Bounds on Causal Effects and Application to High Dimensional Data

Ang Li, Judea Pearl

1 Department of Computer Science, University of California, Los Angeles, California, USA
angli@cs.ucla.edu, judea@cs.ucla.edu

Abstract

This paper addresses the problem of estimating causal effects when adjustment variables in the back-door or front-door criterion are partially observed. For such scenarios, we derive bounds on the causal effects by solving two non-linear optimization problems, and demonstrate that the bounds are sufficient. Using this optimization method, we propose a framework for dimensionality reduction that allows one to trade bias for estimation power, and demonstrate its performance using simulation studies.

Introduction

Estimating causal effects has been encountered in many areas of industry, marketing, and health science, and it is the most critical problem in causal inference. Pearl’s back-door and front-door criteria, along with the adjustment formula (Pearl 1995), are powerful tools for estimating causal effects. In this paper, the problem of estimating causal effects when adjustment variables in the back-door or front-door criterion are partially observable, or when the adjustment variables have high dimensionality, is addressed.

Consider the problem of estimating the causal effects of $X$ on $Y$ when a sufficient set $W \cup U$ of confounders is partially observable (see Figure 1). Because $W \cup U$ is assumed to be sufficient, the causal effects are identified from measurements on $X, Y, W,$ and $U$ and can be written as

$$P(y|do(x)) = \sum_{w,u} P(y|x, w, u)P(w, u)$$

$$= \sum_{w,u} \frac{P(x, y, w, u)P(w, u)}{P(x, w, u)}.$$

However, if $U$ is unobserved, $d$-separation tells us immediately that adjusting for $W$ is inadequate by leaving the back-door path $X \leftarrow U \rightarrow Y$ unblocked. Therefore, regardless of sample size, the causal effects of $X$ on $Y$ cannot be estimated without bias. However, it turns out that when given a prior distribution $P(U)$, we can obtain bounds on the causal effects. We will demonstrate later that the midpoints of the bounds are sufficient for estimating the causal effects.

Bounding has been proven to be useful in causal inference. (Balke and Pearl 1997a) provided bounds on causal effects with imperfect compliance, (Tian and Pearl 2000) provided bounds on probabilities of causation, (Cai et al. 2008) provided bounds on causal effects with the presence of confounded intermediate variables, and (Li and Pearl 2019) proposed bounds on the benefit function of a unit selection problem.

Although $P(U)$ is assumed to be given, it is usually known regardless of the model itself (e.g., $U$ stands for gender, gene type, blood type, or age). Alternatively, if costs permit, one can estimate $P(U)$ by re-testing within a small sampled sub-population.

A second problem considered in this paper is that of estimating causal effects when a sufficient set $Z$ of confounders is fully observable (see Figure 2), but with a high dimensionality (e.g., $Z$ has 1024 instantiates). In such a case, a prohibitively large sample size would be required, which is generally recognized to be impractical. We propose a new framework that transforms the problem associated with Figure 2 into an equivalent problem associated with Figure 3 containing $W$ and $U$, which have much smaller dimensionalities (e.g., $W$ and $U$ have 32 instantiates). We then estimate bounds on causal effects of the equivalent problem and take the midpoints as the effect estimates. We demonstrate through a simulation that this method can deliver good estimates of causal effects of the original problem.

Preliminaries & Related Works

In this section, we review the back-door and front-door criteria and their associated adjustment formulas (Pearl 1995). We use the causal diagrams in (Pearl 1995; Spirtes et al. 2000; Pearl 2009; Koller and Friedman 2009).
The back-door and front-door criteria are two powerful tools for estimating causal effects; however, causal effects are not identifiable if the set of adjustment variables $Z$ is not fully observable. (Tian and Pearl 2000) provided the naivest bounds for causal effects (Equation 3), regardless of the causal diagram.

$$P(x, y) \leq P(y|do(x)) \leq 1 - P(x, y').$$ (3)

As the first contribution of this study, we obtain narrower bounds of the causal effects by leveraging another source of knowledge, i.e., a causal diagram behind data combined with measurements of a set $W$ (observable part of $Z$) of covariates and a prior information of a set $U$ (unobservable part of $Z$), in a causal diagram in which the bounds are solutions to two non-linear optimization problems. We illustrate that the midpoints of the bounds are sufficient for estimating the causal effects.

Using this optimization method, our second contribution is the proposal of a new framework for estimating causal effects when a set of fully observable adjustment variables $Z$ has a high dimensionality without any assumption regarding the data-generating process. (Maathuis et al. 2009) proposed a method of estimating causal effects when the number of covariates is larger than the sample size. However, it relies on several assumptions, including the assumption that the distribution of covariates is multivariate normal. The method is limited if the distribution of covariates is unknown or does not have an accuracy estimate owing to the limitation of the sample size.

**Bounds on Causal Effects**

In this section, we demonstrate how bounds on causal effects with partially observable back-door or front-door variables can be obtained through non-linear optimizations.

**Partially Observable Back-Door Variables**

**Theorem 4.** Given a causal diagram $G$ and a distribution compatible with $G$, let $W \cup U$ be a set of variables satisfying the back-door criterion in $G$ relative to an ordered pair $(X, Y)$, where $W \cup U$ is partially observable, i.e., only probabilities $P(X, Y, W)$ and $P(U)$ are given, the causal effects of $X$ on $Y$ are then bounded as follows:

$$LB \leq P(y|do(x)) \leq UB$$

where $LB$ is the solution to the non-linear optimization problem in Equation 4 and $UB$ is the solution to the non-linear...
optimization problem in Equation 5.

\[ LB = \min \sum_{w, u} a_{w, u} b_{w, u}, \quad (4) \]

\[ UB = \max \sum_{w, u} a_{w, u} b_{w, u}, \quad (5) \]

where,
\[ \sum_{u} a_{w, u} = P(x, y, w), \quad \sum_{u} b_{w, u} = P(w), \]
\[ \sum_{u} c_{w, u} = P(x, w) \text{ for all } w \in W; \]

and for all \( w \in W \) and \( u \in U \),
\[ b_{w, u} \geq c_{w, u} \geq a_{w, u}, \]
\[ \max \{0, p(x, y, w) + p(u) - 1\} \leq a_{w, u}, \]
\[ \min \{P(x, y, w), p(u)\} \geq a_{w, u}, \]
\[ \max \{0, p(w) + p(u) - 1\} \leq b_{w, u}, \]
\[ \min \{P(w), p(u)\} \geq b_{w, u}, \]
\[ \max \{0, p(x, w) + p(u) - 1\} \leq c_{w, u}, \]
\[ \min \{P(x, w), p(u)\} \geq c_{w, u}. \]

Partially Observable Front-Door Variables

Theorem 5. Given a causal diagram \( G \) and distribution compatible with \( G \), let \( W \cup U \) be a set of variables satisfying the front-door criterion in \( G \) relative to an ordered pair \((X, Y)\), where \( W \cup U \) is partially observable, i.e., only probabilities \( P(X, Y, W) \) and \( P(U) \) are given and \( P(x, W, U) > 0 \), the causal effects of \( X \) on \( Y \) are then bounded as follows:

\[ LB \leq P(y|do(x)) \leq UB \]

where \( LB \) is the solution to the non-linear optimization problem in Equation 6 and \( UB \) is the solution to the non-linear optimization problem in Equation 7.

\[ LB = \min \sum_{w, u} \frac{b_{w, u}}{P(x)} \sum_{x'} \frac{a_{x', w, u} P(x')}{b_{x', w, u}}, \quad (6) \]

\[ UB = \max \sum_{w, u} \frac{b_{w, u}}{P(x)} \sum_{x'} \frac{a_{x', w, u} P(x')}{b_{x', w, u}}, \quad (7) \]

where,
\[ \sum_{u} a_{x, w, u} = P(x, y, w), \quad \sum_{u} b_{x, w, u} = P(x, w) \]

for all \( x \in X \) and \( w \in W \); and for all \( x \in X, w \in W, \) and \( u \in U \),
\[ b_{x, w, u} \geq a_{x, w, u}, \]
\[ \max \{0, p(x, y, w) + p(u) - 1\} \leq a_{x, w, u}, \]
\[ \min \{P(x, y, w), p(u)\} \geq a_{x, w, u}, \]
\[ \max \{0, p(x, w) + p(u) - 1\} \leq b_{x, w, u}, \]
\[ \min \{P(x, w), p(u)\} \geq b_{x, w, u}. \]

Notably, if any observational data (e.g., \( P(U) \)) are unavailable in the above theorems, we can remove that term, and the rest of non-linear optimization problems still provide valid bounds for the causal effects. In general, midpoints of bounds on causal effects are effective estimates. However, the lower (upper) bounds are also informative, which can be interpreted as the minimal (maximal) causal effects. The proofs of Theorems 4 and 5 are provided in the appendix.

Example

Herein, we present a simulated example to demonstrate that the midpoints of the bounds on the causal effects given by Theorem 4 are adequate for estimating the causal effects.

Causal Effect of a Drug

Drug manufacturers want to know the causal effect of recovery when a drug is taken. Thus, they conduct an observational study. Here, the recovery rates of 700 patients were recorded. A total of 192 patients chose to take the drug and 508 patients did not. The results of the study are shown in Table 1. Blood type (type O or not) is not the only confounder of taking the drug and recovery. Another confounder is age (below the age of 70 or not). The manufacturers have no data associated with age. They only know that 85.43% of people in their region are below the age of 70.

Because both age and blood type are confounders of taking the drug and recovery, and the observational data associated with age are unobservable, the causal effect is not identifiable.

Let \( X = x \) denote the event that a patient took the drug, and \( X = x' \) denote the event that a patient did not take the drug. Let \( Y = y \) denote the event that a patient recovered, and \( Y = y' \) denote the event that a patient did not recover. Let \( W = w \) represent a patient with blood type O, and \( W = w' \) represent a patient without blood type O. Let \( U = u \) represent a patient below the age of 70, and \( U = u' \) represent a patient above the age of 70. The causal diagram is shown in Figure 4.

An option for the manufacturers could be estimating the causal effect through the Tian-Pearl bounds in Equation 3.

\[ \sum_{u} a_{x, w, u} = P(x, y, w), \quad \sum_{u} b_{x, w, u} = P(x, w) \]

Figure 4: Needed the causal effects of \( X \) on \( Y \) when \( U \) is unobserved and independent with \( W \).
We have access to the fully observed observational data, as summarized in Table 1. The error of the estimate of the causal effect is extremely close to the actual causal effect, particularly when the causal effects are close to 0 and 1. The average gap (upper bounds – lower bounds), 0.383, of the proposed bounds among 1000 samples is much smaller than the average gap, 0.487, of the Tian-Pearl bounds among 1000 samples. This means that the midpoints of the proposed bounds are more convincing, because the bounds are narrower.
would be required to estimate the causal effects, which is 

to obtain

Figure 5: Bounds on causal effects of 100 sample distributions with partially observed confounders, where the Tian-Pearl bounds are obtained through Equation 3 and the proposed bounds are obtained through Theorem 4.

**Application to High Dimensionality of Adjustment Variables**

Consider the problem of estimating the causal effects of $X$ on $Y$ when a sufficient set $Z$, which satisfies the back-door or front-door criterion, is fully observable (e.g., see Figure 2) in a causal diagram $G$ but has high dimensionality (e.g., $Z$ has 1024 instantiates), a prohibitive large sample size would be required to estimate the causal effects, which is generally recognized to be impractical. Herein, we propose a new framework to achieve dimensionality reduction.

**Equivalent Causal Diagram with Observational Data**

**Definition 6** (Equivalent causal diagram with observational data). Let $G, G'$ be causal diagrams both containing nodes $X$, $Y$. $O$ are observational data compatible with $G$, and $O'$ are observational data compatible with $G'$. We say that $(G, O)$ is equivalent to $(G', O')$ with respect to $P(y|do(x))$ if the causal effects of $X$ on $Y$ with $(G, O)$ is equal to the causal effects of $X$ on $Y$ with $(G', O')$.

This equivalent tuple $(G', O')$ is easy to obtain. We can simply add two new nodes $W$ and $U$, and remove a node $Z$ in $G$ to obtain $G'$. Let the arrows entering $Z$ in $G$ now enter both $W$ and $U$ in $G'$, and let the arrows exiting $Z$ in $G$ now exit both $W$ and $U$ in $G'$. Finally, add an arrow from $U$ to $W$. It is easy to show that $(G, O)$ and $(G', O')$ are equivalent if the states of $Z$ are the Cartesian product of the states of $W$ and the states of $U$. Formally, we have the following theorem (the proof of the theorem is provided in the appendix).

**Theorem 7**. Let $G$ be a causal diagram containing nodes \{V_1, \ldots, V_{n-3}, X, Y, Z\}. Let $O$ be any observational data compatible with $G$. Suppose there exists a set of variables that satisfies the back-door or front-door criterion relative to $(X, Y)$ in $G$, then, $(G, O)$ is equivalent to $(G', O')$ with respect to $P(y|do(x))$ $(G'$ containing nodes \{V_1, \ldots, V_{n-3}, X, Y, Z, W, U\}; $O'$ are observational data compatible with $G'$), where the number of states in $W$ times the number of states in $U$ is equal to the number of states in $Z$, and the structure of $G'$ and the observational data $O'$ are obtained as follows:

**Structure of $G'$**

Let $\text{Parents}_G(H)$ be the parents of $H$ in causal diagram $G$.

$\text{Parents}_G(U) = \text{Parents}_G(Z)$.

$\text{Parents}_G(W) = \text{Parents}_G(Z) \cup \{U\}$.

$\text{Parents}_G(H) = \text{Parents}_G(H)$ if $Z \notin \text{Parents}_G(H)$ for $H \in \{V_1, \ldots, V_{n-3}, X, Y\}$.

$\text{Parents}_G(H) = \text{Parents}_G(H) \setminus \{Z\} \cup \{W, U\}$ if $Z \in \text{Parents}_G(H)$ for $H \in \{V_1, \ldots, V_{n-3}, X, Y\}$.

Note that, let $Q$ be the set of variables in $G$ that satisfies the back-door or front-door criterion relative to $(X, Y)$, then $Q'$ satisfies the back-door or front-door criterion relative to $(X, Y)$ in $G'$, where $Q' = Q$ if $Z \notin Q$.

$Q' = Q \setminus \{Z\} \cup \{W, U\}$ if $Z \in Q$.

**Observational data**:

Let $p$ be the number of states in $W$, and $q$ be the number of states in $U$.

The states of $Z$ are the Cartesian product of the states of $W$ and the states of $U$.

In detail, $(w_j, u_k)$ is equivalent to $z_{(j-1)q+k}$. $w_j$ is equivalent to $v_{j=1}^p(w_j, u_k) = v_{j=1}^p z_{(j-1)q+k}$ and $u_k$ is equivalent to $v_{j=1}^q(w_j, u_k) = v_{j=1}^q z_{(j-1)q+k}$, i.e., $P(w_j, u_k, V) = P(z_{(j-1)q+k}, V)$ for any $V \subseteq \{V_1, \ldots, V_{n-3}, X, Y\}$.

For example, consider the causal diagram in Figure 2 and the observational data (in the form of conditional probability tables (CPTs), where $X, Y$ are binary, and $Z$ has 4 states.) in Table 3. The causal effect, $P(y|do(x))$, through the adjustment formula in Equation 1, is 0.47. Based on the construction in Theorem 7 (see the appendix for details), we have the causal diagram in Figure 2 with the observational data in Table 3 is equivalent to the causal diagram in Figure 3 with the observational data in Table 4 (all nodes are binary), and we can verify that the causal effect, $P(y|do(x))$, in the causal diagram in Figure 3 with the observational data in Table 4 is also 0.47.

Table 3: Observational data in CPTs compatible with the causal diagram in Figure 2.

| $P(z_1)$ | $P(y|x, z_1)$ |
|---|---|
| 0.3 | 0.2 |

| $P(z_2)$ | $P(y|x, z_2)$ |
|---|---|
| 0.2 | 0.7 |

| $P(z_3)$ | $P(y|x, z_3)$ |
|---|---|
| 0.2 | 0.1 |

| $P(z_4)$ | $P(y|x, z_4)$ |
|---|---|
| 0.3 | 0.6 |

| $P(x_1)$ | $P(y|x, z_1)$ |
|---|---|
| 0.1 | 0.5 |

| $P(x_2)$ | $P(y|x, z_2)$ |
|---|---|
| 0.4 | 0.5 |

| $P(x_3)$ | $P(y|x, z_3)$ |
|---|---|
| 0.5 | 0.5 |

| $P(x_4)$ | $P(y|x, z_4)$ |
|---|---|
| 0.7 | 0.4 |

Notably, the equivalent tuple is not unique and is transitive (i.e., if $(G, O)$ is equivalent to $(G', O')$, and $(G', O')$ is equivalent to $(G'', O'')$, then $(G, O)$ is equivalent to $(G'', O'')$).
Algorithm 1: Generate Equivalent Tuple

Input: A n nodes, \((X_1, X_2, \ldots, X_{n-3}, X, Y, Z)\), causal diagram \(G\) and compatible \(O\); \(p\), the number of states in \(W\) in \(G'\) of the equiv. tuple \((G', O')\).

Output: A \(n + 1\) nodes, \((X_1, X_2, \ldots, X_{n-3}, X, Y, W, U)\), causal diagram \(G'\);

Mapping relation \(M_1\): state of \(W \rightarrow\) state of \(Z\);

Mapping relation \(M_2\): state of \(U \rightarrow\) state of \(Z\).

1: \(m = \text{num}_{\text{states}}\) in \(G(Z)\);
2: if \(m \mod p == 0\) then
3: \(q = m/p;\)
4: else
5: \(q = m/p + 1;\)
6: end if
7: if \(Z \in \text{Parents}_G(H)\) then
8: \(\text{parents}_G(H) = \text{parents}_G(H) \setminus \{Z\} \cup \{W, U\};\)
9: else
10: \(\text{parents}_G(H) = \text{parents}_G(H);\)
11: end if
12: end for
13: \(\text{num}_{\text{states}}\) in \(G'(W) = p;\)
14: \(\text{num}_{\text{states}}\) in \(G'(U) = q;\)
15: \(\text{parents}_G'(W) = \text{parents}_G(Z) \cup \{U\};\)
16: \(\text{parents}_G'(U) = \text{parents}_G(Z);\)
17: for \(i = 1\) to \(p\) do
18: \(M_{1}(w_i) = \sqrt[p]{q}^k_{k=1} z_{(i-1)q+k};\)
19: end for
20: for \(i = 1\) to \(q\) do
21: \(M_{2}(u_i) = \sqrt[q]{p}^j_{j=1} z_{(j-1)q+i};\)
22: end for

Finally, let's consider how many samples are required for each method. According to (Roscoe 1975), each state needs at least 30 samples, and therefore, the exact solution by Equation 1 requires \(2 \times 2 \times 256 \times 30 = 30720\) samples. However, the proposed bounds based on Theorem 4 only requires \(\max(2 \times 2 \times 16, 16) \times 30 = 1920\) samples. If the sample size is still unacceptable, we can use another equivalent tuple with \(W\) having 8 states and \(U\) having 32 states, then we only require \(\max(2 \times 2 \times 8, 32) \times 30 = 960\) samples to obtain the bounds on the causal effects.

Simulation Results

Similarly to the previous simulation, we further illustrate that the bounds on the causal effects of the proposed framework are sufficient for estimating the original causal effects.

Once again, by employing the simplest causal diagram in Figure 2, where \(X\) and \(Y\) are binary and \(Z\) has 256 states. We randomly generated 100 sample distributions compatible with the causal diagram (the algorithm for generating the distributions are shown in the appendix). The average

**Dimensionality Reduction**

Now, considering the problem in the beginning of the section Application to High Dimensionality of Adjustment Variables. First, we transform the causal diagram \(G\) with the compatible observational data \(O\) into an equivalent tuple \((G', O')\) using Algorithm 1 based on the construction in Theorem 7 (note that the algorithm only construct the structure of the \(G'\) and assigning the meaning of the states for \(W, U\), the corresponding observational data \(O'\) are then easy to obtain), then the new problem \((G', O')\) has the same causal effects of \(X\) on \(Y\) as in \((G, O)\). By picking the dimensionality of \(W\) in Algorithm 1), we can control the dimensionality of the new problem.

Note that, if \(Z = (Z_1, Z_2, \ldots, Z_n)\) in \(G\) is a set of variables, we can repeat Algorithm 1 for each variable in \(Z\), and finally obtain \(W = (W_1, W_2, \ldots, W_m)\) and \(U = (U_1, U_2, \ldots, U_m)\), where the multiplication of the number of states in \(W\) is equal to \(p\).

We then treat the new problem \((G', O')\) as a partially observable back-door or front-door variables problem in Sections and , where \(P(X, Y, W)\) and \(P(U)\) are given, and we can then obtain the bounds of the causal effects through Theorems 4 and 5. We claim that the midpoints of the bounds are good estimates of the original causal effects. In addition, the bounds themselves will help make decisions.

**Example**

Consider the problem in Figure 2, where \(X\) and \(Y\) are binary and \(Z\) has 256 states. We randomly generated a distribution \(P(X, Y, Z)\) that is compatible with the causal diagram. Because we know the exact distribution, we can easily obtain the causal effects through Equation 1. The causal effect \(P(y|do(x))\) is 0.5527 (the algorithm for generating the distribution is shown in the appendix).

Now, we transform the causal diagram with the observational data into an equivalent tuple \((G', O')\) (\(G'\) is shown in Figure 3) using Algorithm 1 \((p = 16)\). We obtain the variable \(W\) of 16 states and the variable \(U\) of 16 states in \(G'\) \((w_j, u_k)\) is equivalent to \(z_{(j-1)q+k}\). We are then forced to use only observational data \(P(X, Y, W)\) and \(P(U)\) (the construction of \(P(X, Y, W)\) and \(P(U)\) is shown in the appendix), and based on Theorem 4, with the “SLSQP” solver, we obtain the bounds on the causal effect \(p(y|do(x))\), which are 0.4595 \(\leq P(y|do(x)) \leq 0.7012\). We see the midpoint, 0.5804, is extremely close to the actual causal effect, 0.5527.

### Table 4: Observational data in CPTs compatible with the causal diagram in Figure 3.

| \(P(u)\) | 0.5 |
| \(P(u|u)\) | 0.6 |
| \(P(u|w)\) | 0.4 |
| \(P(x|u, w)\) | 0.1 |
| \(P(x|u, w')\) | 0.4 |
| \(P(x|u')\) | 0.5 |
| \(P(x|u', w)\) | 0.7 |
| \(P(x|u', w')\) | 0.5 |
| \(P(y|x, u, w)\) | 0.2 |
| \(P(y|x', u, w)\) | 0.3 |
| \(P(y|x, u, w')\) | 0.7 |
| \(P(y|x', u, w')\) | 0.1 |
| \(P(y|x, u', w)\) | 0.6 |
| \(P(y|x', u', w)\) | 0.5 |
| \(P(y|x', u', w')\) | 0.5 |
| \(P(y|x', u, w')\) | 0.4 |
Next, bounds in Theorems 4 and 5 are given by non-linear optimizations. Therefore, the quality of the bounds also depends on the optimization solver. The examples and simulated results in this paper are all obtained from the simplest “SLSQP” solver from 1988. The quality of the bounds can be improved if more advanced solvers are applied. Inspired by the idea of Balke’s linear programming (Balke and Pearl 1997b), we may obtain parametric solutions to non-linear optimizations in Theorems 4 and 5, we then do not need a non-linear optimization solver. However, the problem related to a non-linear optimization solver is not the scope of this paper.

Moreover, the midpoint of the bounds are used in this paper, however, the information that the bounds provided are far more than the midpoint. The lower (upper) bound represents the minimal (maximal) causal effects. One can define their own way to use the bounds, but this is not the scope of this paper.

In addition, the constraints in Theorems 4 and 5 are only based on the basic back-door or front-door criterion. We can also add constraints of independencies in a specific graph. For instance, $W$ and $U$ are independent in the causal diagram of Figure 4, we can then add the constraints that reflect $P(W)$ and $P(U)$ as being independent. The greater the number of constraints that are added to the optimizations, the better the bounds we can obtain.

Moreover, if one believes they have a sufficient sample size to estimate causal effects with high dimensionality adjustment variables, the framework in the section Application to High Dimensionality of Adjustment Variables could be evidence validating whether the sample size is indeed sufficient.

Next, in the section Application to High Dimensionality of Adjustment Variables, we transformed $(G, O)$ into $(G', O')$ to obtain the bounds on causal effects with high dimensionality adjustment variables. However, for a tuple $(G, O)$, multiple equivalent tuples exist by picking a different $p$ in Algorithm 1, and each of the equivalent tuple has bounds for the original causal effects. We can compute bounds for as many equivalent tuples as we want and take the maximal lower bounds and the minimal upper bounds.

Finally, based on numerous experiments, we realized that when $P(U)$ or $P(W)$ is specific (i.e., closer to 0 or 1), the proposed bounds are almost identified (i.e., the bounds shrink to point estimates). Therefore, in practice, we can always pick the equivalent tuple to transform, in which the $P(U)$ or $P(W)$ is close to 0 or 1.

**Conclusion**

We demonstrated how to estimate causal effects when adjustment variables in the back-door or front-door criterion are partially observable by bounding the causal effects using solutions to non-linear optimizations. We provided examples and simulated results illustrating that the proposed method is sufficient to estimate the causal effects. We also proposed a framework for estimating causal effects when the adjustment variables have a high dimensionality. In summary, we analyzed and demonstrated how causal effects can be gained in practice using a causal diagram.
Acknowledgements
This research was supported in parts by grants from the National Science Foundation [#IIS-2106908], Office of Naval Research [#N00014-17-S-12091 and #N00014-21-1-2351], and Toyota Research Institute of North America [#PO-000897].

References
Proof of Theorem 4

Theorem 4. Given a causal diagram $G$ and a distribution compatible with $G$, let $W \cup U$ be a set of variables satisfying the back-door criterion in $G$ relative to an ordered pair $(X, Y)$, where $W \cup U$ is partially observable, i.e., only probabilities $P(X, Y, W)$ and $P(U)$ are given, the causal effects of $X$ on $Y$ are then bounded as follows:

$$LB \leq P(y|do(x)) \leq UB$$

where $LB$ is the solution to the non-linear optimization problem in Equation 9 and $UB$ is the solution to the non-linear optimization problem in Equation 10.

$$LB = \min \sum_{w,u} a_{w,u} b_{w,u}, \quad (9)$$

$$UB = \max \sum_{w,u} a_{w,u} b_{w,u}, \quad (10)$$

where,

$$\sum_u a_{w,u} = P(x, y, w), \sum_u b_{w,u} = P(w), \quad \sum_u c_{w,u} = P(x, w) \text{ for all } w \in W;$$

and for all $w \in W$ and $u \in U$,

$$b_{w,u} \geq c_{w,u} \geq a_{w,u},$$

$$\max \{|0, P(x, y, w) + p(u) - 1\} \leq a_{w,u},$$

$$\min \{P(x, y, w), p(u)\} \geq a_{w,u},$$

$$\max \{|0, P(w) + p(u) - 1\} \leq b_{w,u},$$

$$\min \{P(w), p(u)\} \geq b_{w,u},$$

$$\max \{|0, P(x, w) + p(u) - 1\} \leq c_{w,u},$$

$$\min \{P(x, w), p(u)\} \geq c_{w,u}. $$

Proof. To show that the LB and UB bound the actual causal effects, we only need to show that there exists a point in feasible space of the non-linear optimization that $\sum_{w,u} a_{w,u} b_{w,u} c_{w,u}$ is equal to the actual causal effects.

Since $W \cup U$ satisfies the back-door criterion, by adjustment formula in Equation 1, we have,

$$P(y|do(x)) = \sum_{w,u} P(y|x, w, u) P(w, u)$$

$$= \sum_{w,u} \frac{P(x, y, w, u) P(w, u)}{P(x, w, u)}$$

Let

$$a_{w,u} = P(x, y, w, u)$$

$$b_{w,u} = P(w, u)$$

$$c_{w,u} = P(x, w, u)$$

We now show that the above set of $a_{w,u}, b_{w,u}, c_{w,u}$ are in feasible space.

We have,

for $w \in W$,

$$\sum_u a_{w,u} = \sum_u P(x, y, w, u) = P(x, y, w),$$

$$\sum_u b_{w,u} = \sum_u P(w, u) = P(w),$$

$$\sum_u c_{w,u} = \sum_u P(x, w, u) = P(x);$$

and,

for all $w \in W$ and $u \in U$,

$$b_{w,u} = P(w, u) \geq P(x, w, u) = a_{w,u},$$

$$c_{w,u} = P(x, w, u) \leq \min \{P(x, y, w), p(u)\},$$

$$b_{w,u} = P(w, u) \leq \min \{P(w), p(u)\},$$

$$c_{w,u} = P(x, w, u) \leq \min \{P(x, w), p(u)\},$$

Therefore, the above set of $a_{w,u}, b_{w,u}, c_{w,u}$ are in feasible space, and thus, the UB and LB bound the actual causal effects.

Proof of Theorem 5

Theorem 5. Given a causal diagram $G$ and distribution compatible with $G$, let $W \cup U$ be a set of variables satisfying the front-door criterion in $G$ relative to an ordered pair $(X, Y)$, where $W \cup U$ is partially observable, i.e., only probabilities $P(X, Y, W)$ and $P(U)$ are given and $P(x, w, U) > 0$, the causal effects of $X$ on $Y$ are then bounded as follows:

$$LB \leq P(y|do(x)) \leq UB$$

where $LB$ is the solution to the non-linear optimization problem in Equation 11 and $UB$ is the solution to the non-linear optimization problem in Equation 12.

$$LB = \min \sum_{w,u} b_{x,w,u} P(x) \sum_{x'} a_{x',w,u} P(x'), \quad (11)$$

$$UB = \max \sum_{w,u} b_{x,w,u} P(x) \sum_{x'} a_{x',w,u} P(x'), \quad (12)$$

where,

$$\sum_u a_{x,w,u} = P(x, y, w), \sum_u b_{x,w,u} = P(x, w)$$

for all $x \in X \cup W$; and for all $x \in X, w \in W$, and $u \in U$,

$$b_{x,w,u} \geq a_{x,w,u},$$

$$\max \{|0, P(x, y, w) + p(u) - 1\} \leq a_{x,w,u},$$

$$\min \{P(x, y, w), p(u)\} \geq a_{x,w,u},$$

$$\max \{|0, P(x, w) + p(u) - 1\} \leq b_{x,w,u},$$

$$\min \{P(x, w), p(u)\} \geq b_{x,w,u}.$$
Proof. To show that the LB and UB bound the actual causal effects, we only need to show that there exists a point in feasible space of the non-linear optimization that \[ \sum_{w,u} \frac{b_{w,u}}{P(x)} \sum_{x'} \frac{a_{w,u}P(x')}{b_{y',w,u}} \] is equal to the actual causal effects.

Since \( W \cup U \) satisfies front-door criterion and \( P(u, W, U) > 0 \), by adjustment formula in Equation 2, we have,

\[
P(y|do(x)) = \sum_{w,u} P(w,u|x) \sum_{x'} P(y|x', w, u)P(x') \]

\[
= \sum_{w,u} \frac{P(x,w,u)}{P(x)} \sum_{x'} \frac{P(x', y, w, u)P(x')}{P(x', w, u)}.
\]

Let

\[ a_{x,w,u} = P(x, y, w, u), \]
\[ b_{x,w,u} = P(x, w, u). \]

Similarly to the proof of Theorem 4, it is easy to show that the above set of \( a_{x,w,u}, b_{x,w,u} \) are in feasible space, and therefore, LB and UB bound the actual causal effects. \( \square \)

Proof of Theorem 7

Theorem 7. Let \( G \) be a causal diagram containing nodes \( \{V_1, ..., V_{n-3}, X, Y, Z\} \). Let \( O \) be any observational data compatible with \( G \). Suppose there exists a set of variables that satisfies the back-door or front-door criterion relative to \( (X, Y) \) in \( G \), then \( (G, O) \) is equivalent to \( (G', O') \) \( (G' \) containing nodes \( \{V_1, ..., V_{n-3}, X, Y, W, U\} \); \( O' \) is observational data compatible with \( G' \), where the number of states in \( W \) times the number of states in \( U \) is equal to the number of states in \( Z \), and the structure of \( G' \) and the observational data \( O' \) are obtained as follows:

Structure of \( G' \):

Let \( Parents_G(H) \) be the parents of \( H \) in causal diagram \( G \).

\[
Parents_G(U) = Parents_G(Z), \quad Parents_G(W) = Parents_G(Z) \cup \{U\},
\]

\[
Parents_G(H) = Parents_G(H) \text{ if } Z \notin Parents_G(H) \text{ for } H \in \{V_1, ..., V_{n-3}, X, Y\},
\]

\[
Parents_G(H) = Parents_G(H) \setminus \{Z\} \cup \{W, U\} \text{ if } Z \in Parents_G(H) \text{ for } H \in \{V_1, ..., V_{n-3}, X, Y\}.
\]

Note that, let \( Q \) be the set of variables in \( G \) that satisfies the back-door or front-door criterion relative to \( (X, Y) \), then \( Q' \) satisfies the back-door or front-door criterion relative to \( (X, Y) \) in \( G' \), where

\[ Q' = Q \text{ if } Z \notin Q, \]
\[ Q' = Q \setminus \{Z\} \cup \{W, U\} \text{ if } Z \in Q. \]

Observational data:

Let the number of states in \( W \) be \( p \), and let the number of states in \( U \) be \( q \).

The states of \( Z \) is the Cartesian product of the states of \( W \) and the states of \( U \).

In detail, \((w_j, u_k)\) is equivalent to \( z_{(j-1)q+k} \); \( w_j \) is equivalent to \( \vee_{k=1}^q (w_j, u_k) = \vee_{k=1}^q z_{(j-1)q+k} \), and \( u_k \) is equivalent to \( \vee_{j=1}^p (w_j, u_k) = \vee_{j=1}^p z_{(j-1)q+k} \), i.e., \( P(w_j, u_k, V) = P(z_{(j-1)q+k}, V) \) for any \( V \subseteq \{V_1, ..., V_{n-3}, X, Y\} \).

Proof. First, we show that \( Q' \) satisfies the back-door or front-door criterion relative to \( (X, Y) \) in \( G' \).

If \( Q \) satisfies the back-door criterion relative to \( (X, Y) \) in \( G \), we need to show that,

- no node in \( Q' \) is a descendant of \( X \).
- \( Q' \) blocks every path between \( X \) and \( Y \) that contains an arrow into \( X \).

It is easy to show that if there is a node in \( Q' \) that is a descendant of \( X \) in \( G' \), then there is a node in \( Q \) that is a descendant of \( X \) in \( G \). And if there is a path between \( X \) and \( Y \) that contains an arrow into \( X \) does not blocked by \( Q' \) in \( G' \), then there is a path between \( X \) and \( Y \) that contains an arrow into \( X \) does not blocked by \( Q \) in \( G \). Thus, \( Q' \) satisfies the back-door criterion relative to \( (X, Y) \) in \( G' \). Similarly, we can show that if \( Q \) satisfies the front-door criterion relative to \( (X, Y) \) in \( G \), then \( Q' \) satisfies the front-door criterion relative to \( (X, Y) \) in \( G' \).

Now, we show that \( (G, O) \) is equivalent to \( (G', O') \), i.e., show that \( P(y|do(x)) \) is the same between \( (G, O) \) and \( (G', O') \). Suppose \( Q \) satisfies the back-door criterion relative to \( (X, Y) \) in \( G \). By adjustment formula in Equation 1, we have,

\[
P(y|do(x)) = \sum_{q \in Q} P(y|x, q) \times P(q) = \sum_{q \in \tilde{Q}} P(x,y, q) \times P(q).
\]

And in \( G' \),

\[
P(y|do(x)) = \sum_{q \in \tilde{Q}'} P(y|x, q) \times P(q) = \sum_{q \in \tilde{Q}'} P(x,y, q) \times P(q),
\]

it is obviously that these two causal effects are the same, because \( P(w_j, u_k, V) = P(z_{(j-1)q+k}, V) \) for any \( V \subseteq \{V_1, ..., V_{n-3}, X, Y\} \).

Similarly, we can show that if \( Q \) satisfies the front-door criterion relative to \( (X, Y) \) in \( G \), \( (G, O) \) is equivalent to \( (G', O') \). \( \square \)

Simulation Algorithm for Generating Sample Distributions

The two sample distributions generated in the paper (in two Simulation Results sections) were generated by Algorithm 2 with \( D \) equal to the uniform distribution.
Algorithm 2: Generate-cpt()

**Input:** $n$ causal diagram nodes ($X_1, \ldots, X_n$); Distribution $D$.

**Output:** $n$ conditional probability tables for $P(X_i|\text{Parents}(X_i))$.

1: for $i$ = 1 to $n$ do
2: $s = \text{num\_instantiates}(X_i)$;
3: $p = \text{num\_instantiates}(\text{Parents}(X_i))$;
4: for $k$ = 1 to $p$ do
5: $\text{sum} = 0$;
6: for $j$ = 1 to $s$ do
7: $a_j = \text{sample}(D)$;
8: $\text{sum} = \text{sum} + a_j$;
9: end for
10: for $j$ = 1 to $s$ do
11: $P(x_i|\text{Parents}(X_i)k) = a_j/\text{sum}$;
12: end for
13: end for

**Construction of the Data in Table 4**

$P(u, w) = P(z_1)$,

$P(u, w') = P(z_2)$,

$P(u', w) = P(z_3)$,

$P(u', w') = P(z_4)$,

$P(u) = P(u, w) + P(u, w')$ = $P(z_1) + P(z_2) = 0.5$,

$P(w|u) = P(u, w)/P(u)$ = $P(z_1)/P(u) = 0.3/0.5 = 0.6$,

$P(w|u') = P(u', w)/P(u')$ = $P(z_3)/(1 - P(u)) = 0.2/0.5 = 0.4$,

$P(x|u, w) = P(x|z_1) = 0.1$,

$P(x|u, w') = P(x|z_2) = 0.4$,

$P(x|u', w) = P(x|z_3) = 0.5$,

$P(x|u', w') = P(x|z_4) = 0.7$,

$P(y|x, u, w) = P(y|x, z_1) = 0.2$,

$P(y|x, u, w') = P(y|x, z_2) = 0.3$,

$P(y|x, u', w) = P(y|x, z_2) = 0.7$,

$P(y|x, u', w') = P(y|x, z_2) = 0.1$,

$P(y|x, u', w) = P(y|x, z_3) = 0.6$,

$P(y|x', u, w) = P(y|x', z_3) = 0.5$,

$P(y|x, u', w') = P(y|x, z_4) = 0.5$,

$P(y|x', u', w') = P(y|x', z_4) = 0.4$.

**Construction of the Distribution in the Example of Dimensionality Reduction**

Here is how the data used in the example of Dimensionality Reduction were generated (both $P(X, Y, Z)$ and $P(X, Y, W), P(U)$). Instead of providing the resulting 1024 rows of the observational data, we provide the details for regenerating the observational data as following steps.

- Generate $P(X, Y, Z)$ using Algorithm 2.
- Let $P(X, Y, w_j, u_k) = P(X, Y, z_{j-1}*16+k)$.
- Let $P(X, Y, w_j) = \sum_{k=1}^{p} P(X, Y, w_j, u_k)$.
- Let $P(X, Y, u_k) = \sum_{j=1}^{q} P(X, Y, w_j, u_k)$.
- Let $P(u_k) = \sum_{X, Y} P(X, Y, u_k)$.

For example,

$P(u_1) = \sum_{X, Y} P(X, Y, u_1)$

$= \sum_{j=1}^{16} P(x, y, z_{j-1}*16+1) + \sum_{j=1}^{16} P(x, y, z_{j-1}*16+1)$

$+ \sum_{j=1}^{16} P(x', y, z_{j-1}*16+1) + \sum_{j=1}^{16} P(x', y', z_{j-1}*16+1)$

$= \sum_{j=1}^{16} P(x, y, z_{j-1}*16+1) + \sum_{j=1}^{16} P(x, y, z_{j-1}*16+1)$

$+ \sum_{j=1}^{16} P(x', y, z_{j-1}*16+1) + \sum_{j=1}^{16} P(x', y', z_{j-1}*16+1)$.

$P(x, y, w_1) = \sum_{k=1}^{16} P(x, y, w_1, u_k)$

$= \sum_{k=1}^{16} P(x, y, z_k).$