Recovering from Selection Bias in Causal and Statistical Inference (Supplemental Material)

Elias Bareinboim **Jin Tian** **Judea Pearl**

Appendix 1 (background)

The basic semantical framework in our analysis rests on structural causal models (Pearl 2000, Ch. 7). In this framework, each child-parent relationship is a structural function, and a collection of these functions induces observational and interventional distributions. We make use of a criterion for reading constraints imposed over these distributions in the induced graph known as *d-separation*, as defined next.

Definition (d-separation (Pearl 1988))**.** *A set Z of nodes is said to block a path p in a causal graph G if either*

1. *p contains at least one arrow-emitting node in Z, or*
2. *p contains at least one collision node that is outside Z and has no descendant in Z.*

If Z blocks all *paths from set X to set Y, it is said to "d-separate X and Y," and then, it can be shown that variables X and Y are independent given Z, written $(X \perp\!\!\!\perp Y|Z)$.*

We also make use of the do-calculus, which is a collection of syntactic rules that permit the manipulation of causal expressions involving the do-operator (Pearl 1995). Let X, Y, and Z be arbitrary disjoint sets of nodes in a causal graph G. An expression of the type $Q = P(y|do(x), z)$ is said to be compatible with G if the interventional distribution described by Q can be generated by parameterizing the graph with a set of functions and exogenous variables.

The following rules are valid for every interventional distribution compatible with G (Pearl 2000, pp. 85–86):

Rule 1 (Insertion/deletion of observations):

$$P(y|do(x), z, w) = P(y|do(x), w) \quad \text{if } (Y \perp\!\!\!\perp Z|X, W)_{G_{\overline{X}}}$$

Rule 2 (Action/observation exchange):

$$P(y|do(x), do(z), w) = P(y|do(x), z, w) \quad \text{if } (Y \perp\!\!\!\perp Z|X, W)_{G_{\overline{X}\underline{Z}}}$$

Rule 3 (Insertion/deletion of actions):

$$P(y|do(x), do(z), w) = P(y|do(x), w) \quad \text{if } (Y \perp\!\!\!\perp Z|X, W)_{G_{\overline{X}\,\overline{Z(W)}}},$$

where $Z(W)$ is the set of Z-nodes that are not ancestors of any W-node in $G_{\overline{\mathbf{X}}}$.

Appendix 2 (proofs)

Theorem 1. *The distribution $P(y|x)$ is s-recoverable from G_s if and only if $(S \perp\!\!\!\perp Y|X)$.*

Proof. (if) It is obvious that if X d-separates S from Y in G_s, $P(y|x)$ is s-recoverable.

(only if) We show that whenever there exists an open path between S and Y that is not blocked by X, two distributions P_1, P_2 compatible with the causal model can be constructed such that they agree in the probability distribution under selection bias, $P_1(V \mid S = 1) = P_2(V \mid S = 1)$, and disagree in the target distribution $Q = P(Y \mid X)$, i.e., $P_1(Y \mid X) \neq P_2(Y \mid X)$.

Let P_1 be compatible with the graph $G_1 = G_s$, and P_2 with the subgraph G_2 where the edges pointing to S are removed (see (Tian 2002, Lemma 8)). Notice that P_2 harbors an additional independence relative $(V \perp\!\!\!\perp S)_{P_2}$, where V represents all variables in G_s but the selection mechanism S. We will set the parameters of P_1 through its factors and then compute the parameters of P_2 by enforcing $P_2(V \mid S = 1) = P_1(V \mid S = 1)$. Since $P_2(V|S = 1) = P_2(V)$, we will have $P_1(V|S = 1) = P_2(V)$.

Given a Markovian data-generating model (Pearl 2000), P_1 can be parametrized through its factors in the Markovian decomposition $P_1(S = 1 \mid Pa_s), P_1(X \mid Pa_x), \ldots$, more generally, $P_1(V_i \mid PA_i)$ for each family in the graph. Recoverability should hold for any parametrization, so we assume that all variables are binary. In turn, we examine the possible ways of how S is connected to Y while conditioned on X.

Case 1. Firstly, let us consider the case in which $Y \in Pa_S$, which implies that S is not separable from Y in G_s. We follow the construction given in Lemma 1. Let U be the set of nodes that connect X to Y. The distribution of Y is a function of the values of X if we sum out all variables in U, $P_1(Y|X) = \sum_U \prod_{X,U,Y} P_1(V_i|Pa_i))$, so without loss of generality we can parametrize this distribution directly. Now, we can write the conditional distribution in the second causal model as follows:

$$P_2(Y|X) = P_1(Y|X, S = 1) = \frac{P_1(Y, X, S = 1)}{P_1(X, S = 1)} \quad (1)$$

$$= \frac{P_1(S = 1|Y)P_1(Y|X)}{P_1(S = 1|Y)P_1(Y|X) + P_1(S = 1|\overline{Y})P_1(\overline{Y}|X)}, \quad (2)$$

where the first equality is enforced by construction, the second and third follow from the axioms of probability.

Consider the subgraph G' such that all $V \setminus \{X, Y, U, S\}$ are disconnected from $\{X, Y, U, S\}$, where we can parametrize the complete model as in (Tian 2002, Lemma 8). Now we compare $P_2(Y|X)$ with $P_1(Y|X)$. The equality constraint imposed over these quantities can be seen as a line in the parameter space of higher dimension, which has measure zero. This implies that for almost all parametrizations, $P_1(Y|X)$ and eq. (2) will not be the same. For instance, we can set the distribution of every family in G' but the selection node equal to $1/2$, and set the distribution $P_1(S = 1|Y) = \alpha, P_1(S = 1|\overline{Y}) = \beta$, for $0 < \alpha, \beta < 1$ and $\alpha \neq \beta$. The result follows since the other parameters of P_2 are free and can be chosen to match P_1, and $P_2(Y|X) = \alpha/(\alpha + \beta)$ and $P_1(Y|X) = 1/2$.

Case 2. Let us consider the case in which there exists an open directed path from Y to S, which means that it does not pass through X (i.e., only the values of X will end up being used in the construction). Let Z be the immediate child of Y in this path and assume the distance from Z to S is arbitrary. Let W be the set of nodes that connect Z to S and U be the set of nodes that connect X to Y. Consider the induced subgraph G' such that all nodes in G_s but $V \setminus \{X, U, Y, Z, W, S\}$ are disconnected from $\{X, U, Y, Z, W, S\}$.

Following eq. (1), $P_2(Y|X) = \frac{P_1(Y,X,S=1)}{P_1(X,S=1)}$, we can rewrite the numerator of the r.h.s. in expanded form as

$$\begin{aligned} P_1(Y, X, S = 1) &= \sum_{U,Z,W} P_1(X, U, Y, W, Z, S = 1) \\ &= \sum_{U,Z,W} P_1(X|Pa_x) \dots P_1(S = 1 \mid Pa_s) \\ &= \sum_{U,Z,W} \prod_{V \cup S} P_1(V_i \mid Pa_i) \\ &= \sum_{U} \prod_{U} P_1(V_i \mid Pa_i) \sum_{Z,W} \prod_{V \cup S \setminus U} P_1(V_i \mid Pa_i) \end{aligned} \quad (3)$$

Given a topological order compatible with G', the families in U are functions of X but not of Z, W, Y, S, and since the same value of X is instantiated in the numerator and the denominator in eq. (1), these factors cancel out. So, we consider only the second sum in eq. (3). Now, we can rewrite

$$\begin{aligned} &\sum_{Z,W} \prod_{V \cup S \setminus U} P_1(V_i \mid Pa_i) \\ &= \sum_{Z} \prod_{V \setminus U \cup W} P_1(V_i \mid Pa_i) \sum_{W} \prod_{W \cup S} P_1(V_i \mid Pa_i) \end{aligned} \quad (4)$$

The sum over the factors relative to W in eq. (4) is a function of Z (since $Z \in An(S)$), so define $f(Z) = \sum_W \prod_{S \cup W} P_1(V_i|Pa_i)$. The distribution of Y is a function of the value of X since we sum out all values of U, let us call it $P(Y|\tilde{X})$. Define $\alpha_z(Y) = P(Z|Y)$, and since Y is not affected by Z, we can rewrite eq. (4) as $P(Y|\tilde{X}) \sum_Z \alpha_z(Y) f(Z)$. Given these observations, we rewrite $P_2(Y|X)$ (eq. (1)) as follows

$$\frac{P_1(Y|\tilde{X}) \sum_Z \alpha_z(Y) f(Z)}{\big(P_1(Y|\tilde{X}) \sum_Z \alpha_z(Y) f(Z)\big) + \big(P_1(\overline{Y}|\tilde{X}) \sum_Z \alpha_z(\overline{Y}) f(Z)\big)},$$

which we want to compare with $P_1(Y|\tilde{X})$.

By construction of G', $f(Z)$ and $\alpha_z(Y)$ as a convolution, it is the case that the expressions for Q_1 and Q_2 cannot be simplified in the general case. We explore the fact that the equality constraint between these two quantities (for all values of X and Y) imposes weak constraints in the high dimensional parameter space and valid parametrizations have Lebesgue mass zero; i.e., for almost all parameters that we chose the equality between Q_1 and Q_2 will not hold, we chose explicitly one of such parameters. So, first make $P_1(Y|\tilde{X}) = 1/2$ for all values of $Y, \tilde{X}$, which implies

$$P_2(Y|\tilde{X}) = \frac{\sum_Z \alpha_z(Y) f(Z)}{\big(\sum_Z \alpha_z(Y) f(Z)\big) + \big(\sum_Z \alpha_z(\overline{Y}) f(Z)\big)} \quad (5)$$

We can compose the linear transformations encoded in $f(Z)$, which is from the parameter space of $W \cup S$ to Z, that is, $[0, 1]^{2^{|W|+1}} \to [0, 1]^{|Z|}$. Consider a topological order $W_1 < W_2 < ... < W_{|W|} < S$ relative to $W \cup S$. We rearrange the product $\sum_W \prod_{S \cup W} P_1(V_i|Pa_i)$ as 2×2 matrices relative to each factor $P(W_i|W_{i+1})$ (each row sums to 1 satisfying the integrality constraint) and $P(S = 1|W_{|W|})$ is a column-vector 2×1 for each value of $W_{|W|}$.

Let the matrix of the first distribution relative to W_1 be $M = [p, 1-p; 1-q, q]$, for some $0 < p, q < 1$, which will be instantiated below. We can decompose M in its canonical form, i.e., in terms of its eigenvectors, $[1, -(p-1)/(q-1)], [1, 1]$, and eigenvalues $[1, p+q-1]$. The product in $f(Z)$ is a composition of linear transformations, which is also a linear transformation. We make each distribution to follow the same form given by M, so this composition is equivalent to the product of the matrix with the eigenvectors times the power to $k = |W|$ of the matrix with the eigenvalues in the diagonal times the inverse of the matrix with the eigenvectors, let us call it M_c. After some trivial (but tedious) algebra, we obtain:

$$\begin{aligned} M_c(1,1) &= 1 - \frac{(1-p)\big((p+q-1)^k - 1\big)}{p+q-2} \\ M_c(1,2) &= \frac{(1-p)\big((p+q-1)^k - 1\big)}{p+q-2} \\ M_c(2,1) &= \frac{(1-q)\big((p+q-1)^k - 1\big)}{p+q-2} \\ M_c(2,2) &= 1 - \frac{(1-q)\big((p+q-1)^k - 1\big)}{p+q-2} \end{aligned} \quad (6)$$

Set $(p = 3/5, 1 - q = 2/5)$, it is not difficult to check that this assignment yields a valid parametrization for the distribution, we have

$$\begin{aligned} M_c(1,1) = M_c(2,2) &= 1 - \frac{1}{2}(1 - (\frac{1}{5})^k) \\ M_c(1,2) = M_c(2,1) &= \frac{1}{2}(1 - (\frac{1}{5})^k) \end{aligned} \quad (7)$$

Now, let $(P(S = 1|W_{|W|}) = 2/3, P(S = 1|\overline{W_{|W|}}) = 1/2)$, and we can see that $f(Z) = 7/12 + \epsilon, f(\overline{Z}) = 7/12 - \epsilon$, where $\epsilon = (1/5)^k$. We can chose $\alpha_z(y) = 1/3, \alpha_z(\overline{y}) = 3/4$. Finally, we can evaluate eq. (5) and note that $Q_2 = 1/2 - (2/7)\epsilon$, which is never equal to $1/2$ (= Q_1) given that the graph is finite.

Case 3. Let us consider the case in which the path from Y to S pass through an ancestor of Y. Let us call $A = An(Y) \setminus \{Y\}$. Since $A \setminus X$ is not d-separated from Y given X in G_s, there is a path p from $Z \in A \setminus X$ to Y that is not blocked by X. Without loss of generality, let us consider the closest Z in this path. There are two possible cases to consider: p might be a directed path from Z to Y that does not contain X as an intermediate (e.g., $Z \to \ldots \to Y$); or, p might contain converging arrows into X ($Z \to \ldots \to X \leftarrow \ldots \to Y$).

Subcase 3a. We start with when p is a directed path. Let U be the set of nodes that connect X to Y, W the nodes that connect Z to Y (given X), and R the nodes that connect Z to S. Consider the induced subgraph G' of G_s such that all nodes except $\{X, U, Z, W, R, Y\}$ are removed from G_s (i.e., $V \setminus \{X, U, Z, W, R, Y\}$ can be parametrized as random coins, see (Tian 2002, Lemma 8)). Since $Z \in An(S)$, let us call p' the path connecting Z to $An(S) \setminus An(Y)$ in G_s (i.e., $Z \to \ldots \to S$). Add p' with all its nodes to G'. Note that Z is such that it blocks the concatenation of p and p'. Note that this concatenation is such that it has two emanating arrows from Z (i.e., $p \leftarrow Z \to p'$). Now, we can transform G_s while staying in the same equivalence class . In order to do so, reverse the direction of all arrows in p such that Z is no longer in $An(Y)$. Now, the same parametrization as discussed in case 2 is valid for this case.

Subcase 3b. Consider the case in which p contain converging arrows into X. Let us consider the variables X, Y, Z, and let L be the common ancestor that, together with Z, has converging arrows into X in p. The construction here will be similar to the previous case except for two main differences.

First, the path p can be seen as the concatenation of four segments $p_1, ..., p_4$ such that p_1 is the segment $L \to \ldots \to Y$, p_2 is the segment $L \to \ldots \to X$, p_3 is the segment $Z \to \ldots \to X$, and p_4 the segment $Z \to \ldots \to S$. Note that by construction, there might exist only chains along each of these segments, so to avoid algebraic clutter we assume that those are segments of length one, but it is trivial to stretch those segments following the same structure given in case 2 for $f(Z)$. When we have multiple X's in p, we will have the concatenation of several segments p_3 and p_4, and it will also be simple to extend the construction given for $f(Z)$ for this case. Remarkably, these segments capture precisely the forbidden subgraph that precludes s-recoverability when p has converging arrows to X. Second, no directed path between X and Y is used in the construction of the counterexample and the induced subgraph G' without these paths can also be generated by the original model (Tian 2002, Lemma 8).

We follow similar structure as in case 2. Following eq. (1), $P_2(Y|X) = \frac{P_1(Y,X,S=1}{P_1(X,S=1)}$, we can rewrite the numerator as

$$\sum_L P_1(Y|L)P_1(L) \sum_Z P_1(X|Z,L)P_1(Z)P_1(S=1|Z) \quad (8)$$

Define $\alpha_L(Y) = P_1(Y|L)P_1(L)$ and note that the second sum is not affected by Y but it is a function of L, so define $f(L) = \sum_Z P_1(X|Z,L)P_1(Z)P_1(S=1|Z)$, and write

$$P_2(Y|X) = \frac{\sum_L \alpha_L(Y) f(L)}{\left(\sum_L \alpha_L(Y) f(L)\right) + \left(\sum_L \alpha_L(\overline{Y}) f(L)\right)} \quad (9)$$

Define another function of L that sums out S, $g(L) = \sum_Z P_1(X|Z,L)P_1(Z)$, and note that $P_1(Y|X)$ is the same as eq. (9) with the function f replaced with g. This expression cannot be simplified in general since there is a dependence across the two functions. To see that, consider the following parametrization: $\alpha_L(Y) = \alpha_{\overline{L}}(Y) = 1/3, \alpha_L(\overline{Y}) = 1/9, \alpha_{\overline{L}}(\overline{Y}) = 2/9$, $P(Z) = 1/2$; $P_1(X|Z,L) = 1/2 + \epsilon, P_1(X|\overline{Z},L) = 1/2 - \epsilon$, $P_1(X|Z,\overline{L}) = P_1(X|\overline{Z},\overline{L}) = 1/2$, for $0 < \epsilon < 1/2$. Call $P(S = 1|Z) = \alpha, P(S = 1|\overline{Z}) = \beta$, and pick any α, β such that $\alpha > \beta$. After some trivial (but tedious) algebra, we have $P_1(Y|X) = 2/3$ and

$$P_2(Y|X) = \frac{2}{3}\left(\frac{\alpha + \beta + \epsilon(\alpha - \beta)}{\alpha + \beta + \frac{8}{9}\epsilon(\alpha - \beta)}\right), \quad (10)$$

which are always different. QED.

□

Remark. We considered Markovian models in Theorem 1, but the extension for Semi-Markovians is straightforward. This is so because the latent variables impose no constraints over the distribution of the observables, which means that there are even more degrees of freedom that can be used to produce a parametrization following the lack of separability.

Theorem 2. *If there is a set* $\mathbf{C}$ *that is measured in the biased study with* $\{\mathbf{X}, Y\}$ *and in the population level with* $\mathbf{X}$ *such that* $(Y \perp\!\!\!\perp S|\{\mathbf{C}, \mathbf{X}\})$*, then* $P(y|\mathbf{x})$ *is recoverable as*

$$P(y|x) = \sum_c P(y|x, c, S = 1)P(c|x). \quad (11)$$

Proof. We can condition $P(y|x)$ on the set C and write

$$P(y|x) = \sum_c P(y|x, c)P(c|x) \quad (12)$$

$$= \sum_c P(y|x, c, S = 1)P(c|x), \quad (13)$$

where the last line follows since C is such that $(Y \perp\!\!\!\perp S|\{\mathbf{C}, \mathbf{X}\})$. QED. □

Lemma 2. *If* $Y \perp\!\!\!\perp S|(C, X)$*, then* $Y \perp\!\!\!\perp S|(C', X)$*, where* $C' = C \cap An(Y \cup S \cup X)$ *(Acid and de Campos 1996).*

Lemma 3. *Given three sets of nodes* X*,* Y*, and* Z*, and a set* $C \subseteq An(X \cup Y \cup Z)$*,* $X \perp\!\!\!\perp Y|(Z \cup C)$ *if and only if* $Z \cup C$ *separates* X *from* Y *in undirected graph* $(G_{An(X \cup Y \cup Z)})^m$*, the moral graph of* $G_{An(X \cup Y \cup Z)}$ *(Acid and de Campos 1996).*

Theorem 3. *There exists some set* $C \subseteq T \cap M$ *such that* $Y \perp\!\!\!\perp S|\{C, X\}$ *if and only if the set* $(C' \cup X)$ *d-separates* S *from* Y *where* $C' = [(T \cap M) \cap An(Y \cup S \cup X)] \setminus (Y \cup S \cup X)$*.*

Proof. The "if" part is trivial as it gives a set that d-separates S from Y.

(only if) If there exists a set $C \subseteq T \cap M$ (that is disjoint from Y, S, X) such that $Y \perp\!\!\!\perp S|(C, X)$ then the set $C'' = C \cap An(Y \cup S \cup X)$ satisfies $Y \perp\!\!\!\perp S|(C'', X)$ based on Lemma 2. From Lemma 3 we have that $C'' \cup X$ separates S from Y in the undirected graph $(G_{An(Y \cup S \cup X)})^m$. In an undirected graph, if $(C'' \cup X) \subseteq (C' \cup X)$, is a separator, then $C' \cup X$ must be a separator. Using Lemma 3 again, we obtain that $(C' \cup X)$ d-separates S from Y in G. QED. □

Lemma 4. *Let C_1 be a minimal set satisfying $Y \perp\!\!\!\perp S|(C_1, X)$, C_1^o be any subset of C_1 (including empty set), and $C_1^m = C_1 \setminus C_1^o$. If C_2 is a minimal set satisfying $C_1^m \perp\!\!\!\perp S|(C_1^o, X, C_2)$, then we must have $Y \perp\!\!\!\perp S|(C_2, C_1, X)$ and $Y \perp\!\!\!\perp S|(C_2, C_1^o, X)$.*

Proof. Since C_1 is minimal, by Lemma 2 we obtain $C_1 \subseteq An(Y \cup S \cup X)$. Similarly we have $C_2 \subseteq An(S \cup X \cup C_1)$, and therefore $C_2 \subseteq An(Y \cup S \cup X)$. Since $Y \perp\!\!\!\perp S|(C_1, X)$, by Lemma 3 we have that $C_1 \cup X$ separates S from Y in the undirected graph $(G_{An(Y \cup S \cup X)})^m$. Since $C_2 \subseteq An(Y \cup S \cup X)$ we have that $C_1 \cup X \cup C_2$ separates S from Y in the undirected graph $(G_{An(Y \cup S \cup X)})^m$. Then by Lemma 3 we obtain $Y \perp\!\!\!\perp S|(C_2, C_1, X)$. Given $Y \perp\!\!\!\perp S|(C_2, C_1^m, C_1^o, X)$, and $C_1^m \perp\!\!\!\perp S|(C_1^o, X, C_2)$, we obtain $Y \perp\!\!\!\perp S|(C_2, C_1^o, X)$ by the contraction axiom. □

Lemma 5. *For sets W, X, let C_1 be a nonempty minimal set satisfying $W \perp\!\!\!\perp S|(C_1, X)$. Let C_1^o be any subset of C_1, and $C_1^m = C_1 \setminus C_1^o$. We have*

$$P(w|x) = \sum_{c_1} P(w|x, c_1, S = 1)P(c_1|x). \tag{14}$$

Then

1. *$C_1 \perp\!\!\!\perp S|X$ does not hold.*
2. *Let $C_2 \subseteq M$ be a minimal set satisfying $C_1 \perp\!\!\!\perp S|(X, C_2)$. Then $W \perp\!\!\!\perp S|(C_2, X)$. Therefore,*

$$P(c_1|x) = \sum_{c_2} P(c_1|x, c_2, S = 1)P(c_2|x). \tag{15}$$

$$P(w|x) = \sum_{c_2} P(w|x, c_2, S = 1)P(c_2|x). \tag{16}$$

That is, if $P(c_1|x)$ is recovered via Theorem 2, then $P(w|x)$ must be recovered via Theorem 2.

3. *$C_1^m \perp\!\!\!\perp S|(C_1^o, X)$ does not hold.*
4. *Let $C_2 \subseteq M$ be a minimal set satisfying $C_1^m \perp\!\!\!\perp S|(C_1^o, X, C_2)$. Then $W \perp\!\!\!\perp S|(C_2, C_1^o, X)$. Therefore,*

$$P(c_1^m|c_1^o, x) = \sum_{c_2} P(c_1^m|c_1^o, x, c_2, S = 1)P(c_2|c_1^o, x). \tag{17}$$

$$P(w|x) = \sum_{c_1^o, c_2} P(w|c_1^o, x, c_2, S = 1)P(c2, c_1^o|x). \tag{18}$$

That is, if $P(c_1^m|c_1^o, x)$ is recovered via Theorem 2, then $P(w|x)$ must be recovered via Theorem 2.

Proof. 1. If $C_1 \perp\!\!\!\perp S|X$, from $W \perp\!\!\!\perp S|(C_1, X)$ and the contraction graphoid axiom, we obtain $W \perp\!\!\!\perp S|X$. This contradicts with C_1 being minimal.

2. Given $C_2 \subseteq M$ being a minimal set satisfying $C_1 \perp\!\!\!\perp S|(X, C_2)$, we obtain $W \perp\!\!\!\perp S|(C_2, X)$ by Lemma 4.
3. If $C_1^m \perp\!\!\!\perp S|(C_1^o, X)$, from $W \perp\!\!\!\perp S|(C_1^m, C_1^o, X)$ and the contraction graphoid axiom, we obtain $S \perp\!\!\!\perp W|(C_1^o, X)$. This contradicts with C_1 being minimal.
4. Given $C_2 \subseteq M$ being a minimal set satisfying $C_1^m \perp\!\!\!\perp S|(C_1^o, X, C_2)$, we obtain $W \perp\!\!\!\perp S|(C_2, C_1^o, X)$ by Lemma 4.

□

Definition 3. *We say that $P(w|z)$ is C-recoverable if and only if it is recovered by the procedure $RC(w, z)$.*

Theorem 4. *For $X \subseteq T$, $Y \notin T$, $Q = P(y|x)$ is C-recoverable if and only if it is recoverable by Theorem 2, that is, if and only if there exists a set $C \subseteq T \cap M$ such that $(Y \perp\!\!\!\perp S|C, X\})$ (where C could be empty). If s-recoverable, $P(y|x)$ is given by $P(y|x) = \sum_c P(y|x, c, S = 1)P(c|x)$.*

Proof. (if) If there exists a set $C \subseteq T \cap M$ such that $Y \perp\!\!\!\perp S|(C, X)$, then it is clear $RC(Y, X)$ will recover $P(y|x)$.

(only if) Assume there exists no set $C \subseteq T \cap M$ such that $Y \perp\!\!\!\perp S|(C, X)$. If there exists no set $C \subseteq M$ such that $Y \perp\!\!\!\perp S|(C, X)$, then $RC(Y, X)$ will output FAIL. Assume for every minimal set $C_1 \subseteq M$ satisfying $Y \perp\!\!\!\perp S|(C_1, X)$, there exist some variables in C_1 that are not in T. We need to prove $RC(Y, X)$ will not recover $P(y|x)$.

The only way for $RC(Y, X)$ to recover $P(y|x)$ is by the following

$$P(y|x) = \sum_{c_1} P(y|x, c_1, S = 1)P(c_1|x), \tag{19}$$

such that $R(C_1, X)$ recovers $P(c_1|x)$ for some C_1. By Lemma 5, the only way for $R(C_1, X)$ to recover $P(c_1|x)$ is that there exists some $C_1^o \subset C_1$ (C_1^o could be empty set) for which there exists a minimal set $C_2 \subseteq M$ satisfying $C_1^m \perp\!\!\!\perp S|(C_1^o, X, C_2)$ where $C_1^m = C_1 \setminus C_1^o$, such that either $C_2 \cup C_1^o \cup X \subseteq T$ rendering $P(c_1^m|c_1^o, x)$ being recovered via Theorem 2 or $R(C_2, C_1^o \cup X)$ recovers $P(c_2|c_1^o, x)$ (and $R(C_1^o, X)$ recovers $P(c_1^o|x)$). But $P(c_1^m|c_1^o, x)$ being recovered via Theorem 2 would contradict with our assumption since by Lemma 5 it means $P(y|x)$ will be recovered via Theorem 2.

These same arguments apply to $R(C_2, C_1^o \cup X)$. By repeated application of Lemma 5, we have that if $RC(Y, X)$ succeeds in recovering $P(y|x)$, then there exist a sequence of function calls $R(C_1, X), R(C_2, C_1^o \cup X), \ldots, R(C_k, C_{k-1}^o \cup \ldots \cup C_1^o \cup X)$ that ends with $R(C_k, C_{k-1}^o \cup \ldots \cup C_1^o \cup X)$ succeeding in computing $R(C_k|C_{k-1}^o \cup \ldots \cup C_1^o \cup X)$ by recovering $R(C_k^m|C_k^o, C_{k-1}^o \cup \ldots \cup C_1^o \cup X)$ via Theorem 2. Then by reasoning backwards using Lemma 5, we have that $R(C_{k-1}^m|C_{k-1}^o, C_{k-2}^o \cup \ldots \cup C_1^o \cup X)$ must be recovered via Theorem 2, and so on, until we obtain $P(c_1^m|c_1^o, x)$ must be recovered via Theorem 2 and finally $P(y|x)$ must be recovered via Theorem 2. This would contradict with our assumption. Therefore $RC(Y, X)$ will not recover $P(y|x)$. QED. □

Definition 4 (Selection-backdoor criterion)**.** *Let a set Z of variables be partitioned into $Z^+ \cup Z^-$ such that Z^+ contains all non-descendants of X and Z^- the descendants of X. Z is said to satisfy the selection backdoor criterion (s-backdoor, for short) relative to an ordered pairs of variables (X, Y) and an ordered pair of sets (M, T) in a graph G_s if Z^+ and Z^- satisfy the following conditions:*

(i) Z^+ blocks all back door paths from X to Y;

(ii) X and Z^+ block all paths between Z^- and Y, namely, $(Z^- \perp\!\!\!\perp Y|X, Z^+)$;

(iii) X and Z block all paths between S and Y, namely, $(Y \perp\!\!\!\perp S|X, Z)$;

(iv) $Z \cup \{X, Y\} \subseteq M$, and $Z \subseteq T$.

Theorem 5 (Selection-backdoor adjustment)**.** *If a set Z satisfies the s-backdoor criterion relative to the pairs (X, Y) and (M, T) (as in def. 2), then the causal effect of X on Y is identifiable and recoverable and is given by the formula*

$$P(y|do(x)) = \sum_{z^+} P(y|x, z, S = 1)P(z) \tag{20}$$

Proof. We first condition the effect of X on Y on Z^+ and write

$$P(y|do(x)) = \sum_{z^+} P(y|do(x), z^+)P(z^+|do(x)) \tag{21}$$

We can rewrite the effect in eq. (21) as

$$P(y|do(x)) = \sum_{z^+} P(y|do(x), z^+)P(z^+) \tag{22}$$

$$= \sum_{z^+} P(y|x, z^+)P(z^+), \tag{23}$$

where eq. (22) follows by the third rule of the do-calculus together with the fact that $(Z^+ \perp\!\!\!\perp X)_{G_{\overline{X}}}$ (since by construction Z^+ contains only non-descendants of Y), and eq. (23) follows by the second rule of the do-calculus together with condition (i).

We can rewrite the second term in eq. (23) summing over Z^- and pull the sum out, which yield

$$P(y|do(x)) = \sum_{z^+, z^-} P(y|x, z^+)P(z^+, z^-) \tag{24}$$

By the contraction graphoid axiom conditions (ii) and (iii) entail $(Y \perp\!\!\!\perp S, Z^-|X, Z^+)$, so we can add $\{Z^-, S\}$ to the first term of eq. (24) and obtain

$$P(y|do(x)) = \sum_{z^+, z^-} P(y|x, z^+, z^-, S = 1)P(z^+, z^-). \tag{25}$$

Note that eq. (25) is identifiable and its recoverability is given by condition 4. QED. □

Appendix 3 (algorithm)

In the sequel, we provide a procedure for listing all recoverable distributions in the form of $P(y, B|A)$. Note that $P(y, B|A)$ being recoverable implies other distributions such as $P(y|A, D)$ is recoverable for all $D \subseteq B$.

Procedure Sink-Recover(G, Y)

1. Remove $V \setminus An(Y \cup S)$ from G.

2. Eliminate S.

 (a) If $Y \in Pa_S$, exit with failure.

 (b) Otherwise, $P(Y, V \setminus Pa_S \setminus \{Y\}|Pa_S)$ is s-recoverable, and remove S from G.

3. Eliminate non-ancestors of Y from the graph one by one. Given $P(Y, B|A)$ s-recoverable, iterate in reverse topological order, for each sink node Z.

 (a) If $Y \notin Pa_Z$, $P(Y, B \setminus Pa_Z|A \cup Pa_Z \setminus Z)$ is s-recoverable, and remove Z from G.

 (b) Otherwise, exit if no non-ancestors of Y can be removed.

4. Now all non-ancestors of Y have been removed and we have $P(Y, B|A)$ s-recoverable.

 (a) For $C \subseteq An(Y) \setminus \{Y\}$, if $(Y \perp\!\!\!\perp A - C|C)$, then $P(Y|C)$ s-recoverable.

The procedure operates traversing the graph and trying to recover distributions in the form $P(y, B|A)$ until the current node can no longer be separated from Y given its parents (and respective ancestors), or it ends listing all distributions and reaching Y itself. It is not difficult to see that whenever the algorithm exits with failure, one of the separability conditions discussed in the proof of Theorem 1 is violated, which means that a counterexample for s-recoverability can be produced.

Interestingly, the Sink-Recover() can be easily modified to list odds ratios (OR), extending the query-specific treatment given in (Bareinboim and Pearl 2012). Note that the symmetry of the functional form of the OR can be exploited in this case so that the separability test in the procedure can be relaxed. Under this relaxation the current Z must be separated from X *or* Y rather than always Y.

References

Acid, S., and de Campos, L. 1996. An algorithm for finding minimum d-separating sets in belief networks. In *Proceedings of the 12th Annual Conference on Uncertainty in Artificial Intelligence*, 3–10. San Francisco, CA: Morgan Kaufmann.

Bareinboim, E., and Pearl, J. 2012. Controlling selection bias in causal inference. In Girolami, M., and Lawrence, N., eds., *Proceedings of The Fifteenth International Conference on Artificial Intelligence and Statistics (AISTATS 2012)*, 100–108. JMLR (22).

Pearl, J. 1988. *Probabilistic Reasoning in Intelligent Systems*. San Mateo, CA: Morgan Kaufmann.

Pearl, J. 1995. Causal diagrams for empirical research. *Biometrika* 82(4):669–710.

Pearl, J. 2000. *Causality: Models, Reasoning, and Inference*. New York: Cambridge University Press. Second ed., 2009.

Tian, J. 2002. *Studies in Causal Reasoning and Learning*. Ph.D. Dissertation, Computer Science Department, University of California, Los Angeles, CA.