf{p}'_x)] \\ &= E[Y|x+1, \mathbf{p}'_x] - E[Y|x, \mathbf{p}'_x] \\ &= \delta(x+1) + \sum_{i=1}^{k} \alpha_{p_i y} p_i^{'} - (\delta x + \sum_{i=1}^{k} \alpha_{p_i y} p_i^{'}) \\ &= \delta. \end{aligned}$$

Above we used Lemma 3.8, the linearity of expectation, joint independence of errors, and the fact that the errors have mean zero. □

Inspection of the proof of Proposition 4.12 tells us that we, in fact, only require the generating mechanism of Y to have a linear form in order to obtain a constant CDE.

Corollary 4.13. *Let X and Y be distinct nodes in DAG $\mathcal{G} = (\mathbf{V}, \mathbf{E})$ and suppose that the generating equation of Y is linear. Then, $CDE(x, \mathbf{p}', Y) = \delta$, where δ is the edge coefficient of $X \to Y$ in the linear generating equation of Y.*

4.2.2 Different Definitions of the Natural Direct Effect

Note that our definition of $ANDE(x, Y)$ is different from the definition of the *natural direct effect*, or *pure direct effect* elsewhere in the causal literature. In 1992, Robins and Greenland proposed a measure of the direct causal effect

called the pure direct effect (Robins and Greenland, 1992). Pearl adopted their definition, but renamed the effect as the natural direct effect in (Pearl, 2001).

Definition 4.14 (Robins and Greenland, 1992; Pearl, 2001)**.** *Let X and Y be two distinct nodes in a DAG $\mathcal{G} = (\mathbf{V}, \mathbf{E})$. The* ***pure****, or* ***natural direct effect*** *of X on Y, when switching from $X = x$ to $X = x + 1$ is defined as*

$$PDE(x, Y) \coloneqq E[Y_{x+1,\mathbf{P}'_x}] - E[Y_x].$$

In the above definition, we may replace $E[Y_x]$ with $E[Y|do(x)]$. However, in general $E[Y_{x+1,\mathbf{P}'_x}] \neq E[Y|do(x+1), do(\mathbf{P}' = E[\mathbf{P}'|do(x)])]$. Hence, the pure direct effect is not equal to the $ANDE$. The counterfactual, $Y_{x+1,\mathbf{P}'_x}$, represents the value of each individual response Y when we intervene and set $X = x + 1$ and we set $\mathbf{P}'$ to the exact individual value $\mathbf{P}'$ would have taken if we intervened to set $X = x$. Therefore, for different individuals in the population, $\mathbf{P}'$ is set to different values in the pure direct effect, unlike in the ANDE definition.

The counterfactual, $Y_{x+1,\mathbf{P}'_x}$, is a so-called "cross-world" counterfactual, because it combines information at the individual level from two different "worlds": the world where the do-intervention sets $X = x$ and the world where the do-intervention sets $X = x + 1$. Identifying the pure direct effect therefore requires making assumptions about cross-world independencies which cannot be verified in practice, and which cannot be graphically represented using a DAG. Hence, we do not rely on this definition of the direct effect. Our definition of the ANDE strives to achieve a similar purpose as the pure direct effect while being fully characterized in the "single-world" scenarios.

However, the pure direct effect definition does allow for an elegant decomposition of causal effects (VanderWeele, 2013). Furthermore, the pure direct effect is able to represent the natural variability among individual responses to an intervention on an exposure which is why there has been much research on the conditions for its identifiability (Robins et al., 2009; Robins and Richardson, 2010), or its bounds (Kaufman et al., 2009; Imai et al., 2013; Tchetgen and Phiri, 2014).

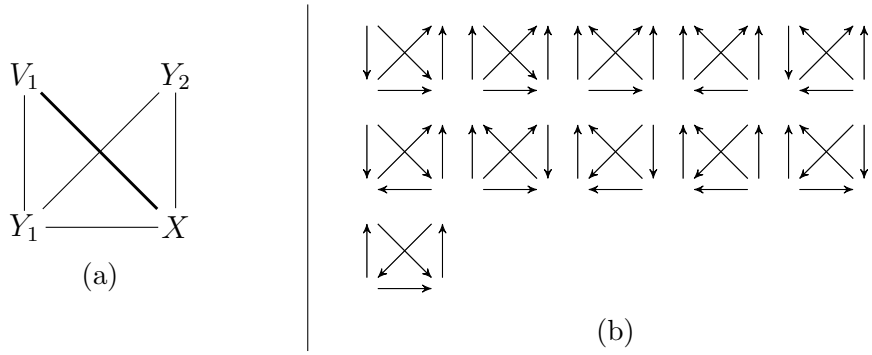

Figure 4: (a) CPDAG $\mathcal{C}$, (b) all DAGs in $[\mathcal{C}]$, reprinted from Perković (2020).

5 Direct Causal Effects in CPDAGs and MPDAGs

It is often not possible to learn the underlying causal DAG from observational data alone. When all variables in the causal system are observed, one can at most learn a *completed partially directed acyclic graph*, (CPDAG, Meek, 1995; Andersson et al., 1997; Spirtes et al., 2000; Chickering, 2002). A CPDAG $\mathcal{C}$ uniquely represents an equivalence class of DAGs called the Markov equivalence class, which we label as $[\mathcal{C}]$. The defining property of the Markov equivalence class of DAGs is that all DAGs in the Markov equivalence class share the same d-separations and the same skeleton or underlying adjacencies (Meek, 1995; Lauritzen, 1996; Pearl, 2009).

A CPDAG can contain both directed and undirected edges. An example of a CPDAG is given in Figure 4(a) which is reprinted from Perković (2020). All DAGs in the Markov equivalence class represented by a CPDAG $\mathcal{C}$ consist of the same node and adjacencies as $\mathcal{C}$. Additionally, if an edge $X \to Y$ is in $\mathcal{C}$, then $X \to Y$ is in every DAG in $[\mathcal{C}]$. Conversely, for every undirected edge $X - Y$ in $\mathcal{C}$ there is at least one DAG in $[\mathcal{C}]$ with $X \to Y$ and at least one DAG with edge $X \leftarrow Y$. An example of a Markov equivalence class of DAGs is given in Figure 4(b). This Markov equivalence class corresponds to the CPDAG $\mathcal{C}$ in Figure 4(a).

If in addition to observational data one has background knowledge of some pairwise causal relationships (say from an expert on the domain topic), one can obtain a *maximally oriented partially directed acyclic graph* (MPDAG) which uniquely represents a refinement of the Markov equivalence class of

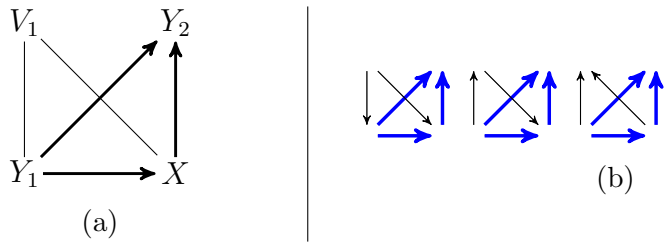

Figure 5: (a) MPDAG $\mathcal{G}$, (b) all DAGs in $[\mathcal{G}]$, reprinted from (Perković, 2020).

DAGs (Meek, 1995). For instance, the graph in Figure 5(a) is an MPDAG $\mathcal{G}$ (this figure is reprinted from Perković, 2020). The MPDAG that results from adding the background knowledge of $Y_1 \to X$, $X \to Y_2$ and $Y_1 \to Y_2$ to CPDAG $\mathcal{G}$ in Figure 4.

The equivalence class of DAGs represented by $\mathcal{G}$, which we denote with $[\mathcal{G}]$ is given in Figure 5(b). The class of DAGs in 5(b) are exactly those DAGs in the Markov equivalence class in Figure 4(b) that include the edge orientations $Y_1 \to X$, $X \to Y_2$ and $Y_1 \to Y_2$.

When is a partially directed acyclic graph a CPDAG or an MPDAG? (Meek, 1995) gives a set of four graphical orientation rules that any partially directed acyclic graph must satisfy to be a CPDAG or an MPDAG. We print these rules in Figure 6. While we do not discuss these rules for construction of CPDAGs and MPDAGs in detail in this paper, it is worth noting that these rules exploit acyclicity of the underlying DAG and use known conditional independences to orient as many edges as possible.

Since all DAGs represented by a CPDAG or MPDAG $\mathcal{G}$ share the same d-separations, by the Markov condition (Definition 3.2) any observational density f compatible with a DAG in $[\mathcal{G}]$ is also compatible with all other DAGs in $[\mathcal{G}]$ (Lauritzen, 1996). Hence, we say that a density f is the observational density compatible with CPDAG or MPDAG $\mathcal{G} = (\mathbf{V}, \mathbf{E})$ if f is compatible with a DAG in $[\mathcal{G}]$.

5.1 Identifiability of Direct Causal Effects

If a CPDAG or an MPDAG is something we can learn from data and expert knowledge, the question then becomes, when can we identify causal effects given one of these types of graphs?

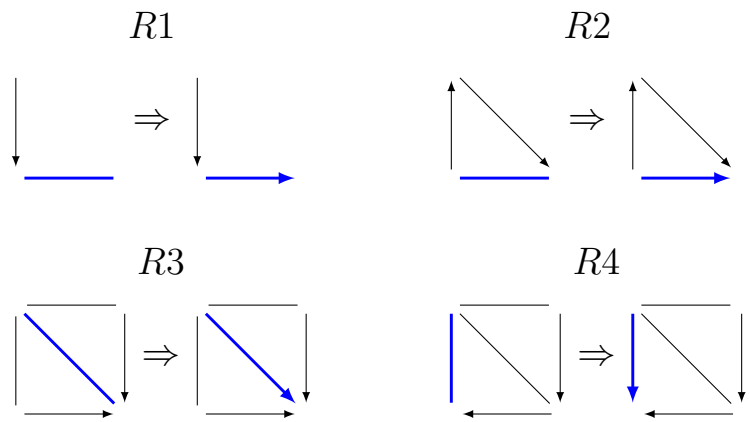

Figure 6: The orientation rules of Meek (1995). If the graph on the left-hand side of a rule is an induced subgraph of a PDAG $\mathcal{G}$, then *orient* the blue undirected edge ($-$) as shown on the right-hand side of the rule. Hence, the graphs on the left-hand side of each rule are not allowed to be induced subgraphs of an MPDAG or a CPDAG. This particular figure is reprinted from Henckel et al. (2019).

We consider a causal effect (total, controlled, average natural) to be identifiable given a CPDAG or MPDAG $\mathcal{G}$, if this causal effect is equal for all DAGs in $[\mathcal{G}]$. In other words, a causal effect is identifiable in CPDAG or MPDAG $\mathcal{G}$ if it can be uniquely computed from the observational distribution compatible with $\mathcal{G}$.

Lemma 5.1. *Let X and Y be two distinct nodes in a CPDAG or MPDAG $\mathcal{G}$. If there is no edge between X and Y in $\mathcal{G}$, or if $X \leftarrow Y$ is in $\mathcal{G}$, then the CDE of X on Y in $\mathcal{G}$ is equal to zero.*

Proof. Suppose that there is no edge between X and Y in $\mathcal{G}$. Then, there is no edge between X and Y in any DAG in $[\mathcal{G}]$. Hence, by Proposition 4.8, the CDE of X on Y is zero in every DAG in $[\mathcal{G}]$. Therefore, the set of CDEs of X on Y is zero for all DAGs in $[\mathcal{G}]$, and so the CDE of X on Y is also zero in $\mathcal{G}$. □

Lemma 5.2. *Let X and Y be two distinct nodes in a CPDAG or MPDAG $\mathcal{G}$. If $X - Y$ is in $\mathcal{G}$, then the CDE of X on Y is not identifiable given $\mathcal{G}$.*

Proof. Suppose that $X - Y$ is in $\mathcal{G}$. Then, $\mathcal{G}$ represents at least one DAG with $X \to Y$. In this DAG, the edge $X \to Y$ implies that the effect of X

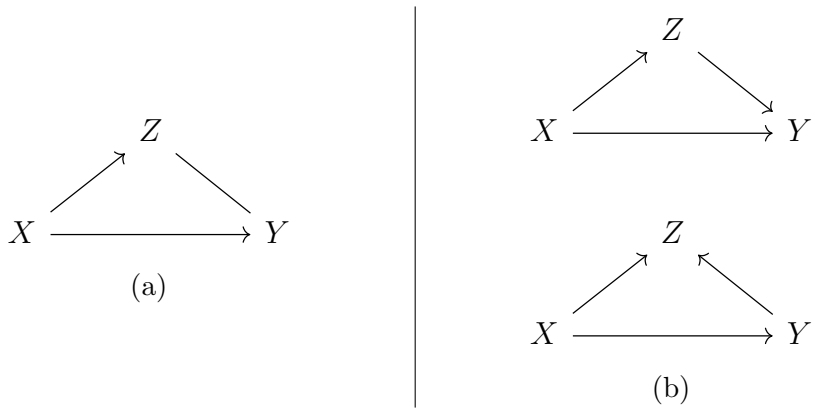

Figure 7: (a) MPDAG $\mathcal{G}$, (b) The set of DAGs in $[\mathcal{G}]$, labeled $\mathcal{G}_1$ and $\mathcal{G}_2$ from top to bottom.

on Y is non-zero (when assuming faithfulness). That is, the presence of a directed edge between X and Y implies $E[Y|do(x)] \neq E[Y]$ in general.

Similarly, edge $X - Y$ in $\mathcal{G}$ implies that there exists at least one DAG in $[\mathcal{G}]$ that contains $Y \to X$ instead of $X \to Y$. Edge $Y \to X$ implies that the CDE of X on Y is zero by Proposition 4.8. Now, because there are two distinct CDEs of X on Y that can be computed given $[\mathcal{G}]$, the CDE of X on Y is not identifiable in $\mathcal{G}$. □

Above, we discuss cases when there is no edge between X and Y in CPDAG or MPDAG $\mathcal{G}$ and when $X - Y$, or $X \leftarrow Y$ is in $\mathcal{G}$. What about the case when $X \to Y$ is in $\mathcal{G}$? Is the CDE of X on Y identifiable in $\mathcal{G}$ in this case? Recently, Perković (2020) proved the following result.

Theorem 5.3 (see Proposition 3.2 and Theorem 3.3 of Perković, 2020)**.** *Let $\mathbf{X}$ and $\mathbf{Y}$ be disjoint node sets in a CPDAG or MPDAG $\mathcal{G} = (\mathbf{V}, \mathbf{E})$. Then the interventional marginal density $f(\mathbf{y}|do(\mathbf{x}))$ is identifiable in $\mathcal{G}$, that is, $f(\mathbf{y}|do(\mathbf{x}))$ is uniquely computable from an observational distribution f compatible with $\mathcal{G}$ if and only if every proper possibly causal path from $\mathbf{X}$ to $\mathbf{Y}$ starts with a directed edge in $\mathcal{G}$.*

Next, we use Theorem 5.3 to prove the following result on the identifiability of the CDE in a CPDAG or MPDAG.

Theorem 5.4. *Let X and Y be two distinct nodes in a CPDAG or MPDAG $\mathcal{G}$. Suppose that the edge $X \to Y$ is in $\mathcal{G}$. The CDE of X on Y is identifiable*

in $\mathcal{G}$ if and only if there is no undirected edge connected to Y in $\mathcal{G}$. That is, the CDE of X on Y is identifiable in $\mathcal{G}$ if and only if $\text{PossPa}(Y, \mathcal{G}) = \text{Pa}(Y, \mathcal{G})$.

Proof. If $\text{PossPa}(Y, \mathcal{G}) = \text{Pa}(Y, \mathcal{G})$, then the parent set of Y is the same for every DAG in $[\mathcal{G}]$. Then by the definition of the CDE, the CDE of X on Y must be equal for every DAG in $\mathcal{G}$.

To show the converse, we prove the contrapositive: if $\text{PossPa}(Y, \mathcal{G}) \neq \text{Pa}(Y, \mathcal{G})$, then the CDE of X on Y is not identifiable, by constructing an example. If $\text{PossPa}(Y, \mathcal{G}) \neq \text{Pa}(Y, \mathcal{G})$, then there is an undirected edge connected to Y in $\mathcal{G}$. Consider the MPDAG $\mathcal{G}$ in Figure 7(a) and the DAGs $\mathcal{G}_1$ and $\mathcal{G}_2$ in $[\mathcal{G}]$ given in Figure 7(b). Suppose that the true SCM is given by,

$$\begin{aligned} X &\leftarrow u_x \\ Z &\leftarrow X + u_z \\ Y &\leftarrow X + Z + u_y \end{aligned} \tag{7}$$

where u_x, u_z, u_y are jointly independent $\mathcal{N}(0,\ 1)$ random variables. Note that this implies that the causal DAG is $\mathcal{G}_1$. We can compute the CDE of X on Y in $\mathcal{G}_1$, labeled CDE_1:

$$\begin{aligned} CDE_1 &= E[Y|do(x+1), do(z)] - E[Y|do(x), do(z)] \\ &= (x+1) + z - (x+z) \\ &= 1. \end{aligned}$$

Let us now compute CDE_2 which denotes the CDE of X on Y in $\mathcal{G}_2$:

$$\begin{aligned} CDE_2 &= E[Y|do(x+1)] - E[Y|do(x)] \\ &= E[X + Z + u_y|do(x+1)] - E[X + Z + u_y|do(x)] \\ &= E[X + X + u_z + u_y|do(x+1)] - E[X + X + u_z + u_y|do(z)] \\ &= 2(x+1) - 2x \\ &= 2. \end{aligned}$$

As $1 \neq 2$, we have found a example that shows that if $\text{PossPa}(Y, \mathcal{G}) \neq \text{Pa}(Y, \mathcal{G})$, then the CDE of X on Y is not identifiable in $\mathcal{G}$. □

Remark 5.5. *From Definition 4.1, we know that the ACE of X on Y is identifiable if $f(y|do(x))$ is identifiable in an MPDAG or CPDAG $\mathcal{G}$. Furthermore, Theorem 5.3 gives the conditions for when $f(y|do(x))$ is identifiable*

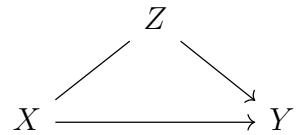

Figure 8: An MPDAG $\mathcal{G}$.

in $\mathcal{G}$. We can notice that these conditions are different from those of Theorem 5.4. In particular, Theorem 5.4 implies that the CDE of X on Y may be identifiable in a CPDAG or MPDAG $\mathcal{G}$ even when the ACE of X on Y is not identifiable in $\mathcal{G}$.

Example 5.6. *For an example of the case when the CDE of X on Y is identifiable in a CPDAG or MPDAG, but the ACE of X on Y is not identifiable case consider MPDAG $\mathcal{G}$ in Figure 8. The interventional density $f(y|do(x))$ is not identifiable in $\mathcal{G}$, due to the possibly causal path $X - Z \to Y$. However, since there is no undirected edge connected to Y in $\mathcal{G}$, $f(y|do(x), do(z))$ is identifiable in $\mathcal{G}$ and therefore, the CDE of X on Y is identifiable given $\mathcal{G}$.*

Consider now the ANDE of X on Y in CPDAG or MPDAG $\mathcal{G} = (\mathbf{V}, \mathbf{E})$. If $\mathbf{V}$ is generated by a linear SCM compatible with a DAG in $[\mathcal{G}]$, the following result for the ANDE follows directly from Proposition 4.12.

Corollary 5.7. *Let X and Y be two distinct nodes in a CPDAG or MPDAG $\mathcal{G} = (\mathbf{V}, \mathbf{E})$. Suppose that $X \to Y$ is in $\mathcal{G}$ and that $\mathbf{V}$ is generated by a linear SCM compatible with a DAG in $[\mathcal{G}]$. The ANDE of X on Y is identifiable in $\mathcal{G}$ if and only if there is no undirected edge connected to Y in $\mathcal{G}$. That is, the ANDE of X on Y is identifiable in $\mathcal{G}$ if and only if the CDE of X on Y is identifiable in $\mathcal{G}$. Furthermore, if the ANDE of X on Y is identifiable, it is equal to the CDE of X on Y in $\mathcal{G}$.*

6 Discussion and Future Work

We gave graphical definitions of the CDE and the ANDE, as well as graphical conditions for when these two causal effects are identifiable in a DAG (see Section 4). We further explored identfiability of the CDE given an equivalence class of DAGs that can be represented by a CPDAG or MPDAG $\mathcal{G}$ (see

Section 5). In case of a linear generating mechanism, we have seen that the CDE and ANDE become equivalent (Proposition 4.3 and Corollary 5.7).

For future work, we consider exploring an analogous result to Theorem 5.4 for CDEs in the case of the ANDE. Furthermore, it is of interest to further compare the identifiability conditions for the CDE, ANDE and ACE. For example, we conjecture that the ANDE of X on Y is identifiable given a CPDAG or MPDAG $\mathcal{G}$ if and only if the CDE and the ACE of X on Y are identifiable in $\mathcal{G}$.

Another avenue of interest are considerations of estimating the set of all possible causal effects (controlled or average natural) given a CPDAG or MPDAG $\mathcal{G}$, when the causal effect is not uniquely identifiable. The *Intervention-calculus when the DAG is Absent* (IDA) algorithm, developed in (Maathuis et al., 2009) is a method for determining the set of possible total causal effects given a CPDAG or MPDAG $\mathcal{G}$. The algorithm works by determining the total causal effect in every DAG in $[\mathcal{G}]$ and returns this set of possible total effects. We believe that this algorithm can be leveraged to develop a similar principle that could be applied to the determine a set of possible CDEs or ANDEs given $\mathcal{G}$.

Furthremore, even though we focus on identifying direct causal effects given a DAG, CPDAG, or MPDAG, it is possible to identify the CDE by only knowing the local neighborhood of Y (see Section 5). Hence, a similar type of algorithm to IDA, but applied to the *Markov blanket* of Y rather than the entire graph should be enough to estimate the CDE. The Markov blanket of a node contains its parents, children and the parents of its children. Examples of causal discovery algorithms that focus on Markov blanket discovery specifically are given in (Pellet and Elisseeff, 2008; Aliferis et al., 2010).

References

Aliferis, C. F., Statnikov, A., Tsamardinos, I., Mani, S., and Koutsoukos, X. D. (2010). Local causal and Markov blanket induction for causal discovery and feature selection for classification part I: Algorithms and empirical evaluation. *Journal of Machine Learning Research*, 11(Jan):171–234.

Andersson, S. A., Madigan, D., and Perlman, M. D. (1997). A characteriza-

tion of Markov equivalence classes for acyclic digraphs. *Annals of Statistics*, 25:505–541.

Chickering, D. M. (2002). Optimal structure identification with greedy search. *Journal of Machine Learning Research*, 3(Nov):507–554.

Henckel, L., Perković, E., and Maathuis, M. H. (2019). Graphical criteria for efficient total effect estimation via adjustment in causal linear models. arXiv:1907.02435.

Huang, Y. and Valtorta, M. (2006). Pearl's calculus of intervention is complete. In *Proceedings of UAI 2006*, pages 217–224.

Imai, K., Tingley, D., and Yamamoto, T. (2013). Experimental designs for identifying causal mechanisms. *Journal of the Royal Statistical Society: Series A*, 176(1):5–51.

Jaber, A., Zhang, J., and Bareinboim, E. (2019). Causal identification under Markov equivalence: Completeness results. In *Proceedings of ICML 2019*, volume 97, pages 2981–2989.

Kalisch, M. and Bühlmann, P. (2014). Causal structure learning and inference: A selective review. *Quality Technology & Quantitative Management*, 11(1):3–21.

Kaufman, S., Kaufman, J. S., and MacLehose, R. F. (2009). Analytic bounds on causal risk differences in directed acyclic graphs involving three observed binary variables. *Journal of Statistical Planning and Inference*, 139(10):3473–3487.

Lauritzen, S. L. (1996). *Graphical models*. Clarendon Press; Oxford University Press, New York, New York, first edition.

Maathuis, M. H., Kalisch, M., and Bühlmann, P. (2009). Estimating high-dimensional intervention effects from observational data. *Annals of Statistics*, 37:3133–3164.

Meek, C. (1995). Causal inference and causal explanation with background knowledge. In *Proceedings of UAI 1995*, pages 403–410.

Pearl, J. (1986). Fusion, propagation, and structuring in belief networks. *Artificial Intelligence*, 29:241–288.

Pearl, J. (2001). Direct and indirect effects. In *Proceedings of UAI 2001*, pages 411–420.

Pearl, J. (2009). *Causality: Models, reasoning, and inference*. Cambridge University Press, New York, NY, second edition.

Pellet, J.-P. and Elisseeff, A. (2008). Using Markov blankets for causal structure learning. *Journal of Machine Learning Research*, 9(Jul):1295–1342.

Perković, E. (2020). Identifying causal effects in maximally oriented partially directed acyclic graphs. In *Proceedings of UAI 2020*. To appear.

Perković, E., Kalisch, M., and Maathuis, M. H. (2017). Interpreting and using CPDAGs with background knowledge. In *Proceedings of UAI 2017*.

Petersen, M. L., Sinsi, S. E., and van der Laan, M. J. (2006). Estimation of direct causal effects. *Epidemiology*, 17(3):276–284.

Richardson, T. S., Evans, R. J., Robins, J. M., and Shpitser, I. (2017). Nested Markov properties for acyclic directed mixed graphs. arXiv preprint arXiv:1701.06686.

Richardson, T. S. and Spirtes, P. (2002). Ancestral graph Markov models. *Annals of Statistics*, 30:962–1030.

Robins, J., Richardson, T., and Spirtes, P. (2009). On identification and inference for direct effects. *Epidemiology*.

Robins, J. M. (1986). A new approach to causal inference in mortality studies with a sustained exposure period-application to control of the healthy worker survivor effect. *Mathematical Modelling*, 7:1393–1512.

Robins, J. M. and Greenland, S. (1992). Identifiability and exchangeability for direct and indirect effects. *Epidemiology*, 3(2):143–155.

Robins, J. M. and Richardson, T. S. (2010). Alternative graphical causal models and the identification of direct effects. *Causality and psychopathology: Finding the determinants of disorders and their cures*, pages 103–158.

Shpitser, I. and Pearl, J. (2006). Identification of joint interventional distributions in recursive semi-Markovian causal models. In *Proceedings of AAAI 2006*, pages 1219–1226.

Spirtes, P., Glymour, C., and Scheines, R. (2000). *Causation, prediction, and search*. MIT Press, Cambridge, MA, second edition.

Tchetgen, E. J. T. and Phiri, K. (2014). Bounds for pure direct effect. *Epidemiology*, 25(5):775.

VanderWeele, T. J. (2013). A three-way decomposition of a total effect into direct, indirect, and interactive effects. *Epidemiology*, 24(2):224.

Wright, S. (1934). The method of path coefficients. *Annals of Mathematical Statistics*, 5(3):161–215.

Zhang, J. (2008). Causal reasoning with ancestral graphs. *Journal of Machine Learning Research*, 9:1437–1474.