SUPPLEMENTARY MATERIAL

A TWIN NETWORKS AND COUNTERFACTUALS

In addition to probabilistic and causal reasoning about interventions, ioSCMs allow for counterfactual reasoning. Given an ioSCM M with graph $G^+ = (V\dot{\cup}U\dot{\cup}J, E^+)$, a set $W \subseteq V \cup J$ and the corresponding intervened ioSCM $M_{\mathrm{do}(W)}$ with graph $G^+_{\mathrm{do}(W)}$ one can construct a (merged) *twin ioSCM* M_{twin} similarly to the acyclic case (see [24]), or a single world intervention graph (SWIG, see [30]). This is done by identifying/merging the corresponding nodes, mechanisms and variables from the non-descendants of W, i.e., $\mathrm{NonDesc}^{G^+}(W)$ and $\mathrm{NonDesc}^{G^+_{\mathrm{do}(W)}}(W)$, which are unchanged by the action $\mathrm{do}(W)$. Then one has the two different branches $\mathrm{Desc}^{G^+}(W)$ and $\mathrm{Desc}^{G^+_{\mathrm{do}(W)}}(W)$ in the network. This construction then allows one to formulate counterfactual statements like in the acyclic case (see [24]), but now for general ioSCMs. E.g., one could state the assumption of *strong ignorability* (see [24, 31]) as:

$$\left(Y^{\mathrm{do}(\varnothing)}, Y^{\mathrm{do}(X)}\right) \underset{G_{\text{twin}}}{\overset{\sigma}{\perp\!\!\!\perp}} X \,|\, Z,$$

or the *conditional ignorability* (see [31, 32]) as:

$$Y^{\mathrm{do}(X)} \underset{G_{\text{twin}}}{\overset{\sigma}{\perp\!\!\!\perp}} X \,|\, Z.$$

All the causal reasoning rules derived in this paper can thus also be applied to reason about counterfactuals.

B MARGINALIZATION OF DIRECTED MIXED GRAPHS

For completeness, we provide here the definition of marginalization of directed mixed graph. For more details and the relationship with the marginalization of an mSCM (or as a straightforward generalization, an ioSCM), we refer the reader to [10].

Definition B.1 (Marginalization of DMGs)**.** *Let $G = (V, E, B)$ be a directed mixed graph (DMG) with set of nodes V, directed edges E and bidirected edges B. Let $W \subseteq V$ be a subset of nodes. We define the* marginalized DMG *$G^{\setminus W} := G' = (V', E', B')$ ("marginalizing out W"), also called* latent projection of G onto $V \setminus W$, *with set of nodes $V' := V \setminus W$ via the following rules (for $v_1, v_2 \in V \setminus W = V'$):*

1. *$v_1 \longrightarrow v_2 \in E'$ iff there exist $k \geq 0$ nodes $w_1, \dots, w_k \in W$ such that the directed walk:*

$$v_1 \longrightarrow w_1 \longrightarrow \cdots \longrightarrow w_k \longrightarrow v_2$$

lies in G (the corner case $v_1 \longrightarrow v_2 \in E$ also applies).

2. *$v_1 \longleftrightarrow v_2 \in B'$ iff there exist $k \geq 0$ nodes $w_1, \dots, w_k \in W$ and an index $0 \leq m \leq k$ such that a walk of the form:*

$$v_1 \longleftarrow w_1 \longleftarrow \cdots \longleftarrow w_m \longrightarrow \cdots \longrightarrow w_k \longrightarrow v_2$$

lies in G with $m \geq 1$ or a walk of the form:

$$v_1 \underbrace{\longleftarrow w_1 \longleftarrow \cdots \longleftarrow w_m}_{m \geq 0} \longleftrightarrow \underbrace{w_{m+1} \longrightarrow \cdots \longrightarrow w_k \longrightarrow}_{k-m \geq 0} v_2$$

lies in G (including the corner cases $v_1 \longleftrightarrow v_2 \in B$ and $v_1 \longleftarrow w \longrightarrow v_2$ in G with $w \in W$).

C CONDITIONAL INDEPENDENCE AND ITS ALTERNATIVE WITH CONFOUNDED INPUTS

Here we want to give a generalization of [3, 29] in the flavor of definition 3.1. The main point is that the approaches of conditional independence for families of distributions/Markov kernels in [3, 29] implicitely assume that the input variables J are jointly confounded. The definition 3.1 of conditional independence, in contrast, assumes (via the product distributions) that the variables J are jointly independent. The approach in definition 3.1 can be easily adapted to the confounded input setting as follows.

C.1 INPUT CONFOUNDED CONDITIONAL INDEPENDENCE

Definition C.1 (Input confounded conditional independence)**.** *Let $\mathcal{X}_V := \prod_{v \in V} \mathcal{X}_v$ and $\mathcal{X}_J := \prod_{j \in J} \mathcal{X}_j$ be the product spaces of any measurable spaces and*

$$\mathbb{P}_V(X_V|X_J)$$

a Markov kernel (i.e. a family of distributions on $\mathcal{X}_V$ measurably[5] *parametrized by $\mathcal{X}_J$). For subsets $A, B, C \subseteq V\dot{\cup}J$ we write:*

$$X_A \underset{\mathbb{P}_V(X_V|X_J),\bullet}{\perp\!\!\!\perp} X_B \,|\, X_C$$

if and only if for every joint *distribution $\mathbb{P}_J$ on $\mathcal{X}_J$ we have:*

$$X_A \underset{\mathbb{P}_{V \cup J}}{\perp\!\!\!\perp} X_B \,|\, X_C,$$

which means that for all measurable $F \subseteq \mathcal{X}_A$ we have:

$$\mathbb{P}_{V \cup J}(X_A \in F|X_B, X_C) = \mathbb{P}_{V \cup J}(X_A \in F|X_C) \quad \mathbb{P}_{V \cup J}\text{-a.s.},$$

[5] We require that for every measurable $F \subseteq \mathcal{X}_V$ the map $\mathcal{X}_J \to [0, 1]$ given by $x_J \mapsto \mathbb{P}_V(X_V \in F|X_J = x_J)$ is measurable.

where $\mathbb{P}_{V\cup J}(X_{V\cup J}) := \mathbb{P}_V(X_V|X_J) \otimes \mathbb{P}_J(X_J)$, the distribution given by $X_J \sim \mathbb{P}_J$ and then $X_V \sim \mathbb{P}_V(\_|X_J)$.

Lemma C.2. *Let the situation be like in C.1 and assume all spaces $\mathcal{X}_v$, $v \in V$, to be standard measurable spaces. Let A, B, C be pairwise disjoint, $A \cap J = \emptyset$ and $J \subseteq B \cup C$. Then every statement implies the one below:*

1. *There is a version of $\mathbb{P}_V(X_A|X_B, X_C)$ such that for all $x_B, x'_B \in \mathcal{X}_B$, $x_C \in \mathcal{X}_C$:*
$$\begin{aligned}&\mathbb{P}_V(X_A|X_B = x_B, X_C = x_C)\\ &= \mathbb{P}_V(X_A|X_B = x'_B, X_C = x_C).\end{aligned}$$
2. $X_A \underset{\mathbb{P}_V(X_V|X_J),\bullet}{\perp\!\!\!\perp} X_B \mid X_C$.
3. $X_A \underset{\mathbb{P}_V(X_V|X_J)}{\perp\!\!\!\perp} X_B \mid X_C$ *(using definition 3.1).*
4. $X_A \underset{\mathbb{P}_V(X_V|X_J)\otimes\delta_{x_J}(X_J)}{\perp\!\!\!\perp} X_B \mid X_C$ *for every $x_J \in \mathcal{X}_J$.*

If there is a Markov kernel $\mathbb{P}(X_A|X_C)$ that is a version of $\mathbb{P}_{V\cup J}(X_A|X_C)$ for every Dirac delta distribution $\mathbb{P}_J = \delta_{x_J}$ (e.g. if $J \subseteq C$) then the last point also implies the first.

Proof. 1. $\implies$ 2.: Functional dependence only on x_C.
2. $\implies$ 3. $\implies$ 4.: Every product distribution is a joint distribution and every Dirac delta distribution is a product distribution.
1. $\impliedby$ 4.: Let $N \subseteq \mathcal{X}_{B\cup C}$ be the measurable set on which the Markov kernels $\mathbb{P}_V(X_A|X_B, X_C)$ and $\mathbb{P}(X_A|X_C)$ (considered as functions of (x_B, x_C)) differ. For every $x_J \in \mathcal{X}_J$ we have by assumption:
$$X_A \underset{\mathbb{P}_V(X_V|X_J)\otimes\delta_{x_J}(X_J)}{\perp\!\!\!\perp} X_B \mid X_C.$$
This shows that:
$$\mathbb{P}_V(X_A|X_B = x_B, X_C = x_C) = \mathbb{P}(X_A|X_C = x_C)$$
for (x_B, x_C) outside of a $\mathbb{P}_V(X_{(B\cup C)\setminus J}|X_J = x_J)$-zero set, for which we can take the section N_{x_J} of N. This implies that N is a $\mathbb{P}_V(X_{(B\cup C)\setminus J}|X_J)$-zero set. So $\mathbb{P}(X_A|X_C)$ is a version of $\mathbb{P}_V(X_A|X_B, X_C)$ and satisfies 1.. □

Remark C.3.
1. *The existence of the Markov kernel $\mathbb{P}(X_A|X_C)$ under the assumption 4. in lemma C.2 always/only holds up to measurability questions, because for every fixed $\mathbb{P}_J$ the regular conditional probability distribution $\mathbb{P}_{V\cup J}(X_A|X_B, X_C)$ always exists in standard measurable spaces and agrees with $\mathbb{P}_{V\cup J}(X_A|X_C)$ (by the assumption 4.). The existence of the Markov kernel $\mathbb{P}(X_A|X_C)$ follows for standard measurable spaces $\mathcal{X}_v$, $v \in V$, if either:*
   - *(a) $J \subseteq C$ and assumption 4. holds, or:*
   - *(b) $\mathcal{X}_J$ is discrete and assumption 2. holds, or:*
   - *(c) $\mathbb{P}_V(X_V|X_J)$ comes as $\mathbb{P}_U(X_V|X_J)$ from an ioSCM and assumptions 2.-4. even hold in form of the corresponding σ-separation statement in the induced DMG G.*

   We plan in future work to address all these subtleties in more detail.
2. *Lemma C.2 shows that definition C.1 (and also already definition 3.1) generalizes the one from [29] (when applied symmetrized). The clear correspondence/generalization is that for any (not necessarily disjoint) $A, B, C \subseteq V \cup J$:*
$$\begin{aligned} & X_A \underset{[29]}{\perp\!\!\!\perp} X_B \mid X_C \\ :\iff \quad & X_A \underset{\mathbb{P}_V(X_V|X_J),\bullet}{\perp\!\!\!\perp} X_{B\cup J} \mid X_C \\ \vee \quad & X_B \underset{\mathbb{P}_V(X_V|X_J),\bullet}{\perp\!\!\!\perp} X_{A\cup J} \mid X_C. \end{aligned}$$
3. *Thm. 4.4 in [3] shows that definitions 3.1, C.1 also generalize the one from [3] in the same sense.*
4. *In contrast with [3,6,29], definition C.1 can accommodate any variable from V or J at any position of the conditional independence statement.*
5. *Also note that $\perp\!\!\!\perp_{\mathbb{P}_V(X_V|X_J),\bullet}$ is well-defined for any measurable spaces and is not restricted to discrete variables or distributions/Markov kernels that come with densities.*
6. *Furthermore, $\perp\!\!\!\perp_{\mathbb{P}_V(X_V|X_J),\bullet}$ satisfies the* separoid axioms *(see [6,13,25] or see rules 1-5 in Lem. 4.5 for $\perp\!\!\!\perp_{\mathbb{P}_V(X_V|X_J),\bullet}$). Indeed, every single $\perp\!\!\!\perp_{\mathbb{P}_{V\cup J}}$ satisfies the separoid axioms (see [3,6]) and an arbitrary intersection of separoids is again a separoid (see [7]):*
$$\left\langle \underset{\mathbb{P}_V(X_V|X_J),\bullet}{\perp\!\!\!\perp} \right\rangle = \bigcap_{\mathbb{P}_J} \left\langle \underset{\mathbb{P}_{V\cup J}}{\perp\!\!\!\perp} \right\rangle.$$

C.2 INPUT CONFOUNDED GLOBAL MARKOV PROPERTY

We can also prove a global Markov property for the input confounded version of conditional independence. For this we need to modify the graphical structures a bit and introduce a few more notations. Note that all spaces are assumed to measurable (but not necessarily standard).

Definition C.4 (Input confounded ioSCM)**.** *Let $M = (G^+, \mathcal{X}, \mathbb{P}_U, g)$ be an ioSCM with graph $G^+ = (V\dot\cup U\dot\cup J, E^+)$. The corresponding* input confounded ioSCM *$M_\bullet$ is then constructed from M by the following changes:*

1. *$V_\bullet := V \cup J$ and $U_\bullet := U$,*
2. *$J_\bullet := \{\bullet\}$ with a new node $\bullet$ with space $\mathcal{X}_\bullet := \mathcal{X}_J$,*

3. $E^+_\bullet := E^+ \cup \{\bullet \longrightarrow j \,|\, j \in J\}$,
4. *add* $g_{\{j\}}$*, the canonical projection from* $\mathcal{X}_\bullet$ *onto* $\mathcal{X}_j$*, to* g *for* $j \in J$.

With this setting $M_\bullet$ *is a well-defined ioSCM.*
Furthermore, let $G_\bullet$ *be the* input confounded induced DMG*, i.e. the induced DMG of* $G^+_\bullet$ *where* $\bullet$ *is marginalized out. In other words,* $G_\bullet$ *arises from the induced DMG* G *of* G^+ *by just adding* $j_1 \leftrightarrow j_2$ *for all* $j_1, j_2 \in J$*,* $j_1 \neq j_2$*, to* G*.*

Theorem C.5 (Input confounded directed global Markov property)**.** *Let* M *be an ioSCM with input confounded induced DMG* $G_\bullet$*. Then for all subsets* $A, B, C \subseteq V \cup J$ *we have the implication:*

$$A \underset{G_\bullet}{\overset{\sigma}{\perp\!\!\!\perp}} B \,|\, C \implies X_A \underset{\mathbb{P}_U(X_V \,|\, \mathrm{do}(X_J)),\bullet}{\perp\!\!\!\perp} X_B \,|\, X_C.$$

In words, if A *and* B *are* σ*-separated by* C *in* $G_\bullet$ *then the corresponding variables* X_A *and* X_B *are conditionally independent given* X_C *for any distribution* $\mathbb{P}_U(X_V \,|\, \mathrm{do}(X_J)) \otimes \mathbb{P}_J(X_J)$ *for any* joint *distribution* $\mathbb{P}_J$ *on* $\mathcal{X}_J$*.*

Proof. This directly follows from the σ-separation criterion/global Markov property 5.2 applied to the input confounded ioSCM $M_\bullet$ and $G^+_\bullet$, or, alternatively, again from the mSCM-version proven in [10,11] for each fixed joint distribution $\mathbb{P}_J$ on $\mathcal{X}_J = \mathcal{X}_\bullet$. Note that $G_\bullet$ is a marginalization of $G^+_\bullet$ and σ-separation is stable under marginalization. $\square$

D THE EXTENDED IOSCM - PROOFS

Proposition D.1. *Let* $M = (G^+, \mathcal{X}, \mathbb{P}_U, g)$ *be an ioSCM with* $G^+ = (V \dot\cup U \dot\cup J, E^+)$ *and* $\hat{M}$ *the extended ioSCM. Let* $A, B, C \subseteq V$ *be pairwise disjoint set of nodes and* $x_{C \cup J} \in \mathcal{X}_{C \cup J}$*. Then we have the equations:*

$$\begin{aligned}&\mathbb{P}_U(X_A | X_B, \mathrm{do}(X_{C\cup J} = x_{C \cup J}))\\ =\ &\mathbb{P}_U(X_A | X_B, I_C = x_C, X_J = x_J)\\ =\ &\mathbb{P}_U(X_A | X_B, I_C = x_C, X_C = x_C, X_J = x_J).\end{aligned}$$

Proof. Consider the first equality. For any subset $D \subseteq V$ the variable $X_D^{\mathrm{do}(X_{C\cup J} = x_{C\cup J})}$ was recursively defined in $M_{\mathrm{do}(C)}$ via g using $G^+_{\mathrm{do}(C)}$, whereas the variable $X_D^{\mathrm{do}((I_C, I_{V\setminus C}, X_J) = (x_C, \varnothing_{V\setminus C}, x_J))}$ was recursively defined in $\hat{M}$ via the same g but using $I(x_C, \varnothing_{V\setminus C})$ and $G^+_{\mathrm{do}(I(x_C, \varnothing_{V\setminus C}))}$. Since $x_C \in \mathcal{X}_C$ we have that $I(x_C, \varnothing_{V\setminus C}) = C$ and thus $G^+_{\mathrm{do}(I(x_C, \varnothing_{V\setminus C}))} = G^+_{\mathrm{do}(C)}$. It directly follows that:

$$X_D^{\mathrm{do}(X_{C\cup J} = x_{C\cup J})} = X_D^{\mathrm{do}((I_C, I_{V\setminus C}, X_J) = (x_C, \varnothing_{V\setminus C}, x_J))}.$$

This shows the equality of top and middle line. For the equality between the middle and bottom line note that:

$$I_C = x_C \overset{x_C \in \mathcal{X}_C}{\Longrightarrow} X_C = x_C. \qquad \square$$

E THE THREE MAIN RULES OF CAUSAL CALCULUS - PROOFS

Theorem E.1 (The three main rules of causal calculus)**.** *Let* M *be an ioSCM with set of observed nodes* V *and intervention nodes* J *and induced DMG* G*. Let* $X, Y, Z \subseteq V$ *and* $J \subseteq W \subseteq V \cup J$ *be subsets.*

1. *Insertion/deletion of observation:*
   If $\quad Y \underset{G}{\overset{\sigma}{\perp\!\!\!\perp}} X | Z, \mathrm{do}(W) \quad$ *then:*
   $$\mathbb{P}(Y | X, Z, \mathrm{do}(W)) = \mathbb{P}(Y | Z, \mathrm{do}(W)).$$
2. *Action/observation exchange:*
   If $\quad Y \underset{G}{\overset{\sigma}{\perp\!\!\!\perp}} I_X | X, Z, \mathrm{do}(W) \quad$ *then:*
   $$\mathbb{P}(Y | \mathrm{do}(X), Z, \mathrm{do}(W)) = \mathbb{P}(Y | X, Z, \mathrm{do}(W)).$$
3. *Insertion/deletion of actions:*
   If $\quad Y \underset{G}{\overset{\sigma}{\perp\!\!\!\perp}} I_X | Z, \mathrm{do}(W) \quad$ *then:*
   $$\mathbb{P}(Y | \mathrm{do}(X), Z, \mathrm{do}(W)) = \mathbb{P}(Y | Z, \mathrm{do}(W)).$$

Proof.

1. Thm. 5.2 applied to $G_{\mathrm{do}(W)}$ gives:
   $$Y \underset{G}{\overset{\sigma}{\perp\!\!\!\perp}} X | Z, \mathrm{do}(W) \overset{5.2}{\Longrightarrow} Y \underset{\mathbb{P}}{\perp\!\!\!\perp} X | Z, \mathrm{do}(W).$$
   The latter directly gives the claim:
   $$\mathbb{P}(Y | X, Z, \mathrm{do}(W)) = \mathbb{P}(Y | Z, \mathrm{do}(W)).$$
2. The σ-separation criterion 5.2 w.r.t. to $\hat{G}_{\mathrm{do}(W)}$ gives:
   $$Y \underset{G}{\overset{\sigma}{\perp\!\!\!\perp}} I_X | X, Z, \mathrm{do}(W) \overset{5.2}{\Longrightarrow} Y \underset{\mathbb{P}}{\perp\!\!\!\perp} I_X | X, Z, \mathrm{do}(W).$$
   Together with Prp. 6.2 (applied to $M_{\mathrm{do}(W)}$) we have:
   $$\begin{aligned}&\mathbb{P}(Y | \mathrm{do}(X), Z, \mathrm{do}(W))\\ \overset{6.2}{=}\ &\mathbb{P}(Y | I_X, X, Z, \mathrm{do}(W))\\ \underset{Y \perp\!\!\!\perp I_X | X, Z, \mathrm{do}(W)}{=}\ &\mathbb{P}(Y | X, Z, \mathrm{do}(W)).\end{aligned}$$
3. As before we have:
   $$Y \underset{G}{\overset{\sigma}{\perp\!\!\!\perp}} I_X | Z, \mathrm{do}(W) \overset{5.2}{\Longrightarrow} Y \underset{\mathbb{P}}{\perp\!\!\!\perp} I_X | Z, \mathrm{do}(W).$$
   And again: $\quad \mathbb{P}(Y | \mathrm{do}(X), Z, \mathrm{do}(W))$
   $$\begin{aligned}\overset{6.2}{=}\ &\mathbb{P}(Y | I_X, Z, \mathrm{do}(W))\\ \underset{Y \perp\!\!\!\perp I_X | Z, \mathrm{do}(W)}{=}\ &\mathbb{P}(Y | Z, \mathrm{do}(W)). \qquad \square\end{aligned}$$

F ADJUSTMENT CRITERIA

F.1 PROOFS

Theorem F.1 (General adjustment criterion and formula)**.** *Let the setting be like in 8.1. Assume that data was collected under selection bias, $\mathbb{P}(V|S=s,\mathrm{do}(W))$ (or under $\mathbb{P}(V|\,\mathrm{do}(W))$ and $S=\emptyset$), and there are unbiased samples from $\mathbb{P}(Z|C,\mathrm{do}(W))$. Further assume that the variables satisfy:*

1. $(Z_0,L) \overset{\sigma}{\underset{G}{\perp\!\!\!\perp}} I_X \mid C,\mathrm{do}(W)$*, and*
2. $Y \overset{\sigma}{\underset{G}{\perp\!\!\!\perp}} (I_X,Z_+) \mid C,X,Z_0,L,\mathrm{do}(W)$*, and*
3. $Y \overset{\sigma}{\underset{G}{\perp\!\!\!\perp}} S \mid C,X,Z,\mathrm{do}(W)$*, and*
4. $L \overset{\sigma}{\underset{G}{\perp\!\!\!\perp}} X \mid C,Z,\mathrm{do}(W)$.

Then one can estimate the conditional causal effect $\mathbb{P}(Y|C,\mathrm{do}(X),\mathrm{do}(W))$ via the adjustment formula*:*

$$\mathbb{P}(Y|C,\mathrm{do}(X),\mathrm{do}(W)) = \int \mathbb{P}(Y|X,Z,C,S=s,\mathrm{do}(W))\, d\mathbb{P}(Z|C,\mathrm{do}(W)).$$

Proof. Since C, $\mathrm{do}(W)$ occur everywhere as a conditioning set, we will suppress C, $\mathrm{do}(W)$ in the following everywhere. Then note that the σ-separation criterion 5.2 implies the corresponding conditional independencies in the following when indicated. The adjustment formula then derives from the following computations:

$$\begin{aligned}
&\mathbb{P}(Y|\,\mathrm{do}(X)) \\
=\; &\int \mathbb{P}(Y|Z_0,L,\mathrm{do}(X))\, d\mathbb{P}(Z_0,L|\,\mathrm{do}(X)) \\
\overset{6.2}{=}\; &\int \mathbb{P}(Y|I_X,X,Z_0,L)\, d\mathbb{P}(Z_0,L|I_X) \\
\overset{Y \perp\!\!\!\perp I_X|X,Z_0,L;\ (Z_0,L)\perp\!\!\!\perp I_X}{=}\; &\int \mathbb{P}(Y|X,Z_0,L)\, d\mathbb{P}(Z_0,L) \\
\overset{\int d\mathbb{P}(Z_+|Z_0,L)=1}{=}\; &\int\int \mathbb{P}(Y|X,Z_0,L)\, d\mathbb{P}(Z_+|Z_0,L)\, d\mathbb{P}(Z_0,L) \\
\overset{Y \perp\!\!\!\perp Z_+|X,Z_0,L}{=}\; &\int \mathbb{P}(Y|X,Z_0,Z_+,L)\, d\mathbb{P}(Z_+,Z_0,L) \\
\overset{Z=Z_+\cup Z_0}{=}\; &\int \mathbb{P}(Y|X,Z,L)\, d\mathbb{P}(Z,L) \\
=\; &\int\int \mathbb{P}(Y|X,Z,L)\, d\mathbb{P}(L|Z)\, d\mathbb{P}(Z) \\
\overset{L \perp\!\!\!\perp X|Z}{=}\; &\int\int \mathbb{P}(Y|L,X,Z)\, d\mathbb{P}(L|X,Z)\, d\mathbb{P}(Z) \\
=\; &\int \mathbb{P}(Y|X,Z)\, d\mathbb{P}(Z) \\
\overset{Y \perp\!\!\!\perp S|X,Z}{=}\; &\int \mathbb{P}(Y|X,Z,S)\, d\mathbb{P}(Z). \quad \square
\end{aligned}$$

F.2 SPECIAL CASES

Corollary F.2. *Let the notations be like in 8.1 and 8.2 and $W=J=\emptyset$. We have the following special cases, which in the* acyclic *case will reduce to the ones given by the indicated references:*

1. *General selection-backdoor (see [4]): $C=\emptyset$, and*
   (a) $(Z_0,L) \overset{\sigma}{\underset{G}{\perp\!\!\!\perp}} I_X$*, and*
   (b) $Y \overset{\sigma}{\underset{G}{\perp\!\!\!\perp}} (I_X,Z_+) \mid X,Z_0,L$*, and*
   (c) $Y \overset{\sigma}{\underset{G}{\perp\!\!\!\perp}} S \mid X,Z$*, and*
   (d) $L \overset{\sigma}{\underset{G}{\perp\!\!\!\perp}} X \mid Z$*, implies:*
   $$\mathbb{P}(Y|\,\mathrm{do}(X)) = \int \mathbb{P}(Y|X,Z,S=s)\, d\mathbb{P}(Z).$$
2. *Selection-backdoor (see [1]): $C=L=\emptyset$, and*
   (a) $Z_0 \overset{\sigma}{\underset{G}{\perp\!\!\!\perp}} I_X$*, and*
   (b) $Y \overset{\sigma}{\underset{G}{\perp\!\!\!\perp}} (I_X,Z_+,S) \mid X,Z_0$ *implies:*
   $$\mathbb{P}(Y|\,\mathrm{do}(X)) = \int \mathbb{P}(Y|X,Z,S=s)\, d\mathbb{P}(Z).$$
3. *Extended backdoor[6] (see [26, 32]): $C=S=\emptyset$,*
   (a) $(Z_0,L) \overset{\sigma}{\underset{G}{\perp\!\!\!\perp}} I_X$*, and*
   (b) $Y \overset{\sigma}{\underset{G}{\perp\!\!\!\perp}} (I_X,Z_+) \mid X,Z_0,L$*, and*
   (c) $L \overset{\sigma}{\underset{G}{\perp\!\!\!\perp}} X \mid Z$*, implies:*
   $$\mathbb{P}(Y|\,\mathrm{do}(X)) = \int \mathbb{P}(Y|X,Z)\, d\mathbb{P}(Z).$$
4. *Backdoor (see [21, 22, 24]): $C=S=L=Z_+=\emptyset$,*
   (a) $Z \overset{\sigma}{\underset{G}{\perp\!\!\!\perp}} I_X$*, and*
   (b) $Y \overset{\sigma}{\underset{G}{\perp\!\!\!\perp}} I_X \mid X,Z$*, implies:*
   $$\mathbb{P}(Y|\,\mathrm{do}(X)) = \int \mathbb{P}(Y|X,Z)\, d\mathbb{P}(Z).$$

F.3 MORE ON ADJUSTMENT CRITERIA

The following generalizes the adjustment criterion of type I in [4].

[6] In the acyclic case it was shown in [32] that when L is allowed to represent latent variables in a graph G' that marginalizes to G then this criterion actually characterizes all adjustment sets for G and $\mathbb{P}(Y|\,\mathrm{do}(X))$.

Theorem F.3 (General adjustment without external data)**.** *Let the setting be like in 8.1. Assume that data was collected under selection bias, $\mathbb{P}(V|S=s)$. Further assume that the variables satisfy:*

1. $Y \underset{G}{\overset{\sigma}{\perp\!\!\!\perp}} S \,|\, \mathrm{do}(X)$,
2. $Z_0 \underset{G}{\overset{\sigma}{\perp\!\!\!\perp}} I_X \,|\, S$,
3. $Y \underset{G}{\overset{\sigma}{\perp\!\!\!\perp}} Z_+ \,|\, Z_0, S, \mathrm{do}(X)$,
4. $Y \underset{G}{\overset{\sigma}{\perp\!\!\!\perp}} I_X \,|\, X, Z, S$.

Then one can estimate the causal effect $\mathbb{P}(Y|\,\mathrm{do}(X))$ via the following adjustment formula *from the biased data:*

$$\mathbb{P}(Y|\,\mathrm{do}(X)) = \int \mathbb{P}(Y|X,Z,S=s)\, d\mathbb{P}(Z|S=s).$$

Proof. First note that the σ-separation criterion Theorem 5.2 implies the corresponding conditional independencies in the following when indicated. We implicitly make use of Proposition 6.2 when needed. The adjustment formula then derives from the following computations:

$$\begin{aligned}
& \mathbb{P}(Y|\,\mathrm{do}(X)) \\
\overset{Y \perp\!\!\!\perp S \,|\, \mathrm{do}(X)}{=} & \mathbb{P}(Y|S, \mathrm{do}(X)) \\
\overset{\text{chain rule}}{=} & \int \mathbb{P}(Y|Z_0, S, \mathrm{do}(X)) \\
& d\mathbb{P}(Z_0|S, \mathrm{do}(X)) \\
\underset{6.2}{\overset{Z_0 \perp\!\!\!\perp I_X \,|\, S}{=}} & \int \mathbb{P}(Y|Z_0, S, \mathrm{do}(X))\, d\mathbb{P}(Z_0|S) \\
\overset{\int d\mathbb{P}(Z_+|Z_0,S)=1}{=} & \int \mathbb{P}(Y|Z_0, S, \mathrm{do}(X)) \\
& d\mathbb{P}(Z_+, Z_0|S) \\
\overset{Y \perp\!\!\!\perp Z_+ \,|\, Z_0, S, \mathrm{do}(X)}{=} & \int \mathbb{P}(Y|Z_+, Z_0, S, \mathrm{do}(X)) \\
& d\mathbb{P}(Z_+, Z_0|S) \\
\overset{Z=Z_+\cup Z_0}{=} & \int \mathbb{P}(Y|Z, S, \mathrm{do}(X))\, d\mathbb{P}(Z|S) \\
\underset{6.2}{\overset{Y \perp\!\!\!\perp I_X \,|\, X,Z,S}{=}} & \int \mathbb{P}(Y|Z, S, X)\, d\mathbb{P}(Z|S).
\end{aligned}$$

□

The following theorem generalizes the adjustment criterion of type III in [5]. For this we have to introduce even more adjustment sets: $Z_0^A, Z_0^B, Z_1^A, Z_1^B, Z_2, Z_3$ and L_0, L_1. We write $Z_0 = (Z_0^A, Z_0^B)$, $Z_{\leq 1}^A = (Z_0^A, Z_1^A)$, etc..

Theorem F.4 (General adjustment with partial external data)**.** *Assume that data was collected under selection bias, $\mathbb{P}(V|S=s)$, but we have unbiased data from $\mathbb{P}(Z_{\leq 1}^B)$. Further assume that the variables satisfy:*

1. $(L_0, Z_0) \perp\!\!\!\perp I_X$,
2. $Y \perp\!\!\!\perp Z_1 \,|\, L_0, Z_0, \mathrm{do}(X)$,
3. $Z_{\leq 1}^A \perp\!\!\!\perp S \,|\, Z_{\leq 1}^B$,
4. $L_0 \perp\!\!\!\perp I_X \,|\, Z_{\leq 1}$,
5. $Y \perp\!\!\!\perp S \,|\, Z_{\leq 1}, \mathrm{do}(X)$,
6. $(L_1, Z_2) \perp\!\!\!\perp I_X \,|\, S, Z_{\leq 1}$,
7. $Y \perp\!\!\!\perp Z_3 \,|\, L_1, S, Z_{\leq 2}, \mathrm{do}(X)$,
8. $L_1 \perp\!\!\!\perp I_X \,|\, S, Z$,
9. $Y \perp\!\!\!\perp I_X \,|\, X, S, Z$.

Then we have the adjustment formula: $\mathbb{P}(Y|\,\mathrm{do}(X)) =$

$$\int\int \mathbb{P}(Y|S=s, Z, X)\, d\mathbb{P}(Z \backslash Z_{\leq 1}^B | S=s, Z_{\leq 1}^B)\, d\mathbb{P}(Z_{\leq 1}^B).$$

Note that this formula does not depend on L_0 and L_1. So L_0 and L_1 can be chosen in a graph G' that marginalizes to G.

Proof.

$$\begin{aligned}
& \mathbb{P}(Y|\,\mathrm{do}(X)) \\
\overset{\text{chain rule}}{=} & \int \mathbb{P}(Y|L_0, Z_0, \mathrm{do}(X)) \\
& \quad d\mathbb{P}(L_0, Z_0|\,\mathrm{do}(X)) \\
\underset{6.2}{\overset{(L_0,Z_0) \perp\!\!\!\perp I_X}{=}} & \int \mathbb{P}(Y|L_0, Z_0, \mathrm{do}(X)) \\
& d\mathbb{P}(L_0, Z_0) \\
\underset{Z_{\leq 1}=Z_0\cup Z_1}{\overset{\int d\mathbb{P}(Z_1|L_0,Z_0)=1}{=}} & \int \mathbb{P}(Y|L_0, Z_0, \mathrm{do}(X)) \\
& d\mathbb{P}(L_0, Z_{\leq 1}) \\
\overset{Y \perp\!\!\!\perp Z_1 \,|\, L_0, Z_0, \mathrm{do}(X)}{=} & \int \mathbb{P}(Y|L_0, Z_{\leq 1}, \mathrm{do}(X)) \\
& d\mathbb{P}(L_0, Z_{\leq 1}) \\
\underset{Z_{\leq 1}=Z_{\leq 1}^A \cup Z_{\leq 1}^B}{\overset{\text{chain rule}}{=}} & \int \mathbb{P}(Y|L_0, Z_{\leq 1}, \mathrm{do}(X)) \\
& d\mathbb{P}(L_0|Z_{\leq 1})\, d\mathbb{P}(Z_{\leq 1}^A|Z_{\leq 1}^B) \\
& d\mathbb{P}(Z_{\leq 1}^B) \\
\overset{Z_{\leq 1}^A \perp\!\!\!\perp S \,|\, Z_{\leq 1}^B}{=} & \int \mathbb{P}(Y|L_0, Z_{\leq 1}, \mathrm{do}(X)) \\
& d\mathbb{P}(L_0|Z_{\leq 1})\, d\mathbb{P}(Z_{\leq 1}^A|S, Z_{\leq 1}^B) \\
& d\mathbb{P}(Z_{\leq 1}^B) \\
\underset{6.2}{\overset{L_0 \perp\!\!\!\perp I_X \,|\, Z_{\leq 1}}{=}} & \int \mathbb{P}(Y|L_0, Z_{\leq 1}, \mathrm{do}(X)) \\
& d\mathbb{P}(L_0|Z_{\leq 1}, \mathrm{do}(X)) \\
& d\mathbb{P}(Z_{\leq 1}^A|S, Z_{\leq 1}^B)\, d\mathbb{P}(Z_{\leq 1}^B) \\
\overset{\text{chain rule}}{=} & \int \mathbb{P}(Y|Z_{\leq 1}, \mathrm{do}(X)) \\
& \quad d\mathbb{P}(Z_{\leq 1}^A|S, Z_{\leq 1}^B)\, d\mathbb{P}(Z_{\leq 1}^B)
\end{aligned}$$

$$
\begin{aligned}
&\overset{Y \perp\!\!\!\perp S \mid Z_{\leq 1}, \mathrm{do}(X)}{=} && \int \mathbb{P}(Y|S, Z_{\leq 1}, \mathrm{do}(X)) \\
& && d\mathbb{P}(Z^A_{\leq 1}|S, Z^B_{\leq 1})\, d\mathbb{P}(Z^B_{\leq 1}) \\
&\overset{\text{chain rule}}{=} && \int \mathbb{P}(Y|L_1, Z_2, S, Z_{\leq 1}, \mathrm{do}(X)) \\
& && d\mathbb{P}(L_1, Z_2|S, Z_{\leq 1}, \mathrm{do}(X)) \\
& && d\mathbb{P}(Z^A_{\leq 1}|S, Z^B_{\leq 1})\, d\mathbb{P}(Z^B_{\leq 1}) \\
&\overset{Z_{\leq 2} = Z_{\leq 1} \cup Z_2}{=} && \int \mathbb{P}(Y|L_1, S, Z_{\leq 2}, \mathrm{do}(X)) \\
& && d\mathbb{P}(L_1, Z_2|S, Z_{\leq 1}, \mathrm{do}(X)) \\
& && d\mathbb{P}(Z^A_{\leq 1}|S, Z^B_{\leq 1})\, d\mathbb{P}(Z^B_{\leq 1}) \\
&\underset{6.2}{\overset{(L_1, Z_2) \perp\!\!\!\perp I_X \mid S, Z_{\leq 1}}{=}} && \int \mathbb{P}(Y|L_1, S, Z_{\leq 2}, \mathrm{do}(X)) \\
& && d\mathbb{P}(L_1, Z_2|S, Z_{\leq 1}) \\
& && d\mathbb{P}(Z^A_{\leq 1}|S, Z^B_{\leq 1})\, d\mathbb{P}(Z^B_{\leq 1}) \\
&\underset{\int \mathbb{P}(Z_3|L_1, S, Z_{\leq 2}) = 1}{\overset{Y \perp\!\!\!\perp Z_3 \mid L_1, S, Z_{\leq 2}, \mathrm{do}(X)}{=}} && \int \mathbb{P}(Y|L_1, S, Z_{\leq 2}, Z_3, \mathrm{do}(X)) \\
& && d\mathbb{P}(L_1, Z_2, Z_3|S, Z_{\leq 1}) \\
& && d\mathbb{P}(Z^A_{\leq 1}|S, Z^B_{\leq 1})\, d\mathbb{P}(Z^B_{\leq 1}) \\
&\underset{Z = Z_{\leq 2} \cup Z_3}{\overset{\text{chain rule}}{=}} && \int \mathbb{P}(Y|L_1, S, Z, \mathrm{do}(X)) \\
& && d\mathbb{P}(L_1|S, Z) \\
& && d\mathbb{P}(Z \setminus Z^B_{\leq 1}|S, Z^B_{\leq 1})\, d\mathbb{P}(Z^B_{\leq 1}) \\
&\underset{6.2}{\overset{L_1 \perp\!\!\!\perp I_X \mid S, Z}{=}} && \int \mathbb{P}(Y|L_1, S, Z, \mathrm{do}(X)) \\
& && d\mathbb{P}(L_1|S, Z, \mathrm{do}(X)) \\
& && d\mathbb{P}(Z \setminus Z^B_{\leq 1}|S, Z^B_{\leq 1})\, d\mathbb{P}(Z^B_{\leq 1})
\end{aligned}
$$

$$
\begin{aligned}
&\overset{\text{chain rule}}{=} && \int \mathbb{P}(Y|S, Z, \mathrm{do}(X)) \\
& && d\mathbb{P}(Z \setminus Z^B_{\leq 1}|S, Z^B_{\leq 1})\, d\mathbb{P}(Z^B_{\leq 1}) \\
&\overset{Y \perp\!\!\!\perp I_X \mid X, S, Z}{=} && \int \mathbb{P}(Y|S, Z, X) \\
& && d\mathbb{P}(Z \setminus Z^B_{\leq 1}|S, Z^B_{\leq 1})\, d\mathbb{P}(Z^B_{\leq 1}).
\end{aligned}
$$

□

G IDENTIFYING CAUSAL EFFECTS

Remark G.1 (More remarks about the ID-algorithm)**.**

1. *The extended version of the ID algorithm is equivalent to applying the ID algorithm to the* acyclification $G^{+,\mathrm{acy}}$ *of* G^+*, which here is meant to be the conditional ADMG that arises by adding edges* $v \longrightarrow w'$ *if* $v \notin \mathrm{Sc}^G(w) \ni w'$ *and* $v \longrightarrow w \in G^+$*, and erasing all edges inside* $\mathrm{Sc}^G(w)$*,* $w \in V$ *(see [10]).*
2. *A consolidated district in G then is the same as a district in* G^{acy}*.*
3. *Every apt-order of G is a topological order of* G^{acy}*.*
4. *So identifiability in* G^{acy} *implies identifiability in G.*
5. *This leads to the rule of thumb that causal effects where both cause and effect nodes are inside one strongly connected component of G are not identifiable from observational data alone, and, that the causal effects of sets of nodes between strongly connected components follow rules similar to the acyclic case.*
6. *Similarly, the corner cases for the identification of* conditional *causal effects* $\mathbb{P}(Y|R, \mathrm{do}(W))$ *in G that are not covered by the identification of* $\mathbb{P}(Y, R|\,\mathrm{do}(W))$ *in G follow from the (acyclic) conditional ID-algorithm from [36] applied to* G^{acy} *and then translated back to G by the above correspondences.*

Lemma G.2. *Let* $M = (G^+, \mathcal{X}, \mathbb{P}_U, g)$ *be an ioSCM with* $G^+ = (V \dot{\cup} U \dot{\cup} J, E^+)$ *and* $<$ *an apt-order for* G^+ *and G its induced DMG (with nodes* $V \dot{\cup} J$*). Let* $S \subseteq V$ *be a strongly connected component of G and* $D \subseteq V$ *be any union of consolidated districts in G with* $S \subseteq D$ *(e.g.* $D = \mathrm{Cd}^G(S)$*) and* $P := \mathrm{Pa}^G(D) \setminus D$*. Then we have the equality (indices for emphasis):*

$$
\begin{aligned}
&\mathbb{P}_M(S|\mathrm{Pred}^G_<(S) \cap V, \mathrm{do}(J)) \\
&= \mathbb{P}_{M_{[D]}}(S|\mathrm{Pred}^{G_{[D]}}_<(S) \cap D, \mathrm{do}(P)).
\end{aligned}
$$

Proof. First note that since D is a union of strongly connected components and all other variables in $G_{[D]}$ have no parents the total order $<$ is also an apt-order for $G_{[D]}$. It follows that we have the equality of sets of nodes:

$$
\mathrm{Pred}^{G_{[D]}}_<(S) \cap D = \mathrm{Pred}^G_<(S) \cap D \qquad =: D_<.
$$

Now we introduce the following further abbreviations:

$$
\begin{aligned}
D_> &:= D \setminus (S \cup D_<), \\
P_< &:= \mathrm{Pred}^G_<(S) \cap (P \cap V), \\
P_> &:= (P \cap V) \setminus \mathrm{Pred}^G_<(S), \\
P_J &:= P \cap J, \\
J_< &:= \mathrm{Pred}^G_<(S) \cap J, \\
J_> &:= J \setminus \mathrm{Pred}^G_<(S), \\
R_< &:= \mathrm{Pred}^G_<(S) \cap V \setminus (D \cup P), \\
R_> &:= V \setminus (D \cup P \cup \mathrm{Pred}^G_<(S)).
\end{aligned}
$$

Then we get the relations between the sets of nodes:

$$\begin{aligned}
V &= R_< \,\dot\cup\, D \,\dot\cup\, R_> \,\dot\cup\, P_< \,\dot\cup\, P_> \\
D &= D_< \,\dot\cup\, S \,\dot\cup\, D_>, \\
P &= P_< \,\dot\cup\, P_> \,\dot\cup\, P_J, \\
\mathrm{Pred}^G_<(S) \cap V &= D_< \,\dot\cup\, R_< \,\dot\cup\, P_<, \\
J &= J_< \,\dot\cup\, J_>.
\end{aligned}$$

Since $\mathrm{Pred}^G_\leq(S)$ is ancestral in G and $\mathrm{Pred}^{G_{[D]}}_\leq(S)$ is ancestral in $G_{[D]}$, resp., we can by remark 9.7 arbitrarily intervene on all variables outside of these sets without changing the distributions $\mathbb{P}_M(S|\mathrm{Pred}^G_<(S) \cap V, \mathrm{do}(J))$ and $\mathbb{P}_{M_{[D]}}(S|\mathrm{Pred}^{G_{[D]}}_<(S) \cap D, \mathrm{do}(P))$, resp.. With these remarks and our new notations we have the equalities:

$$\begin{aligned}
&\mathbb{P}_M(S|\mathrm{Pred}^G_<(S) \cap V, \mathrm{do}(J)) \\
&= \mathbb{P}_M(S|D_<, R_<, P_<, \mathrm{do}(J)) \\
&\overset{9.7}{=} \mathbb{P}_M(S|D_<, R_<, P_<, \mathrm{do}(J, R_>, P_>, D_>));
\end{aligned}$$

and:

$$\begin{aligned}
&\mathbb{P}_{M_{[D]}}(S|\mathrm{Pred}^{G_{[D]}}_<(S) \cap D, \mathrm{do}(P)) \\
&= \mathbb{P}_{M_{[D]}}(S|D_<, \mathrm{do}(P_<, P_>, P_J)) \\
&\overset{9.7}{=} \mathbb{P}_{M_{[D]}}(S|D_<, \mathrm{do}(P_<, P_>, P_J, D_>)) \\
&\overset{9.7}{=} \mathbb{P}_M(S|D_<, \mathrm{do}(P_<, P_>, J, D_>, R_<, R_>)).
\end{aligned}$$

So the equality between those expressions and thus the claim follows by the 2nd rule of causal calculus in Theorem 7.2 with the σ-separation statement:

$$S \underset{G}{\overset{\sigma}{\perp\!\!\!\perp}} I_{R_<, P_<} \mid D_<, R_<, P_<, \mathrm{do}(J, R_>, P_>, D_>).$$

To prove the latter note that the intervention $\mathrm{do}(R_>, P_>, D_>)$ allows us to restrict to the ancestral subgraph $\mathrm{Pred}^G_\leq(S) \cup J$. Now let π be a path from an indicator variable from $I_{R_<, P_<}$ to S (in $\mathrm{Pred}^G_\leq(S) \cup J$). Then the path can only be of the form:

$$v_i \cdots v_p \longrightarrow v_d \cdots v_s,$$

with $v_i \in I_{R_<, P_<}$, $v_p \in P_<$, $v_d \in D$, $v_s \in S$, as there cannot be any bidirected edge or directed edge in the other direction between $R_< \cup P_<$ and D by the definition of consolidated districts and $P = \mathrm{Pa}^G(D) \setminus D$. Since we condition on $P_<$ the path π is σ-blocked. $\square$