Recovering from Selection Bias in Causal and Statistical Inference

Elias Bareinboim[1] Jin Tian[2] Judea Pearl[3]

[1,3] Cognitive System laboratory Computer Science Department, University of California, Los Angeles
[2] Department of Computer Science Iowa State University

Presented by Shaolin Xiang and Ziqi Zhang

Table of Contents

Table of Contents

Selection Bias

Definition

Selection bias is induced by preferential selection of units for data analysis, usually governed by unknown factors including treatment, outcome, and their consequences.

Selection Bias

Definition

Selection bias is induced by preferential selection of units for data analysis, usually governed by unknown factors including treatment, outcome, and their consequences.

- When conducting a sample survey, inevitably, we run into selection bias.

Selection Bias

Definition

Selection bias is induced by preferential selection of units for data analysis, usually governed by unknown factors including treatment, outcome, and their consequences.

- When conducting a sample survey, inevitably, we run into selection bias.
- Example: In a study of the effect of training program on earnings. People with higher incomes tends to report their earnings more frequently than those who earn less.

Selection Bias

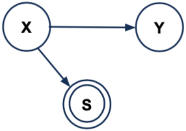
Definition

Selection bias is induced by preferential selection of units for data analysis, usually governed by unknown factors including treatment, outcome, and their consequences.

- When conducting a sample survey, inevitably, we run into selection bias.
- Example: In a study of the effect of training program on earnings. People with higher incomes tends to report their earnings more frequently than those who earn less.
- Consequence: Data gathered from this study will reflect the distortion in the sample proportions and the sample is no longer a faithful representation of the population. Thus, estimates are biased regardless of how many samples were collected.

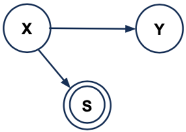
Selection Bias

- Treatment dependent

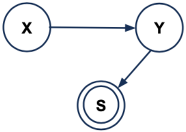


Selection Bias

- Treatment dependent

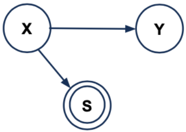


- Outcome dependent

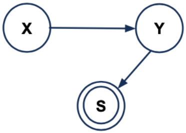


Selection Bias

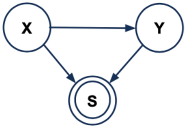
- Treatment dependent



- Outcome dependent

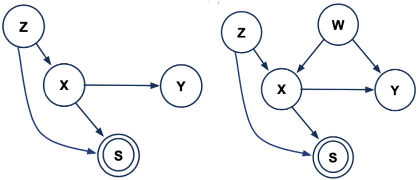


- Both treatment and outcome dependent



Selection Bias

Other examples:



When there is a driver of treatment(denoted by Z), the treatment dependent study where selection bias is also affected by Z.

Questions of Interest

- Problem: We want to compute the population-level conditional distribution $P(y|x)$, but the samples available are collected under selection, that is, only $P(y|x, S = 1)$ is accessible for use. Can we recover $P(y|x)$ from $P(y|x, S = 1)$?

Questions of Interest

- Problem: We want to compute the population-level conditional distribution $P(y|x)$, but the samples available are collected under selection, that is, only $P(y|x, S = 1)$ is accessible for use. Can we recover $P(y|x)$ from $P(y|x, S = 1)$?
- Question:
  - Selection without external data: The dataset is collected under selection bias, $P(\mathbf{v}|S = 1)$; under which condition is $P(y|\mathbf{x})$ recoverable?
  - Selection with external data: The dataset is collected under selection bias, $P(\mathbf{v}|S = 1)$, but there are unbiased samples from $P(\mathbf{t})$, for $\mathbf{T} \subseteq \mathbf{V}$, under which conditions is $P(y|\mathbf{x})$ recoverable?
  - Selection in causal inferences: The data is collected under selection bias, $P(\mathbf{v}|S = 1)$, but there are unbiased samples from $P(\mathbf{t})$, for $\mathbf{T} \subseteq \mathbf{V}$, under which condition is $P(y|do(\mathbf{x}))$ estimable?

Table of Contents

Without External Data

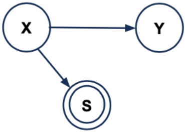
- All data are collected under selection.

Definition: s-Recoverability

Given a causal graph G_s augmented with a node S encoding the selection mechanism, the distribution $Q = P(y|\mathbf{x})$ is said to be s-recoverable from selection biased data in G_s if the assumptions embedded in the causal model renders Q expressible in terms of the distribution under selection bias $P(\mathbf{v}|S = 1)$. Formally, for every two probability distribution P_1 and P_2 compatible with G_s, $P_1(\mathbf{v}|S = 1) = P_2(\mathbf{v}|S = 1) > 0$ implies $P_1(y|\mathbf{x}) = P_2(y|\mathbf{x})$

Example 1

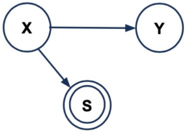
- Consider graph G_s:



Our goal is to establish s-recoverability of $Q = P(y|x)$

Example 1

- Consider graph G_s:

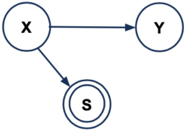


Our goal is to establish s-recoverability of $Q = P(y|x)$

- By d-separation, X separates Y from S, i.e. $Y \perp\!\!\!\perp S|X$

Example 1

- Consider graph G_s:



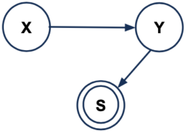
Our goal is to establish s-recoverability of $Q = P(y|x)$

- By d-separation, X separates Y from S, i.e. $Y \perp\!\!\!\perp S|X$
- s-recoverable.

$$P(y|x) = P(y|x, S = 1)$$

Example 2

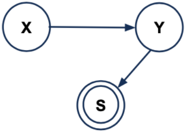
- Consider graph G_s:



Our goal is to establish s-recoverability of $Q = P(y|x)$

Example 2

- Consider graph G_s:

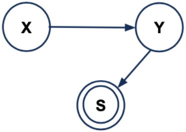


Our goal is to establish s-recoverability of $Q = P(y|x)$

- S is not d-separated from Y if we condition on X

Example 2

- Consider graph G_s:

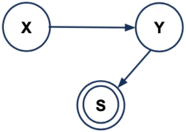


Our goal is to establish s-recoverability of $Q = P(y|x)$

- S is not d-separated from Y if we condition on X
- $P(y|x) = P(y|x, S = 1)$ does not hold in at least one distribution compatible with G_s.

Example 2

- Consider graph G_s:



Our goal is to establish s-recoverability of $Q = P(y|x)$

- S is not d-separated from Y if we condition on X
- $P(y|x) = P(y|x, S = 1)$ does not hold in at least one distribution compatible with G_s.
- Not s-recoverable.

Findings

Corollary 1

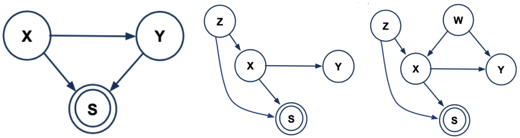
$P(y|x)$ is not s-recoverable when S is dependent on both treatment and outcome.

Theorem 1

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

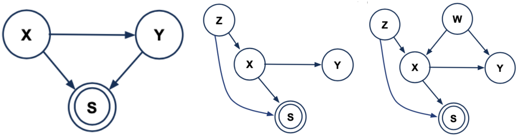
More Examples

- Consider graphs:



More Examples

- Consider graphs:



- not s-recoverable, s-recoverable, not s-recoverable

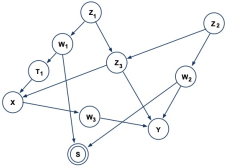
Table of Contents

Recoverability with External Data

A natural question that arises is whether additional measurements in the population level over certain variables can help recovering a given distribution. For example, P (age) can be estimated from census data which is not under selection bias.

Recoverability with External Data

- **Example**:
  Consider Fig. 2 and assume that our goal is to s-recover Q = P(y|x). It follows immediately from Thm. 1 that Q cannot be s-recovered without additional assumptions. However, that the parents of the selection node $Pa_s = \{W1, W2\}$ separates S from all other nodes in the graph.



After conditioning Q on W1 and W2, we obtain:

$P_{y|x} = \sum_{w1,w2} P(y \mid x, w1, w2) P(w1, w2 \mid x) = \sum_{w1,w2} P(y \mid x, w1, w2, S = 1)\ P(w1, w2 \mid x)$, the last equality follows from $(Y \perp\!\!\!\perp S \mid X, W1, W2)$.

Recoverability with External Data

- Our goal is to understand the interplay between measurements taken over two types of variables, M, T $\subseteq$ V , where M are variables collected under selection bias, $P(M \mid S = 1)$, and T are variables collected in the population-level, P(T).
  In other words, we want to understand when (and how) can this new piece of evidence P (T) together with the data under selection (P(M|S = 1)) help in extending the treatment of the previous section for recovering the true underlying distribution Q = P (y|x).

Recoverability with External Data

Definition 2 (s-Recoverability):

Given a causal graph Gs augmented with a node S, the distribution Q = P(y | x) is said to be s-recoverable from selection bias in Gs with external information over T ⊆ V and selection biased data over M ⊆ V (for short, s-recoverable) if the assumptions embedded in the causal model render Q expressible in terms of P(m | S = 1) and P(t), both positive.

When we can choose the variables to be collected. Let *Pas* be the parent set of S. If measurements on the set T = $Pas \cup \{X\}$ can be taken without selection, we can write P (y|x) = $\sum_{pas} P(y \mid x, pas, S = 1)P(pas|x)$, since S is separated from all nodes in the graph given its parent set.

Recoverability with External Data

Theorem 2.

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

$$p(y \mid x) = \sum_c P(y \mid x, c, S = 1) P(c \mid x)$$

In particular, $\{W2, Z3\}$ is such a set, and it allows us to s-recover Q without measuring W1 (W1 $\in$ Pas). Note, however, that the set C = W2 , Z1 , Z2 is not sufficient for s-recoverability. It fails to satisfy the separability condition of the theorem since conditioning on X, W2, Z1, Z2 leaves an unblocked path between S and Y (i.e.,S$\leftarrow$W1 $\rightarrow$T1 $\rightarrow$X$\leftarrow$Z3 $\rightarrow$Y).

Recoverability with External Data

Theorem 3.

There exists some set $C \subseteq T \cap M$ such that $Y \perp\!\!\!\perp S \mid \{C, X\}$ if and only if the set (C'$\cup X$) d-separates S from Y where

$$C' = [(T \cap M) \cap An(Y \cup S \cup X)] \backslash (Y \cup S \cup X)$$

.

In practice, we can restrict ourselves to minimal separators, that is, looking only for minimal set $C' \subseteq T \cap M$ such that $(Y \perp\!\!\!\perp S \mid \{C, X\})$. The algorithm for finding minimal separators has been given in (Acid and de Campos 1996; Tian, Paz, and Pearl 1998).

Recoverability with External Data

Despite the computational advantages given by Thm. 3, Thm. 2 still requires the existence of a separator C measured in both the biased study (M) and in the overall population (T), and it is natural to ask whether this condition can be relaxed. Assume that all we have is a separator $C \subseteq M$, but C (or some of its elements) is not measured in population T, and therefore P (c|x) in eq. (4) still needs to be s-recovered. We could s-recover P (c|x) in the spirit of Thm. 2 as

$$P(c \mid x) = \sum_{c1} P(c \mid x, c1, S = 1) P(c1 \mid x)$$

if there exists a set $C1 \subseteq M \cap T$ such that $(S \perp\!\!\!\perp C \mid X, C1)$.

Recoverability with External Data

extend this idea by considering other possible probabilistic manipulations and embed them in a recursive procedure. For W, $Z \subseteq M$, consider the problem of recovering P (w|z) from P (t) and P (m|S = 1), and define procedure RC(w, z) as follows:

Definition:

1. If $W \cup Z \subseteq T$, then $P(w \mid z)$ is s-recoverable.
2. If $(S \perp\!\!\!\perp W \mid Z)$, then $P(w \mid z)$ is s-recoverable as $P(w \mid z) = P(w \mid z, S = 1)$.
3. For minimal $C \subseteq M$ such that $(S \perp\!\!\!\perp W|(Z \cup C)), P(w \mid z) = \sum_c P(w \mid z, c, S = 1)P(c \mid z)$. If $C \cup Z \subseteq T$, then $P(w \mid z)$ is s-recoverable. Otherwise, call RC(c, z).
4. For some $W` \subset W, P(w \mid z) = P(w` \mid w\backslash w`, z)P(w\backslash w`|z)$. Call $RC(w`, w\ w` \cup z)$ and $RC(w\ w`, z))$.
5. Exit with FAIL (to s-recover P(w|z)) if for a singleton W, none of the above operations are applicable.

Recoverability with External Data

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

Recoverability with External Data

Now we turn our attention to some special cases that appear in practice. Note that, so far, we assumed X being measured in the overall population, but in some scenarios Y 's prevalence might be available instead. So, assume $Y \in T$ but some variables in X are not measured in the population-level. Let $X^0 = X \cap T$and $X^m = X \backslash X^0$,we have

$$P(y \mid x) = \frac{P(x^m \mid y, x^0)p(y \mid x^0)}{\sum_y P(x^m \mid y, x^0)p(y \mid x^0)} (6)$$

Therefore, P(y|x) is recoverable if $P(x^m \mid y, x^0)$ is recoverable.

Recoverability with External Data

Corollary 3

$P(y \mid x)$ is recoverable if there exists a set $C \subseteq T \cap M$(C could be empty) such that$(X^m \perp\!\!\!\perp S | C \cup Y \cup X^0)$. If recoverable, P(y|x) is given by Eq. (6) where

$$P(x^m \mid y, x^0) = \sum_c P(x^m \mid y, x^0, c, S = 1) P(c \mid y, x^0)$$

Corollary 4

P(y|x) is recoverable via Corollary 3 if and only if the set (C' $\cup Y \cup X^0$) d-separates S from X^m where C' = $[(T \cap M) \cap An(Y \cup S \cup X)]$ $(Y \cup S \cup X)$.

Recoverability with External Data

Furthermore, it is worth examining when no data is gathered over X or Y in the population level. In this case, P (y|x) may be recoverable through P (x, y), as shown in the sequel.

Corollary 5

P (y|x) is recoverable if there exists a set $C \subseteq T \cap M$ such that (Y $\cup X \perp\!\!\!\perp S \mid C$). If recoverable, P (y, x) is given by P(y,x) = $\sum_c P(y, x \mid c, S = 1)P(c)$. For instance, P (x, y) is s-recoverable in Fig. 2 if $T \cap M$ contains W2, T1, Z3 or W2, T1, Z1 (without X, Y).

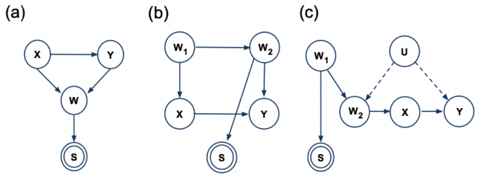
Table of Contents

Recoverability of Causal Effects

- Our goal is to recover the effect of X on Y, $P(y|do(x))$

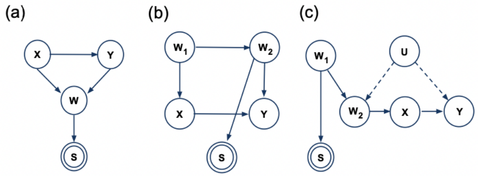
Recoverability of Causal Effects

- Our goal is to recover the effect of X on Y, $P(y|do(x))$
- Consider graphs:



Recoverability of Causal Effects

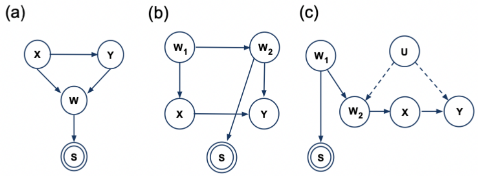
- Our goal is to recover the effect of X on Y, $P(y|do(x))$
- Consider graphs:



- In (a), $P(y|do(x)) = P(y|x)$. But based on Theorem 1, $P(y|x) \neq P(y|x, S = 1)$. So $P(y|do(x))$ is not recoverable.

Recoverability of Causal Effects

- Our goal is to recover the effect of X on Y, $P(y|do(x))$
- Consider graphs:



- In (a), $P(y|do(x)) = P(y|x)$. But based on Theorem 1, $P(y|x) \neq P(y|x, S = 1)$. So $P(y|do(x))$ is not recoverable.
- In (b) and (c), $\{W_2\}$ satisfies the backdoor condition. $P(y|do(x))$ is recoverable or not depends on whether W_2 satisfy conditions similar to those in Theorem 1.

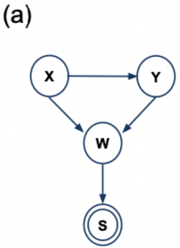
Recoverability of Causal Effects

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, $(Z^- \perp\!\!\!\perp Y \mid X, Z^+)$;
(iii) X and Z block all paths between S and Y, namely, $(Y \perp\!\!\!\perp S \mid X, Z)$;
(iv) $Z \cup \{X, Y\} \subseteq M$, and $Z \subseteq T$.

Recoverability of Causal Effects

- Consider graph G_s:

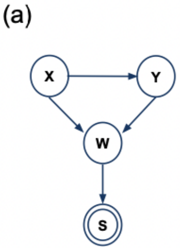


Definition 4 (Selection-backdoor criterion)

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, $(Z^- \perp\!\!\!\perp Y \mid X, Z^+)$;
(iii) X and Z block all paths between S and Y, namely, $(Y \perp\!\!\!\perp S \mid X, Z)$;
(iv) $Z \cup \{X, Y\} \subseteq M$, and $Z \subseteq T$.

Recoverability of Causal Effects

- Consider graph G_s:



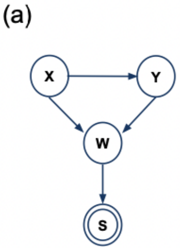
Definition 4 (Selection-backdoor criterion)

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, $(Z^- \perp\!\!\!\perp Y \mid X, Z^+)$;
(iii) X and Z block all paths between S and Y, namely, $(Y \perp\!\!\!\perp S \mid X, Z)$;
(iv) $Z \cup \{X, Y\} \subseteq M$, and $Z \subseteq T$.

- $Y \perp\!\!\!\perp S | \{X, W\}$

Recoverability of Causal Effects

- Consider graph G_s:



Definition 4 (Selection-backdoor criterion)

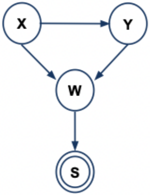
Let a set Z of variables be partitioned into $Z^+ \cup Z^-$ such that Z^+ contains all non-descendants of X and Z^- the descendants of X. Z is said to satisfy the selection backdoor criterion (s-backdoor, for short) relative to an ordered pairs of variables (X, Y) and an ordered pair of sets (M, T) in 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, $(Z^- \perp\!\!\!\perp Y \mid X, Z^+)$;
(iii) X and Z block all paths between S and Y, namely, $(Y \perp\!\!\!\perp S \mid X, Z)$;
(iv) $Z \cup \{X, Y\} \subseteq M$, and $Z \subseteq T$.

- $Y \perp\!\!\!\perp S|\{X, W\}$
- $Z^- = \{W\}$, $Z^+ = \{\}$

Recoverability of Causal Effects

- Consider graph G_s:

(a)



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, $(Z^- \perp\!\!\!\perp Y \mid X, Z^+)$;
(iii) X and Z block all paths between S and Y, namely, $(Y \perp\!\!\!\perp S \mid X, Z)$;
(iv) $Z \cup \{X, Y\} \subseteq M$, and $Z \subseteq T$.

- $Y \perp\!\!\!\perp S | \{X, W\}$
- $Z^- = \{W\}$, $Z^+ = \{\}$
- (ii) violated, Z^- is not separated from Y given $\{X\} \cup Z^+$
- Despite the fact that the relationship between X and Y is unconfounded and $(Y \perp\!\!\!\perp S | \{W, X\})$, it is improper to adjust for $\{W\}$ when computing the target effect.

Recoverability of Causal Effects

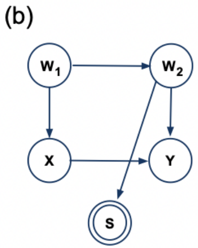
Theorem 5 (Selection-backdoor adjustment).

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

$$P(y \mid do(x)) = \sum_{z} P(y \mid x, \mathbf{z}, S = 1) P(\mathbf{z})$$

Recoverability of Causal Effects

- Consider graph G_s, our goal is to establish $Q = P(y|do(x))$ when external data over $\{W_2\}$ is available in both studies.

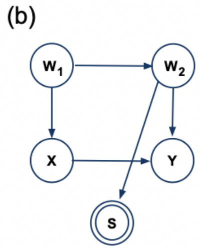


Definition 4 (Selection-backdoor criterion)

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, $(Z^- \perp\!\!\!\perp Y \mid X, Z^+)$;
(iii) X and Z block all paths between S and Y, namely, $(Y \perp\!\!\!\perp S \mid X, Z)$;
(iv) $Z \cup \{X, Y\} \subseteq M$, and $Z \subseteq T$.

Recoverability of Causal Effects

- Consider graph G_s, our goal is to establish $Q = P(y|do(x))$ when external data over $\{W_2\}$ is available in both studies.



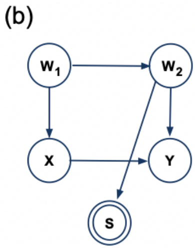
Definition 4 (Selection-backdoor criterion)

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, $(Z^- \perp\!\!\!\perp Y \mid X, Z^+)$;
(iii) X and Z block all paths between S and Y, namely, $(Y \perp\!\!\!\perp S \mid X, Z)$;
(iv) $Z \cup \{X, Y\} \subseteq M$, and $Z \subseteq T$.

- $Z = \{W_2\}$ is s-backdoor admissible and the s-backdoor adjustment is applicable

Recoverability of Causal Effects

- Consider graph G_s, our goal is to establish $Q = P(y|do(x))$ when external data over $\{W_2\}$ is available in both studies.



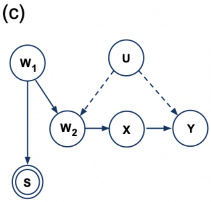
Definition 4 (Selection-backdoor criterion)

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, $(Z^- \perp\!\!\!\perp Y \mid X, Z^+)$;
(iii) X and Z block all paths between S and Y, namely, $(Y \perp\!\!\!\perp S \mid X, Z)$;
(iv) $Z \cup \{X, Y\} \subseteq M$, and $Z \subseteq T$.

- $Z = \{W_2\}$ is s-backdoor admissible and the s-backdoor adjustment is applicable
- If T $= \{W_1\}$, Z $= \{W_1\}$ is backdoor admissible, but it is not s-backdoor admissible since condition (iii) is violated (i.e., $(S \perp\!\!\!\perp Y \mid \{W_1, X\})$ does not hold in G_s)

Recoverability of Causal Effects

- Consider graph G_s, our goal is to establish $Q = P(y|do(x))$ when $T = \{W_2\}$ is available.

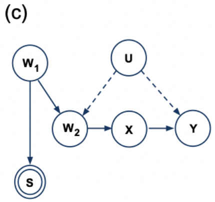


Definition 4 (Selection-backdoor criterion)

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, $(Z^- \perp\!\!\!\perp Y \mid X, Z^+)$;
(iii) X and Z block all paths between S and Y, namely, $(Y \perp\!\!\!\perp S \mid X, Z)$;
(iv) $Z \cup \{X, Y\} \subseteq M$, and $Z \subseteq T$.

Recoverability of Causal Effects

- Consider graph G_s, our goal is to establish $Q = P(y|do(x))$ when $T = \{W_2\}$ is available.



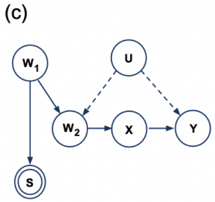
Definition 4 (Selection-backdoor criterion)

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, $(Z^- \perp\!\!\!\perp Y \mid X, Z^+)$;
(iii) X and Z block all paths between S and Y, namely, $(Y \perp\!\!\!\perp S \mid X, Z)$;
(iv) $Z \cup \{X, Y\} \subseteq M$, and $Z \subseteq T$.

- When $Z = \{W_2\}$, S-backdoor criterion fails since condition (iii) is violated (i.e., $(S \perp\!\!\!\perp Y \mid \{W_2, X\})$ does not hold in G_s)

Recoverability of Causal Effects

- Consider graph G_s, our goal is to establish $Q = P(y|do(x))$ when $T = \{W_2\}$ is available.



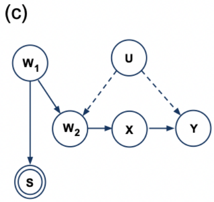
Definition 4 (Selection-backdoor criterion)

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, $(Z^- \perp\!\!\!\perp Y \mid X, Z^+)$;
(iii) X and Z block all paths between S and Y, namely, $(Y \perp\!\!\!\perp S \mid X, Z)$;
(iv) $Z \cup \{X, Y\} \subseteq M$, and $Z \subseteq T$.

- When $Z = \{W_2\}$, S-backdoor criterion fails since condition (iii) is violated (i.e., $(S \perp\!\!\!\perp Y \mid \{W_2 \, , X\})$ does not hold in G_s)
- Alternatively, if consider $Z = \{\}$, condition (i) fails

Recoverability of Causal Effects

- Consider graph G_s, our goal is to establish $Q = P(y|do(x))$ when $T = \{W_2\}$ is available.

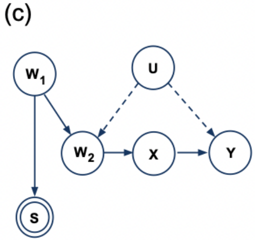


Definition 4 (Selection-backdoor criterion)

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, $(Z^- \perp\!\!\!\perp Y \mid X, Z^+)$;
(iii) X and Z block all paths between S and Y, namely, $(Y \perp\!\!\!\perp S \mid X, Z)$;
(iv) $Z \cup \{X, Y\} \subseteq M$, and $Z \subseteq T$.

- When $Z = \{W_2\}$, S-backdoor criterion fails since condition (iii) is violated (i.e., $(S \perp\!\!\!\perp Y \mid \{W_2 , X\})$ does not hold in G_s)
- Alternatively, if consider $Z = \{\}$, condition (i) fails
- But, $P(y|do(x))$ **is recoverable**.

What is the expression of $P(y|do(x))$ according to the graph?



Hint: Using s-recoverability and do-calculus. And read the paper

Email: shaolinx@uci.edu, zhangz28@uci.edu