Bounds on Causal Effects and Application to High Dimensional Data

Ang Li,[1] **Judea Pearl**[1]

[1] Cognitive Systems Laboratory, Department of Computer Science,
University of California, Los Angeles,
Los Angeles, California, USA.
{angli, 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

$$
\begin{aligned}
P(y|do(x)) &= \sum_{w,u} P(y|x,w,u)P(w,u) \\
&= \sum_{w,u} \frac{P(x,y,w,u)P(w,u)}{P(x,w,u)}.
\end{aligned}
$$

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.
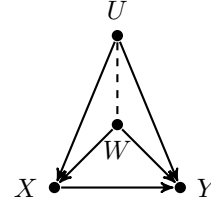
Figure 1: Needed the causal effects of X on Y when U is unobserved. The dot line between U and W means either U affects W, W affects U, or U and W are independent.

Bounding has been proven to be useful in causal inference. (Balke and Pearl 1997) provided bounds on causal effects with imperfect compliance, (Tian and Pearl 2000) proposed 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 subpopulation.

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
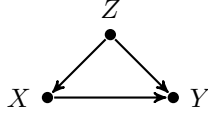
Figure 2: Needed the causal effects of X on Y when Z has high dimensionality.
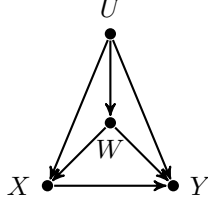

Figure 3: Causal diagram of an equivalent problem.

use the causal diagrams in (Pearl 1995; Spirtes et al. 2000; Pearl 2009; Koller and Friedman 2009).

One key concept of a causal diagram is called d-separation (Pearl 2014).

Definition 1 (d-separation)**.** *In a causal diagram G, a path p is blocked by a set of nodes Z if and only if*

1. *p contains a chain of nodes $A \rightarrow B \rightarrow C$ or a fork $A \leftarrow B \rightarrow C$ such that the middle node B is in Z (i.e., B is conditioned on), or*
2. *p contains a collider $A \rightarrow B \leftarrow C$ such that the collision node B is not in Z, and no descendant of B is in Z.*

If Z blocks every path between two nodes X and Y, then X and Y are d-separated conditional on Z, and thus are independent conditional on Z, denoted as $X \perp\!\!\!\perp Y \mid Z$.

With the concept of d-separation in a causal diagram, Pearl proposed the back-door and front-door criteria as follows:

Definition 2 (Back-Door Criterion)**.** *Given an ordered pair of variables (X, Y) in a directed acyclic graph G, a set of variables Z satisfies the back-door criterion relative to (X, Y), if no node in Z is a descendant of X, and Z blocks every path between X and Y that contains an arrow into X.*

If a set of variables Z satisfies the back-door criterion for X and Y, the causal effects of X on Y are given by the adjustment formula:

$$P(y|do(x)) = \sum_z P(y|x,z)P(z). \tag{1}$$

Definition 3 (Front-Door Criterion)**.** *A set of variables Z is said to satisfy the front-door criterion relative to an ordered pair of variables (X, Y) if*

- *Z intercepts all directed paths from X to Y;*
- *there is no back-door path from X to Z; and*
- *all back-door paths from Z to Y are blocked by X.*

If a set of variables Z satisfies the front-door criterion for X and Y, and if $P(x, Z) > 0$, then the causal effects of X on Y are given by the adjustment formula:

$$P(y|do(x)) = \sum_z P(z|x) \sum_{x'} P(y|x',z)P(x'). \tag{2}$$

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.

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

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 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.

$$LB = \min \sum_{w,u} \frac{a_{w,u}b_{w,u}}{c_{w,u}}, \tag{4}$$

$$UB = \max \sum_{w,u} \frac{a_{w,u}b_{w,u}}{c_{w,u}}, \tag{5}$$

where,

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

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

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.

$$LB = \min \sum_{w,u} \frac{b_{x,w,u}}{P(x)} \sum_{x'} \frac{a_{x',w,u}P(x')}{b_{x',w,u}}, \tag{6}$$

$$UB = \max \sum_{w,u} \frac{b_{x,w,u}}{P(x)} \sum_{x'} \frac{a_{x',w,u}P(x')}{b_{x',w,u}}, \tag{7}$$

where,

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

for all $x \in X$ and $w \in W$;

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

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

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

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

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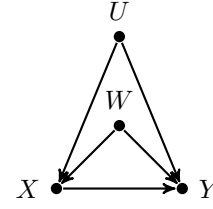
$\min\{P(x,w), p(u)\} \geq b_{x,w,u}$.

Figure 4: Needed the causal effects of X on Y when U is unobserved and independent with W.

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[2].

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. Gender is not the only confounder of taking the drug and recovery. Another confounder is family history. The manufacturers have no data associated with family history. They only know that 85.43% of people in their region have family history.

Because both gender and family history are confounders of taking the drug and recovery, and the observational data associated with family history 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 male patient, and $W = w'$ represent a female patient. Let $U = u$ represent a patient with family history, and $U = u'$ represent a patient without family history. 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

---

[2]See <http://ftp.cs.ucla.edu/pub/stat_ser/r511.pdf> for the appendix.

	Drug	No Drug
Men	23 out of 36 recovered (63.9%)	145 out of 225 recovered (64.4%)
Women	135 out of 156 recovered (86.5%)	152 out of 283 recovered (53.7%)
Overall	158 out of 192 recovered (82.3%)	297 out of 508 recovered (58.5%)

Table 1: Results of an observational study considering gender.

and the observational data from Table 1, where

$$P(x,y) = \sum_w P(y|x,w)P(x|w)P(w)$$
$$= 0.2257,$$
$$1 - P(x,y') = 1 - \sum_w P(y'|x,w)P(x|w)P(w)$$
$$= 0.9514.$$

Therefore, the bounds on the causal effect estimated using Equation 3 are $0.2257 \leq P(y|do(x)) \leq 0.9514$, where the causal information of the covariate W and the prior information $P(U)$ are not used. These bounds are not sufficiently informative to conclude the actual causal effect. Although one may believe that we can use the midpoint of the bounds (i.e., 0.5886), the gap (i.e., $0.9514 - 0.2257 = 0.7257$) between the bounds is not small; hence, this point estimate is unconvincing.

Now, considering the proposed bounds in Theorem 4 with the observational data from Table 1. $W \cup U$ satisfies the back-door criterion, and $P(X,Y,W)$ and $P(U)$ are available. We have 12 optimal variables in each objective function, because W and U are binary. With the help of the "SLSQP" solver (Kraft 1988) in the scipy package (SciPyCommunity 2020), we obtain the bounds on the causal effect, which are $0.4728 \leq P(y|do(x)) \leq 0.9514$. The lower bound actually increased significantly, and reached close to 0.5, which can help make decisions. The midpoint is 0.7121. Our conclusion is then that the causal effect of recovery when taking the drug is 0.7121. We show in the following section that this estimate of the causal effect is extremely close to the actual causal effect.

Informer View of the Causal Effect

An informer with access to the fully observed observational data, as summarized in Table 2 (Note that although it can be verified that the data in Table 2 are compatible with those in Table 1, we will never know these numbers in reality), would easily calculate the causal effect of recovery when taking the drug using the adjustment formula in Equation 1 (shown in Equation 8). The error of the estimate of the causal effect using Theorem 4 is only $(0.7518 - 0.7121)/0.7518 \approx 5.28\%$.

$$P(y|do(x)) = \sum_{w,u} P(y|x,w,u)P(w,u) = 0.7518 \quad (8)$$

	Drug	No Drug
Men with family history	3 out of 4 recovered (75.0%)	141 out of 219 recovered (64.4%)
Men without family history	20 out of 32 recovered (62.5%)	4 out of 6 recovered (66.7%)
Women with family history	135 out of 151 recovered (89.4%)	117 out of 224 recovered (52.2%)
Women without family history	0 out of 5 recovered (0.0%)	35 out of 59 recovered (59.3%)
Overall	158 out of 192 recovered (82.3%)	297 out of 508 recovered (58.5%)

Table 2: Informer view of the observational data considering gender and family history.

Simulation Results

Here, we further illustrate that the midpoints of the proposed bounds on causal effects are sufficient for estimating the causal effects, and the midpoints of the proposed bounds in Theorem 4 are better than the midpoints of the Tian-Pearl bounds in Equation 3 based on a random simulation.

We employ the simplest causal diagram in Figure 1 with binary W, U, such that $W \cup U$ satisfies the back-door criterion. We randomly generated 1000 sample distributions compatible with the causal diagram (the algorithm for generating the sample distributions is shown in the appendix[2]). The average gap (upper bound $-$ lower bound) of the Tian-Pearl bounds among 1000 samples is 0.487, and the average gap of the proposed bounds among 1000 samples is 0.383. We then randomly picked 100 out of 1000 sample distributions to draw the graph of the actual causal effects, the midpoints of the Tian-Pearl bounds, and the midpoints of the proposed bounds. The results are shown in Figure 5.

From Figure 5, although both midpoints of the bounds on the causal effects are good estimates of the actual causal effects, the midpoints of the proposed bounds are much closer to the actual causal effects, 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.

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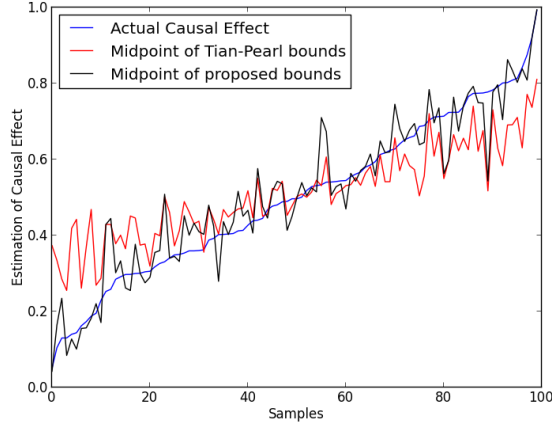
Figure 5: Bounds on causal effects of 100 sample distributions with partially observed confounders, where the Tian-Pearl bounds are obtained through Equation 3 and the proposed bounds are obtained through Theorem 4.

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[2]),

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') respect to $P(y|do(x))$ (G' containing nodes $\{V_1, ..., V_{n-3}, X, Y, W, U\}$; 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 $Parents_G(H)$ be the parents of H in causal diagram G.
$Parents_{G'}(U) = Parents_G(Z),$

$P(z_1)$	0.3		$P(y\|x, z_1)$	0.2
$P(z_2)$	0.2		$P(y\|x', z_1)$	0.3
$P(z_3)$	0.2		$P(y\|x, z_2)$	0.7
$P(z_4)$	0.3		$P(y\|x', z_2)$	0.1
$P(x\|z_1)$	0.1		$P(y\|x, z_3)$	0.6
$P(x\|z_2)$	0.4		$P(y\|x', z_3)$	0.5
$P(x\|z_3)$	0.5		$P(y\|x, z_4)$	0.5
$P(x\|z_4)$	0.7		$P(y\|x', z_4)$	0.4

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

$Parents_{G'}(W) = Parents_G(Z) \cup \{U\},$
$Parents_{G'}(H) = Parents_G(H)$ *if* $Z \notin Parents_G(H)$
for $H \in \{V_1, ..., V_{n-3}, X, Y\}$,
$Parents_{G'}(H) = Parents_G(H) \setminus \{Z\} \cup \{W, U\}$ *if* $Z \in Parents_G(H)$ *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
$Q' = Q$ *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 $\vee_{k=1}^{q}(w_j, u_k) = \vee_{k=1}^{q} z_{(j-1)*q+k}$, 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 $V \subseteq \{V_1, ..., V_{n-3}, X, Y\}$.

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[2] 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.

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'')).

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

$P(u)$	0.5

$P(w\|u)$	0.6
$P(w\|u')$	0.4

$P(x\|u,w)$	0.1
$P(x\|u,w')$	0.4
$P(x\|u',w)$	0.5
$P(x\|u',w')$	0.7

$P(y\|x,u,w)$	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

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

assigning the meaning of the states for W, U, the corresponding observatioal 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 (p in Algorithm 1), we can control the dimensionality of the new problem.

Note that, if $Z = (Z_1, Z_2, ..., Z_m)$ in G is a set of variables, we can repeat Algorithm 1 for each variable in Z, and finally obtain $W = (W_1, W_2, ..., W_m)$ and $U = (U_1, U_2, ..., U_m)$, 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[2]).

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 $z_{(j-1)*16+k}$). 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[2]), and based on Theorem 4, with the "SLSQP" solver, we obtain the bounds on the causal effect $p(y|do(x))$, which are $0.4595 \le P(y|do(x)) \le 0.7012$. We see the midpoint, 0.5804, is extremely close to the actual causal effect, 0.5527.

Finally, lets 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, we then only

---

Algorithm 1: Generate Equivalent Tuple

Input: A n nodes, $(X_1, X_2, ..., X_{n-3}, X, Y, Z)$, 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, $(X_1, X_2, ..., X_{n-3}, X, Y, W, U)$, causal diagram G';
Maping relation M_1 : state of W $\rightarrow$ state of Z;
Maping relation M_2 : state of U $\rightarrow$ state of Z.

1: $m = num\_states\_in\_G(Z)$;
2: **if** $m \mod p == 0$ **then**
3: $\quad q = m/p$;
4: **else**
5: $\quad q = m/p + 1$;
6: **end if**
7: // Set the virtual states for Z s.t. the probability is 0.
8: $num\_states\_in\_G(Z) = p \times q$;
9: **for** H in $\{X_1, ..., X_{n-3}, X, Y\}$ **do**
10: $\quad num\_states\_in\_G'(H) = num\_states\_in\_G(H)$;
11: $\quad$ **if** $Z \in Parents\_in\_G(H)$ **then**
12: $\quad\quad Parents\_in\_G'(H) = Parents\_in\_G(H)\backslash\{Z\}\cup \{W, U\}$;
13: $\quad$ **else**
14: $\quad\quad Parents\_in\_G'(H) = Parents\_in\_G(H)$;
15: $\quad$ **end if**
16: **end for**
17: $num\_states\_in\_G'(W) = p$;
18: $num\_states\_in\_G'(U) = q$;
19: $Parents\_in\_G'(W) = Parents\_in\_G(Z) \cup \{U\}$;
20: $Parents\_in\_G'(U) = Parents\_in\_G(Z)$;
21: **for** $i = 1$ to p **do**
22: $\quad M_1(w_i) = \vee_{k=1}^{q} z_{(i-1)*q+k}$;
23: **end for**
24: **for** $i = 1$ to q **do**
25: $\quad M_2(u_i) = \vee_{j=1}^{p} z_{(j-1)*q+i}$;
26: **end for**

---

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[2]). The average gap (upper bound − lower bound) of the Tian-Pearl bounds among 100 samples is 0.5102, and the average gap of the proposed bounds through Theorems 7 and 4 among 100 samples is 0.0676. We then draw the graph of the actual causal effects, the midpoints of the Tian-Pearl bounds, and the midpoints of the proposed bounds through Theorems 7 and 4. The results are shown in Figure 6.

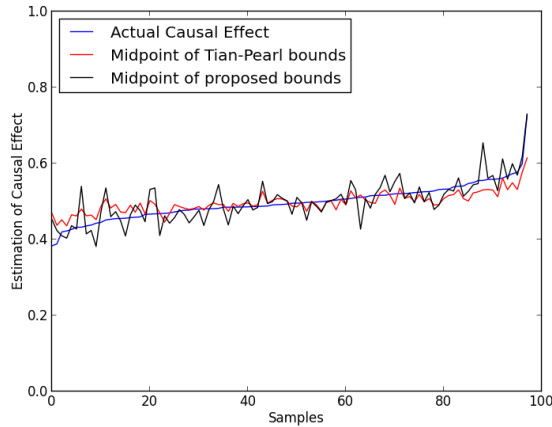From Figure 6, both midpoints of the bounds on the

Figure 6: Bounds on causal effects of 100 sample distributions with high dimensional data, where the Tian-Pearl bounds are obtained through Equation 3 and the proposed bounds are obtained through Theorems 7 and 4.

causal effects are good estimates of the actual causal effects, whereas the midpoints of the proposed bounds are slightly closer to the actual causal effects, particularly when the causal effects are close to 0 and 1. Although the trend of the Tian-Pearl bounds is also close to the actual causal effects, the Tian-Pearl bounds are more likely to be parallel with the x-axis. Here, the Tian-Pearl bounds perform well because, in high-dimensionality cases, the randomly generated distributions are more likely to yield causal effects of approximately 0.5. However, the average gap of the proposed bounds among 100 samples, 0.0676, is much smaller than the average gap of the Tian-Pearl bounds among 100 samples, 0.5102. This means that the midpoints of the proposed bounds are more convincing, because the bounds are narrower.

Discussion

Here, we discuss additional features of bounds on causal effects.

First, if a whole set of back-door or front-door variables are unobserved, the causal effects have the naivest bounds in Equation 3. When the back-door or front-door variables are gradually observed, the bounds of the causal effects become increasingly narrow. Finally, when the back-door or front-door variables are fully observed, the bounds shrink into point estimates, which are identifiable. This also tells us that, when we pick p in Algorithm 1, we should pick the largest p for which the sample size is sufficient to estimate the observational distributions.

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 1995), 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

References

Balke, A.; and Pearl, J. 1997. Bounds on treatment effects from studies with imperfect compliance. *Journal of the American Statistical Association*, 92(439): 1171–1176.

Balke, A. A. 1995. *Probabilistic counterfactuals: semantics, computation, and applications*. University of California, Los Angeles.

Cai, Z.; Kuroki, M.; Pearl, J.; and Tian, J. 2008. Bounds on Direct Effect in the Presence of Confounded Intermediate Variables. *Biometrics*, 64: 695–701.

Koller, D.; and Friedman, N. 2009. *Probabilistic graphical models: Principles and techniques*. MIT press.

Kraft, D. 1988. *A software package for sequential quadratic programming*. Deutsche Forschungs- und Versuchsanstalt für Luft- und Raumfahrt Köln: Forschungsbericht. Wiss. Berichtswesen d. DFVLR.

Li, A.; and Pearl, J. 2019. Unit selection based on counterfactual logic. In *Proceedings of the 28th International Joint Conference on Artificial Intelligence*, 1793–1799. AAAI Press.

Maathuis, M. H.; Kalisch, M.; Bühlmann, P.; et al. 2009. Estimating high-dimensional intervention effects from observational data. *The Annals of Statistics*, 37(6A): 3133–3164.

Pearl, J. 1995. Causal diagrams for empirical research. *Biometrika*, 82(4): 669–688.

Pearl, J. 2009. *Causality*. Cambridge university press, 2nd edition.

Pearl, J. 2014. *Probabilistic reasoning in intelligent systems: Networks of plausible inference*. Morgan Kaufmann.

Roscoe, J. T. 1975. *Fundamental research statistics for the behavioral sciences*. Number v. 2 in Editors' Series in Marketing. Holt, Rinehart and Winston. ISBN 9780030919343.

SciPyCommunity. 2020. Scipy Reference Guide.

Spirtes, P.; Glymour, C. N.; Scheines, R.; and Heckerman, D. 2000. *Causation, prediction, and search*. MIT press.

Tian, J.; and Pearl, J. 2000. Probabilities of causation: Bounds and identification. *Annals of Mathematics and Artificial Intelligence*, 28(1-4): 287–313.