

Unit Selection with Causal Diagram

Ang Li,¹ Judea Pearl,¹

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

Abstract

The unit selection problem aims to identify a set of individuals who are most likely to exhibit a desired mode of behavior, for example, selecting individuals who would respond one way if encouraged and a different way if not encouraged. Using a combination of experimental and observational data, Li and Pearl derived tight bounds on the “benefit function” - the payoff/cost associated with selecting an individual with given characteristics. This paper shows that these bounds can be narrowed significantly (enough to change decisions) when structural information is available in the form of a causal model. We address the problem of estimating the benefit function using observational and experimental data when specific graphical criteria are assumed to hold.

Introduction

In many areas of industry, marketing, and health science, the unit selection dilemma arises. For example, in customer relationship management (Berson, Smith, and Thearling 1999; Lejeune 2001; Hung, Yen, and Wang 2006; Tsai and Lu 2009), it is useful to know which customers are going to churn but might reconsider if encouraged to stay. Due to the high expense of such initiatives, management is forced to limit inducement to customers who are most likely to exhibit the behavior of interest. As another example, companies are interested in identifying users who would click on an advertisement if and only if it is highlighted in online advertising (Yan et al. 2009; Bottou et al. 2013; Li et al. 2014; Sun et al. 2015). The challenge in identifying these users stems from the fact that the desired response pattern is not observed directly but rather is defined counterfactually in terms of what the individual would do under hypothetical unrealized conditions. For example, when we observe that a user has clicked on a highlighted advertisement, we do not know whether they would click on that same advertisement if it were not highlighted.

The benefit function for the unit selection problem was defined by Li and Pearl (Li and Pearl 2019), and it properly captures the nature of the desired behavior. Using a combination of experimental and observational data, Li and Pearl derived tight bounds of the benefit function. The only assumption is that the treatment has no effect on the population specific characteristics. However, Li-Pearl’s derivation does not leverage information from auxiliary covariates, if such is

available. Mueller, Li, and Pearl (Mueller, Li, and Pearl 2021) recently proposed using covariate information and the causal structure to narrow the bounds of probability of necessity and sufficiency. Dawid et al. (Dawid, Musio, and Murtas 2017) also proposed using covariates information to narrow the bounds of probability of necessity. A similar approach might be used for the benefit function. Most crucially, the information provided by covariates and their causal structure may result in a reversal of decision (relative to not considering such covariates).

Consider the following motivating scenario: a carwash company wants to offer a discount to employees of company A. The offer can only be presented to the entire company A; the carwash company will not be able to provide a discount to a specific group inside the company A. The carwash company’s manager seeks to maximize total profit, including nonimmediate profit. The management estimates that the benefit of selecting a complier (i.e., offer the discount to a customer who would use the carwash service if they received the discount, but would not otherwise) is \$100 as the profit is \$140 but the discount is \$40, that of selecting an always-taker (i.e., offer the discount to a customer who would use the carwash service regardless of whether they received the discount) is $-\$60$ as the customer would use the service anyway (so the company loses the value of the discount and an extra cost of \$20 because the always-taker may require additional discounts in the future), that of selecting a never-taker (i.e., offer the discount to a customer who would never use the carwash service regardless of whether they received the discount) is \$0 as the cost of issuing the discount is negligible, and that of selecting a defier (i.e., offer the discount to a customer who would not use the carwash service if they received the discount, but would use the carwash service otherwise) is $-\$140$ as the customer is lost due to the discount. The manager of carwash company has both experimental and observational data related to customer age collected from the company A. If the entire company A’s employees are given the discount, the manager of carwash wants to know what the average profit will be.

Based on Li-Pearl’s model, it is easy to see that the benefit vector for the aforementioned example is $(100, -60, 0, -140)$, and a corresponding benefit function can be defined as the objective function. Li-Pearl’s model can then obtain the bounds of the benefit function using experi-

mental and observational data. The model, however, does not take into account the covariate information (customer age) and the causal structure.

In this paper, we show how the information included in such covariates and their causal structure, can be used to narrow the bounds of the benefit function in Li-Pearl’s model. Most importantly, the narrower bounds can, sometimes, flip the decision.

Preliminaries

In this section, we review Li and Pearl’s benefit function of the unit selection problem (Li and Pearl 2019). Individual behavior was classified into four response types: labeled complier, always-taker, never-taker, and defier. Suppose the benefit of selecting one individual in each category are $\beta, \gamma, \theta, \delta$ respectively (i.e., the benefit vector is $(\beta, \gamma, \theta, \delta)$). They defined the objective function of the unit selection problem as the average benefit gained per individual. Suppose a and a' are binary treatments, r and r' are binary outcomes, and c are population-specific characteristics, the objective function (i.e., benefit function) is following (If the goal is to evaluate the average benefit gained per individual for a specific population c , argmax_c can be dropped.):

$$\text{argmax}_c \beta P(r_a, r'_a | c) + \gamma P(r_a, r_{a'} | c) + \theta P(r'_a, r'_a | c) + \delta P(r'_a, r_{a'} | c).$$

Using a combination of experimental and observational data, Li and Pearl established the most general tight bounds on this benefit function (which we refer to as Li-Pearl’s Theorem in the rest of the paper). The only constraint is that the population-specific characteristics are not a descendant of the treatment.

However, the information of covariates (if available, such as the age in the motivating example in the previous section) are not considered. In this paper, we present three common cases of covariates and their causal structures and theorems that show how the information about the covariates along with their causal structures could narrow the bounds of the benefit function. The improvement of the bounds is sometimes significant and can change the decisions compared to Li-Pearl’s Theorem.

Selection Criteria with Causal Diagrams

We present three common cases of the covariates and their causal structures in this section. For each case, we provide a theorem for estimating the benefit function in such a case. The proof of all theorems is in the appendix. In any causal diagram of this paper, the dot line between A and B represents either A affects B , B affects A , or A and B are independent; the dot line with arrow from A to B represents either A affects B or A and B are independent.

Causal Diagram with Non-descendant Covariates

Theorem 1 provides bounds for the benefit function when a set Z of variables can be measured, which satisfies only one condition: both population-specific variables C and covariates Z contain no descendant of X . This condition is important because if X is set to x and $C \cup Z$ contains a

descendant of X , then $C \cup Z$ could be altered and $P(y_x | z, c)$ would be another unmeasurable counterfactual term. If the descendant is independent of Y_x , then $P(y_x | z, c)$ would be measurable, but the descendant would not contribute to any narrowing of the bounds. These bounds are always contained within the bounds of the benefit function in Li-Pearl’s Theorem.

Theorem 1. *Given a causal diagram G and distribution compatible with G , let $Z \cup C$ be a set of variables that does not contain any descendant of X in G , then the benefit function $f(c) = \beta P(y_x, y'_{x'} | c) + \gamma P(y_x, y_x | c) + \theta P(y'_{x'}, y'_{x'} | c) + \delta P(y_{x'}, y_x | c)$ is bounded as follows:*

$$\begin{aligned} W + \sigma U &\leq f(c) \leq W + \sigma L && \text{if } \sigma < 0, \\ W + \sigma L &\leq f(c) \leq W + \sigma U && \text{if } \sigma > 0, \end{aligned}$$

where σ, W, L, U are given by,

$$\begin{aligned} \sigma &= \beta - \gamma - \theta + \delta, \\ W &= (\gamma - \delta)P(y_x | c) + \delta P(y_{x'} | c) + \theta P(y'_{x'} | c), \\ L &= \sum_z \max \left\{ \begin{array}{l} 0, \\ P(y_x | z, c) - P(y_{x'} | z, c), \\ P(y | z, c) - P(y_{x'} | z, c), \\ P(y_x | z, c) - P(y | z, c) \end{array} \right\} \\ &\quad \times P(z | c), \\ U &= \sum_z \min \left\{ \begin{array}{l} P(y_x | z, c), \\ P(y'_{x'} | z, c), \\ P(y, x | z, c) + P(y', x' | z, c), \\ P(y_x | z, c) - P(y_{x'} | z, c) + \\ + P(y, x' | z, c) + P(y', x | z, c) \end{array} \right\} \\ &\quad \times P(z | c). \end{aligned}$$

Notably, C can be interpreted as the population-specific variables, and Z are the attributes in each population. Moreover, the bounds provided above are always no worse than Li-Pearl’s bound (see proof in the appendix). Besides, if $\sigma = 0$, the Gain Equality is satisfied in Li-Pearl’s model, and the result of the benefit function is no longer bounds, but a point estimate.

Causal Diagram with Mediators

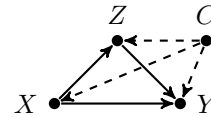


Figure 1: Mediator Z with direct effects of X on Y .

Partial Mediators In Figure 1, partial mediator Z is a descendant of X ; thus, we cannot use Theorem 1. However, the absence of confounders (other than population specific variables C) between Z and Y and between X and Y permits us to bound the benefit function as follows:

Theorem 2. *Given a causal diagram G and distribution compatible with G , let Z be a set of variables such that $\forall x, x' \in X : x \neq x', (Y_x \perp\!\!\!\perp X \cup Z_{x'} \mid Z_x, C)$ in G , and*

C does not contain any descendant of X in G , then the benefit function $f(c) = \beta P(y_x, y_{x'}|c) + \gamma P(y_x, y_{x'}|c) + \theta P(y_{x'}, y_{x'}|c) + \delta P(y_{x'}, y_{x'}|c)$ is bounded as follows:

$$\begin{aligned} W + \sigma U &\leq f(c) \leq W + \sigma L && \text{if } \sigma < 0, \\ W + \sigma L &\leq f(c) \leq W + \sigma U && \text{if } \sigma > 0, \end{aligned}$$

where σ, W, L, U are given by,

$$\begin{aligned} \sigma &= \beta - \gamma - \theta + \delta, \\ W &= (\gamma - \delta)P(y_x|c) + \delta P(y_{x'}|c) + \theta P(y_{x'}|c), \\ L &= \max \left\{ \begin{array}{l} 0, \\ P(y_x|c) - P(y_{x'}|c), \\ P(y|c) - P(y_{x'}|c), \\ P(y_x|c) - P(y|c) \end{array} \right\}, \\ U &= \min \left\{ \begin{array}{l} P(y_x|c), \\ P(y_{x'}|c), \\ P(y, x|c) + P(y', x'|c), \\ P(y_x|c) - P(y_{x'}|c) + \\ + P(y, x'|c) + P(y', x|c), \\ \sum_z \sum_{z' \neq z} \min\{P(y|z, x, c), \\ P(y'|z', x', c)\} \times \\ \min\{P(z_x|c), P(z'_{x'}|c)\} \end{array} \right\}. \end{aligned}$$

Although this lower bound is unchanged from that in Li-Pearl's Theorem, the upper bound contains a vital additional argument (i.e., the last term in the min function of U) to the min function. This new term can significantly reduce the upper bound. The rest of the terms are included because sometimes the bounds of Li-Pearl's Theorem are superior. The following theorem has the same quality.

Pure Mediators Figure 2 is a special case of Figure 1, in which X has no direct effects on Y . The resulting bounds for the benefit function are as follows:

Theorem 3. Given a causal diagram G in Figure 2 and distribution compatible with G , and C does not contain any descendant of X , then the benefit function $f(c) = \beta P(y_x, y_{x'}|c) + \gamma P(y_x, y_{x'}|c) + \theta P(y_{x'}, y_{x'}|c) + \delta P(y_{x'}, y_{x'}|c)$ is bounded as follows:

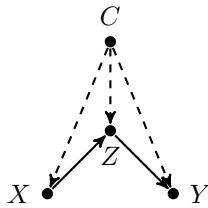


Figure 2: Mediator Z with no direct effects of X on Y .

$$\begin{aligned} W + \sigma U &\leq f(c) \leq W + \sigma L && \text{if } \sigma < 0, \\ W + \sigma L &\leq f(c) \leq W + \sigma U && \text{if } \sigma > 0, \end{aligned}$$

where σ, W, L, U are given by,

$$\begin{aligned} \sigma &= \beta - \gamma - \theta + \delta, \\ W &= (\gamma - \delta)P(y_x|c) + \delta P(y_{x'}|c) + \theta P(y_{x'}|c), \\ L &= \max \left\{ \begin{array}{l} 0, \\ P(y_x|c) - P(y_{x'}|c), \\ P(y|c) - P(y_{x'}|c), \\ P(y_x|c) - P(y|c) \end{array} \right\}, \\ U &= \min \left\{ \begin{array}{l} P(y_x|c), \\ P(y_{x'}|c), \\ P(y, x|c) + P(y', x'|c), \\ P(y_x|c) - P(y_{x'}|c) + \\ + P(y, x'|c) + P(y', x|c), \\ \sum_z \sum_{z' \neq z} \min\{P(y|z, c), \\ P(y'|z', c)\} \times \\ \min\{P(z|x, c), P(z'|x', c)\} \end{array} \right\}. \end{aligned}$$

The core term (i.e., the last term in the min function of U) for Theorem 3 added to the upper bound notably only requires observational data.

Examples

In this section, we will show how the presented theorems can be applied to applications and how the theorems affect judgments using two cases.

Company Selection

Consider the motivating example in the introduction section.

Let $A = a$ denote the event that a customer receives the discount, $A = a'$ denote the event that a customer does not receive the discount, $R = r$ denote the event that a customer uses the services, $R = r'$ denote the event that a customer does not use the services, $C = c$ denote a company A's customer, $Z = z$ denote a younger customer (age below or equal to 50), and $Z = z'$ denote an older customer (age above 50). The model is as shown in Figure 3.

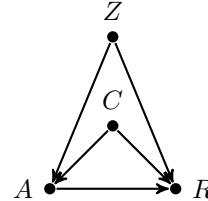


Figure 3: Company selection model.

Based on Li-Pearl's model, it is easy to see that the benefit vector is $(100, -60, 0, -140)$ (see the introduction section). Therefore, the benefit function is:

$$\begin{aligned} \operatorname{argmax}_c & 100P(r_a, r'_{a'}|c) - 60P(r_a, r_{a'}|c) + \\ & + 0P(r'_{a'}, r'_{a'}|c) - 140P(r'_{a'}, r_{a'}|c). \end{aligned} \quad (1)$$

The manager of the carwash company collected the data listed in Tables 1 and 2 from company A. By Li-Pearl's Theorem, the bounds of the benefit function are $[-0.423, 2.832]$ (see the appendix for details), and the midpoint is 1.205. It suggests that the carwash company would gain \$1.205 profit

Table 1: Experimental data collected by the carwash company. 350 customers were forced to receive the discount and 350 customers were forced not to receive the discount.

| | Discount | No Discount |
|---------|---|---|
| Young | 45 out of 101 used the service (44.6%) | 5 out of 101 used the service (5.0%) |
| Elder | 248 out of 249 used the service (99.6%) | 179 out of 249 used the service (71.9%) |
| Overall | 293 out of 350 used the service (83.7%) | 184 out of 350 used the service (52.6%) |

Table 2: Observational data collected by the carwash company. 700 customers were given access to the discount, they can choose whether to obtain the discount by themselves (note that a customer may still not use the service even they obtained the discount by themselves).

| | Discount | No Discount |
|---------|---|---|
| Young | 90 out of 152 used the service (59.2%) | 9 out of 50 used the service (18.0%) |
| Elder | 157 out of 159 used the service (98.7%) | 239 out of 339 used the service (70.5%) |
| Overall | 247 out of 311 used the service (79.4%) | 248 out of 389 used the service (63.8%) |

from each individual from company A if they offer company A's employees the discount. Besides, most of the bounded area is positive, which provided more confidence that the conclusion is correct. However, Li-Pearl's theorem only uses the overall data in Tables 1 and 2 (i.e., customer age is not considered).

Now, if we apply Theorem 1 to the data in Tables 1 and 2, the bounds of the benefit function is $[-0.168, -0.077]$ (see the appendix for details), with the midpoint at -0.123 . This suggests that if the carwash company offers the discount to company A's employees, the carwash company will lose \$0.123 profit per individual. Notably, the upper bound (-0.077) is negative, implying that the carwash company must lose profit if they offers the discount to company A's employees regardless of how the bounds are used.

Effective Patients of a Drug

When a pharmaceutical company develops a new drug, it seeks to identify patients so as to maximize the difference between the number of effective patients and the number of ineffective patients. The causal diagram is shown in Figure 4.

For the benefit vector, the pharmaceutical company assigned 1 to a complier because the complier is the patient cured by the drug, assign -1 to an always-taker, a never-taker, and a defier because they are all ineffective patients. The benefit vector is then $(1, -1, -1, -1)$.

Table 3: Results of an observational study (30 years old male) into a new drug, with post-treatment blood pressure taken into account.

| | Drug | No Drug |
|---------------|----------------------------------|----------------------------------|
| Low BP | 375 out of 405 recovered (92.6%) | 159 out of 481 recovered (33.1%) |
| High BP | 17 out of 183 recovered (9.3%) | 3 out of 6 recovered (50.0%) |
| Combined data | 392 out of 588 recovered (66.7%) | 162 out of 487 recovered (33.3%) |

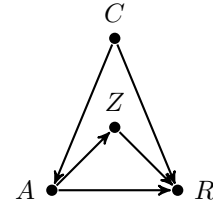


Figure 4: A graphical model representing the effects of a new drug, with A representing drug usage, R representing recovery, Z representing blood pressure (measured at the end of the study), and C representing the population specific variables (gender and age).

Let $A = a$ denote the event that a patient takes the drug, $A = a'$ denote the event that a patient does not take the drug, $R = r$ denote the event that a patient is recovered, $R = r'$ denote the event that a patient is not recovered, $Z = z$ denote low blood pressure (measured at the end of the study), $Z = z'$ denote high blood pressure, and C (a set of variables) denote the population-specific characteristics (gender and age) of a patient. The benefit function is then

$$\begin{aligned} \text{argmax}_c & P(r_a, r'_a | c) - P(r_a, r'_a | c) - \\ & - P(r'_a, r'_a | c) - P(r'_a, r'_a | c). \end{aligned} \quad (2)$$

The pharmaceutical company records the recovery rates of 70000 patients who were given access to the drug (i.e., observational study). For each group of patients who have the same gender and age, they record the number of patients who chose to take the drug and their recovery rates, the number of patients who did not choose to take the drug, and their recovery rates. For example, the results of the 30 years old male patients (1075 patients) are shown in Table 3.

Note that the data in Table 3 is observational data. The experimental data is not available yet. However, the set $\{C\}$ satisfied the back-door criterion for both (A, Z) and (A, R) (Pearl 1995). By Pearl's adjustment formula, the experimental data needed are: $P(r_a | c) = P(r | a, c) = 0.6667$, $P(r'_a | c) = P(r | a', c) = 0.3326$, $P(z_a | c) = P(z | a, c) = 0.6888$, and $P(z'_a | c) = P(z | a', c) = 0.0123$.

First, we apply Li-Pearl's Theorem to the combined data in Table 3 and the above experimental data, the bounds of the benefit function are $[-0.3320, 0.3333]$ (see the appendix for

details), and the midpoint is 0.0007. It suggests that the drug should apply to the 30 years old male because the difference between the number of effective patients and the number of ineffective patients per 30 years old male is positive. Or someone may say that it is hard to decide because the bounded area is roughly half positive and half negative.

Second, we apply the proposed Theorem 2 to the entire data in Table 3 and the above experimental data, the bounds of the benefit function are $[-0.3320, -0.0054]$ (see the appendix for details), and the midpoint is -0.1687 . The upper bound dropped significantly from 0.3333 to -0.0054 . It suggests that the drug should not apply to the 30 years old male, because the difference between the number of effective patients and the number of ineffective patients per 30 years old male is negative. Most importantly, the entire bounded area is negative so that the decision is convincing.

Simulated Results

In this section, we will show how much in general the bounds of the benefit function are improved by Theorems 1, 2, and 3 in three simple causal diagrams.

For each theorem, we randomly generated 100000 sample distributions (observational data and experimental data) compatible with the causal diagram (see the appendix for the generating algorithm). Each sample distribution represents a different instantiate of the population-specific characteristics C in the model. The generating algorithm ensures that the experimental data and observational data satisfy the general relation (i.e., $P(x, y|c) \leq P(y|do(x), c) \leq 1 - P(x, y'|c)$) (Tian and Pearl 2000). We set the benefit vector $(\beta, \gamma, \theta, \delta)$ to be the most common $(1, -1, -1, -1)$ to encourage compliers while avoiding always-takers, never-takers, and defiers. For the sample distribution i , let $[a_i, b_i]$ be the bounds that considered the covariates and the causal diagram from the proposed theorems and $[c_i, d_i]$ be the bounds that did not consider the covariates and the causal diagram from Li-Pearl's Theorem. We summarized the following criteria for each case:

- Average increased lower bound : $\frac{\sum(a_i - c_i)}{100000}$;
- Average decreased upper bound : $\frac{\sum(d_i - b_i)}{100000}$;
- Average gap that did not consider the covariates and the causal diagram : $\frac{\sum(d_i - c_i)}{100000}$;
- Average gap that considered the covariates and the causal diagram : $\frac{\sum(b_i - a_i)}{100000}$;
- Number of sample distributions in which the decision was flipped : $\sum e_i$ where, $e_i = 1$ if $(a_i + b_i) \times (c_i + d_i) < 0$ and $e_i = 0$ otherwise;
- Number of sample distributions in which the bounds that considered the covariates and the causal diagram from proposed Theorems were narrower : $\sum f_i$ where, $f_i = 1$ if $(a_i > c_i)$ or $(b_i < d_i)$ and $f_i = 0$ otherwise.

Non-descendant Covariates

In the case of non-descendant covariates compatible with Theorem 1. We randomly generated 100000 sample distributions compatible with the causal diagram in Figure 5.

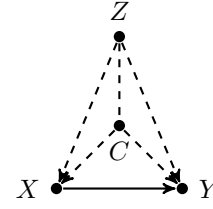


Figure 5: Causal diagram such that $C \cup Z$ is not a descendant of X .

The results between proposed Theorem 1 and Li-Pearl's Theorem are summarized in Table 4. We can see that the average gap that did not consider the covariates and the causal diagram by Li-Pearl's Theorem is 0.4342, while the average gap that considered the covariates and the causal diagram by Theorem 1 is 0.3352, and both the lower bound and upper bound are improved by roughly 0.05. The decisions flipped (i.e., the results of Li-Pearl's Theorem suggest gain profit, while the results of Theorem 1 suggest losing profit, or the reverse) is $920/100000 \approx 1\%$ of the samples, which means that at least 1% of the applications would have the wrong decision if we do not consider the covariates. The bounds that considered the covariates and the causal diagram are narrower in $93688/100000 \approx 93.7\%$ of the samples. Therefore, if a set of Z is available that satisfies Theorem 1, the bounds of the benefit function by the proposed theorem are more useful as the gap is narrower.

Table 4: Simulation results of 100000 sample distributions compatible with the causal diagram in Figure 5.

| Average increased lower bound | Average decreased upper bound | Average gap by Li-Pearl's Theorem |
|-------------------------------|-------------------------------|-----------------------------------|
| 0.0494 | 0.0496 | 0.4342 |
| Average gap by Theorem 1 | Decision flipped | Bounds narrower |
| 0.3352 | 920 | 93688 |

We then randomly picked 100 of 100000 sample distributions to draw the graph of bounds that considered and did not consider the covariates and the causal diagram (To have a better vision, we sorted the sample distributions by the general lower bound that did not considered the covariates and the causal diagram). The results are shown in Figure 6. We can see that the bounds of the benefit function are improved in most of the samples with the causal diagram.

Partial Mediators

In the case of partial mediators compatible with Theorem 2. We randomly generated 100000 sample distributions that are compatible with the causal diagram in Figure 1.

The results between the proposed Theorem 2 and Li-Pearl's Theorem are summarized in Table 5. First, the av-

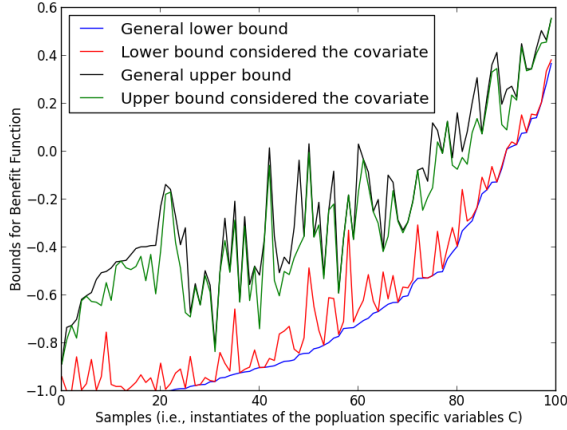


Figure 6: Bounds of the benefit function for 100 samples compatible with the causal diagram of Figure 5, where the general bounds are obtained from Li-Pearl’s Theorem and the bounds that considered the non-descendant covariate and the causal diagram are obtained from Theorem 1.

average increased lower bound is 0 because the lower bound in Theorem 2 is exactly the lower bound in Li-Pearl’s Theorem. The partial mediator cannot improve the lower bound. The average gap is also close between Li-Pearl’s Theorem and proposed Theorem 2 because the bounds of only $12724/100000 \approx 12.7\%$ of samples are narrowed by the proposed Theorem 2. 12.7% is an acceptable number if the costs for considering the partial mediators are acceptable. The actual improvement among the narrowed samples is impressive. We then randomly generated 100000 samples that the bounds are indeed narrowed by the proposed Theorem 2 (same generating algorithm, but we keep generating until we have 100000 narrowed samples). The results of the comparison between the proposed Theorem 2 and Li-Pearl’s Theorem are summarized in Table 6. We can see that the average gap that did not consider the partial mediator and the causal diagram is 0.5531, while the average gap that considered the partial mediator and the causal diagram by Theorem 2 is 0.4768, and the upper bound is improved by roughly 0.0764. Therefore, if a set of Z is available that satisfies Theorem 2 and the costs permitted, we should always consider the partial mediators and using Theorem 2.

We then randomly picked 100 of 100000 narrowed sample distributions to draw the graph of bounds that considered and did not considered the partial mediator and the causal diagram (To have a better vision, we sorted the sample distributions by the general upper bound that did not consider the partial mediator and the causal diagram). The results are shown in Figure 7. We can see that the upper bounds of the benefit function are improved significantly among these narrowed cases.

Pure Mediators

In the case of pure mediators compatible with Theorem 3. We randomly generated 100000 sample distributions compatible

Table 5: Simulation results of 100000 sample distributions compatible with the causal diagram in Figure 1.

| Average increased lower bound | Average decreased upper bound | Average gap by Li-Pearl’s Theorem |
|-------------------------------|-------------------------------|-----------------------------------|
| 0 | 0.00985 | 0.4564 |
| Average gap by Theorem 2 | Decision flipped | Bounds narrower |
| 0.4465 | 139 | 12724 |

Table 6: Simulation results of 100000 narrowed sample distributions compatible with the causal diagram in Figure 1.

| Average increased lower bound | Average decreased upper bound | Average gap by Li-Pearl’s Theorem |
|-------------------------------|-------------------------------|-----------------------------------|
| 0 | 0.0764 | 0.5531 |
| Average gap by Theorem 2 | Decision flipped | |
| 0.4768 | 1033 | |

with the causal diagram in Figure 2.

The results between the proposed Theorem 3 and Li-Pearl’s Theorem are summarized in Table 7. We can see that the average gap that did not consider the pure mediator and the causal diagram by Li-Pearl’s Theorem is 0.5195, while the average gap that considered the pure mediator and the causal diagram by Theorem 3 is 0.3324, and the upper bound is improved by roughly 0.187. The lower bound is not improved, because the lower bound in Theorem 3 is exactly the same as in Li-Pearl’s Theorem. The decisions flipped (i.e., the results of Li-Pearl’s Theorem suggest gain profit, while the results of Theorem 3 suggest losing profit, or the reverse) is $459/100000 \approx 0.46\%$ of the samples, which means that at least 0.46% of the applications would have the wrong decision if we do not consider the pure mediators. The bounds that considered the pure mediator and the causal diagram are narrower in $99996/100000 \approx 99.9\%$ of the samples. Therefore, if a set of Z is available that satisfies Theorem 3, the bounds of the benefit function by the proposed theorem is more useful as the gap is narrower.

We then randomly picked 100 of 100000 sample distributions to draw the graph of bounds that considered and did not consider the pure mediator and the causal diagram (To have a better vision, we sorted the sample distributions by the general upper bound that did not consider the pure mediator and the causal diagram). The results are shown in Figure 8. We can see that the bounds of the benefit function are improved in almost all the samples with the causal diagram.

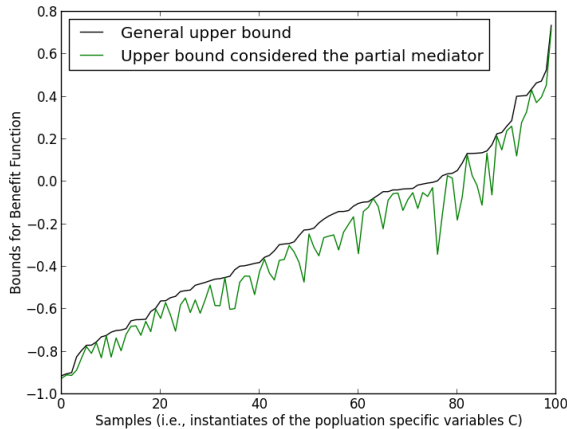


Figure 7: Upper bound of the benefit function for 100 narrowed samples compatible with the causal diagram of Figure 1, where the general upper bounds are obtained from Li-Pearl’s Theorem and the upper bounds that considered the partial mediator and the causal diagram are obtained from Theorem 2.

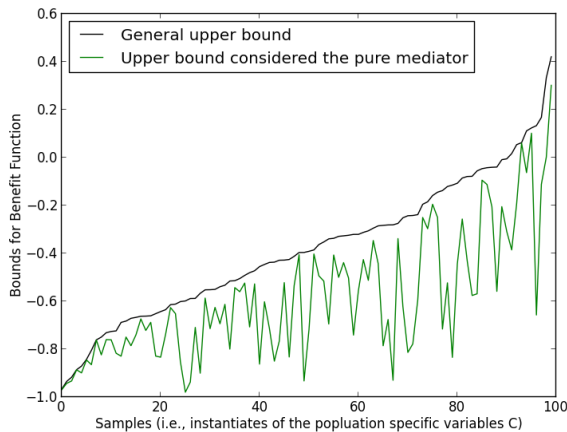


Figure 8: Bounds of the benefit function for 100 samples compatible with the causal diagram of Figure 2, where the general bounds are obtained from Li-Pearl’s Theorem and the bounds that considered the pure mediator and the causal diagram are obtained from Theorem 3.

Table 7: Simulation results of 100000 sample distributions compatible with the causal diagram in Figure 2.

| Average increased lower bound | Average decreased upper bound | Average gap by Li-Pearl’s Theorem |
|-------------------------------|-------------------------------|-----------------------------------|
| 0 | 0.1870 | 0.5195 |
| Average gap by Theorem 3 | Decision flipped | Bounds narrower |
| 0.3324 | 459 | 99996 |

Discussion

In this section, we will discuss one more requirement of covariates Z in Theorem 1. Note that in the motivating example in the introduction section, the discount should apply to the entire company A’s employees; the carwash company can only decide to offer the discount to the entire company A or not to the entire company A. The carwash company cannot offer the discount to a specific age group in company A. Otherwise, if the carwash company can offer the discount to a specific age group, the covariates Z should be considered as the population-specific characteristics and combined into C , and apply Li-Pearl’s Theorem separately to each population-specific group. This requirement is common; for example, an election speech cannot offer to only a specific group of people in a region, and an auto show cannot offer to only a specific group of customers in a region. This requirement does not apply to Theorems 2 and 3 because the mediators happen after the treatment.

Conclusion

We demonstrated how bounds of the benefit function in the unit selection problem could be narrowed if covariates information and their associated causal structures are available. We derived three theorems to narrow the bounds of the benefit function in three common graphical conditions. We illustrated that if costs are permitted, and there are covariates and causal structures available, the proposed theorems should always be applied, as narrower bounds are helping to make accurate decisions. Examples and simulation results are provided to support the proposed theorems.

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

- Berson, A.; Smith, S.; and Thearling, K. 1999. *Building data mining applications for CRM*. McGraw-Hill Professional.
- Bottou, L.; Peters, J.; Quiñero-Candela, J.; Charles, D. X.; Chickering, D. M.; Portugaly, E.; Ray, D.; Simard, P.; and

- Snelson, E. 2013. Counterfactual reasoning and learning systems: The example of computational advertising. *The Journal of Machine Learning Research*, 14(1): 3207–3260.
- Dawid, P.; Musio, M.; and Murtas, R. 2017. The Probability of Causation. *Law, Probability and Risk*, (16): 163–179.
- Hung, S.-Y.; Yen, D. C.; and Wang, H.-Y. 2006. Applying data mining to telecom churn management. *Expert Systems with Applications*, 31(3): 515–524.
- Lejeune, M. A. 2001. Measuring the impact of data mining on churn management. *Internet Research*, 11(5): 375–387.
- 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.
- Li, L.; Chen, S.; Kleban, J.; and Gupta, A. 2014. Counterfactual estimation and optimization of click metrics for search engines. *arXiv preprint arXiv:1403.1891*.
- Mueller, S.; Li, A.; and Pearl, J. 2021. Causes of Effects: Learning individual responses from population data. *arXiv preprint arXiv:2104.13730*.
- Pearl, J. 1995. Causal diagrams for empirical research. *Biometrika*, 82(4): 669–688.
- Sun, W.; Wang, P.; Yin, D.; Yang, J.; and Chang, Y. 2015. Causal inference via sparse additive models with application to online advertising. In *AAAI*, 297–303.
- Tian, J.; and Pearl, J. 2000. Probabilities of causation: Bounds and identification. *Annals of Mathematics and Artificial Intelligence*, 28(1-4): 287–313.
- Tsai, C.-F.; and Lu, Y.-H. 2009. Customer churn prediction by hybrid neural networks. *Expert Systems with Applications*, 36(10): 12547–12553.
- Yan, J.; Liu, N.; Wang, G.; Zhang, W.; Jiang, Y.; and Chen, Z. 2009. How much can behavioral targeting help online advertising? In *Proceedings of the 18th international conference on World Wide Web*, 261–270. ACM.

Appendix

Proof of Theorems

First, we have the following Lemmas 4 and 5 from (Li and Pearl 2019).

Lemma 4. *The c -specific PNS $P(y_x, y'_{x'}|c)$ is bounded as follows:*

$$\begin{aligned} \max \left\{ \begin{array}{l} 0, \\ P(y_x|c) - P(y_{x'}|c), \\ P(y|c) - P(y_{x'}|c), \\ P(y_x|c) - P(y|c) \end{array} \right\} &\leq c\text{-PNS}, \\ \min \left\{ \begin{array}{l} P(y_x|c), \\ P(y'_{x'}|c), \\ P(y, x|c) + P(y', x'|c), \\ P(y_x|c) - P(y_{x'}|c) + \\ +P(y, x'|c) + P(y', x|c) \end{array} \right\} &\geq c\text{-PNS}. \end{aligned}$$

Lemma 5.

$$\begin{aligned} &P(y_x, y'_{x'}|c) - P(y'_{x'}, y_x|c) \\ = &P(y_x|c) - P(y_{x'}|c). \end{aligned}$$

Lemma 6. *Given a causal diagram G and distribution compatible with G , let $Z \cup C$ be a set of variables that does not contain any descendant of X in G , then c -specific PNS $P(y_x, y'_{x'}|c)$ is bounded as follows:*

$$\begin{aligned} \sum_z \max \left\{ \begin{array}{l} 0, \\ P(y_x|z, c) - P(y_{x'}|z, c), \\ P(y|z, c) - P(y_{x'}|z, c), \\ P(y_x|z, c) - P(y|z, c) \end{array} \right\} \\ \times P(z|c) &\leq c\text{-PNS}, \end{aligned} \quad (3)$$

$$\begin{aligned} \sum_z \min \left\{ \begin{array}{l} P(y_x|z, c), \\ P(y'_{x'}|z, c), \\ P(y, x|z, c) + P(y', x'|z, c), \\ P(y_x|z, c) - P(y_{x'}|z, c) + \\ +P(y, x'|z, c) + P(y', x|z, c) \end{array} \right\} \\ \times P(z|c) &\geq c\text{-PNS}. \end{aligned} \quad (4)$$

Proof.

$$\begin{aligned} c\text{-PNS} &= P(y_x, y'_{x'}|c) \\ &= \sum_z P(y_x, y'_{x'}|z, c) \times P(z|c). \end{aligned} \quad (5)$$

From Lemma 4, replace c with (z, c) , we have the following:

$$\begin{aligned} \max \left\{ \begin{array}{l} 0, \\ P(y_x|z, c) - P(y_{x'}|z, c), \\ P(y|z, c) - P(y_{x'}|z, c), \\ P(y_x|z, c) - P(y|z, c) \end{array} \right\} \\ \leq P(y_x, y'_{x'}|z, c), \end{aligned} \quad (6)$$

$$\begin{aligned} \min \left\{ \begin{array}{l} P(y_x|z, c), \\ P(y'_{x'}|z, c), \\ P(y, x|z, c) + P(y', x'|z, c), \\ P(y_x|z, c) - P(y_{x'}|z, c) + \\ +P(y, x'|z, c) + P(y', x|z, c) \end{array} \right\} \\ \geq P(y_x, y'_{x'}|z, c). \end{aligned} \quad (7)$$

Substituting Equations 6 and 7 into Equation 5, Lemma 6 holds.

Note that since we have,

$$\begin{aligned} &\sum_z \max\{0, \\ &P(y_x|z, c) - P(y_{x'}|z, c), \\ &P(y|z, c) - P(y_{x'}|z, c), \\ &P(y_x|z, c) - P(y|z, c)\} \times P(z|c) \\ &\geq \sum_z 0 \times P(z|c) \\ &= 0, \\ &\sum_z \max\{0, \\ &P(y_x|z, c) - P(y_{x'}|z, c), \\ &P(y|z, c) - P(y_{x'}|z, c), \\ &P(y_x|z, c) - P(y|z, c)\} \times P(z|c) \\ &\geq \sum_z [P(y_x|z, c) - P(y_{x'}|z, c)] \times P(z|c) \\ &= P(y_x|c) - P(y_{x'}|c), \end{aligned}$$

$$\begin{aligned} &\sum_z \max\{0, \\ &P(y_x|z, c) - P(y_{x'}|z, c), \\ &P(y|z, c) - P(y_{x'}|z, c), \\ &P(y_x|z, c) - P(y|z, c)\} \times P(z|c) \\ &\geq \sum_z [P(y|z, c) - P(y_{x'}|z, c)] \times P(z|c) \\ &= P(y|c) - P(y_{x'}|c), \end{aligned}$$

$$\begin{aligned} &\sum_z \max\{0, \\ &P(y_x|z, c) - P(y_{x'}|z, c), \\ &P(y|z, c) - P(y_{x'}|z, c), \\ &P(y_x|z, c) - P(y|z, c)\} \times P(z|c) \\ &\geq \sum_z [P(y_x|z, c) - P(y|z, c)] \times P(z|c) \\ &= P(y_x|c) - P(y|c), \end{aligned}$$

then the lower bound in Lemma 6 is guaranteed to be no worse than the lower bound in Lemma 4. Similarly, the upper bound in Lemma 6 is guaranteed to be no worse than the upper bound in Lemma 4. Also note that, since $Z \cup C$ does not contain a descendant of X , the term $P(y_x|z, c)$ refers to experimental data under population z, c . \square

Lemma 7.

$$\begin{aligned} f(c) &= \beta P(y_x, y'_{x'}|c) + \gamma P(y_x, y_{x'}|c) + \\ &\quad + \theta P(y'_{x'}, y'_{x'}|c) + \delta P(y_{x'}, y'_{x'}|c) \\ &= W + \sigma P(y_x, y'_{x'}|c). \end{aligned} \quad (8)$$

where,

$$\begin{aligned} W &= (\gamma - \delta)P(y_x|c) + \delta P(y_{x'}|c) + \theta P(y'_{x'}|c), \\ \sigma &= \beta - \gamma - \theta + \delta. \end{aligned}$$

Proof.

$$\begin{aligned} & f(c) \\ &= \beta P(y_x, y'_{x'}|c) + \gamma P(y_x, y_{x'}|c) + \\ & \quad + \theta P(y'_{x'}, y'_{x'}|c) + \delta P(y'_{x'}, y_{x'}|c) \\ &= \beta P(y_x, y'_{x'}|c) + \gamma [P(y_x|c) - P(y_x, y'_{x'}|c)] + \\ & \quad + \theta [P(y'_{x'}) - P(y_x, y'_{x'}|c)] + \delta P(y'_{x'}, y_{x'}|c) \\ &= \gamma P(y_x|c) + \theta P(y'_{x'}|c) + \\ & \quad + (\beta - \gamma - \theta)P(y_x, y'_{x'}|c) + \delta P(y'_{x'}, y_{x'}|c). \quad (9) \end{aligned}$$

By Lemma 5, we have,

$$P(y'_{x'}, y_{x'}|c) = P(y_x, y'_{x'}|c) - P(y_x|c) + P(y_{x'}|c). \quad (10)$$

Substituting Equation 10 into Equation 9, we have,

$$\begin{aligned} & f(c) \\ &= \gamma P(y_x|c) + \theta P(y'_{x'}|c) + \\ & \quad + (\beta - \gamma - \theta)P(y_x, y'_{x'}|c) + \delta P(y'_{x'}, y_{x'}|c) \\ &= \gamma P(y_x|c) + \theta P(y'_{x'}|c) + \\ & \quad + (\beta - \gamma - \theta)P(y_x, y'_{x'}|c) + \\ & \quad + \delta [P(y_x, y'_{x'}|c) - P(y_x|c) + P(y_{x'}|c)] \\ &= (\gamma - \delta)P(y_x|c) + \delta P(y_{x'}|c) + \theta P(y'_{x'}|c) + \\ & \quad + (\beta - \gamma - \theta + \delta)P(y_x, y'_{x'}|c). \end{aligned}$$

□

Theorem 1. Given a causal diagram G and distribution compatible with G , let $Z \cup C$ be a set of variables that does not contain any descendant of X in G , then the benefit function $f(c) = \beta P(y_x, y'_{x'}|c) + \gamma P(y_x, y_{x'}|c) + \theta P(y'_{x'}, y'_{x'}|c) + \delta P(y'_{x'}, y_{x'}|c)$ is bounded as follows:

$$\begin{aligned} W + \sigma U &\leq f \leq W + \sigma L & \text{if } \sigma < 0, \\ W + \sigma L &\leq f \leq W + \sigma U & \text{if } \sigma > 0, \end{aligned}$$

where σ, W, L, U are given by,

$$\begin{aligned} \sigma &= \beta - \gamma - \theta + \delta, \\ W &= (\gamma - \delta)P(y_x|c) + \delta P(y_{x'}|c) + \theta P(y'_{x'}|c), \\ L &= \sum_z \max \left\{ \begin{array}{l} 0, \\ P(y_x|z, c) - P(y_{x'}|z, c), \\ P(y|z, c) - P(y_{x'}|z, c), \\ P(y_x|z, c) - P(y|z, c) \end{array} \right\} \\ &\quad \times P(z|c), \\ U &= \sum_z \min \left\{ \begin{array}{l} P(y_x|z, c), \\ P(y'_{x'}|z, c), \\ P(y, x|z, c) + P(y', x'|z, c), \\ P(y_x|z, c) - P(y_{x'}|z, c) + \\ + P(y, x'|z, c) + P(y', x|z, c) \end{array} \right\} \\ &\quad \times P(z|c). \end{aligned}$$

Proof. By Lemmas 6 and 7,

substituting Equations 3 and 4 into Equation 8, Theorem 1 holds.

Note that, if we substituting Lemma 4 into Lemma 7, we have the same results as in Li-Pearl's Theorem. We showed that in Lemma 6 that the bounds in Lemma 6 is guaranteed to be no worse than the bounds in Lemma 4, therefore, the bounds in Theorem 1 is guaranteed to be no worse than the bounds in Li-Pearl's Theorem. □

Lemma 8. Given a causal diagram G and distribution compatible with G , let $Z \cup C$ be a set of variables such that $\forall x, x' \in X : x \neq x', (Y_x \perp\!\!\!\perp X \cup Z_{x'} \mid Z_x, C)$ in G , then the c -PNS $P(y_x, y'_{x'}|c)$ is bounded as follows:

$$\max \left\{ \begin{array}{l} 0, \\ P(y_x|c) - P(y_{x'}|c), \\ P(y|c) - P(y_{x'}|c), \\ P(y_x|c) - P(y|c) \end{array} \right\} \leq c\text{-PNS}, \quad (11)$$

$$\min \left\{ \begin{array}{l} P(y_x|c), \\ P(y'_{x'}|c), \\ P(y, x|c) + P(y', x'|c), \\ P(y_x|c) - P(y_{x'}|c) + \\ + P(y, x'|c) + P(y', x|c), \\ \sum_z \sum_{z'} \min\{P(y|z, x, c), \\ P(y'|z', x', c)\} \\ \times \min\{P(z_x|c), P(z'_{x'}|c)\} \end{array} \right\} \geq c\text{-PNS}. \quad (12)$$

Proof.

c -PNS

$$\begin{aligned} &= P(y_x, y'_{x'}|c) \\ &= \sum_z \sum_{z'} P(y_x, y'_{x'}, z_x, z'_{x'}|c) \\ &= \sum_z \sum_{z'} P(y_x, y'_{x'}|z_x, z'_{x'}, c) \times P(z_x, z'_{x'}|c) \\ &\leq \sum_z \sum_{z'} \min\{P(y_x|z_x, z'_{x'}, c), P(y'_{x'}|z_x, z'_{x'}, c)\} \\ &\quad \times \min\{P(z_x|c), P(z'_{x'}|c)\} \\ &= \sum_z \sum_{z'} \min\{P(y_x|z_x, c), P(y'_{x'}|z'_{x'}, c)\} \\ &\quad \times \min\{P(z_x|c), P(z'_{x'}|c)\} \quad (13) \end{aligned}$$

$$\begin{aligned} &= \sum_z \sum_{z'} \min\{P(y|z_x, x, c), P(y'|z'_{x'}, x', c)\} \\ &\quad \times \min\{P(z_x|c), P(z'_{x'}|c)\} \quad (14) \\ &= \sum_z \sum_{z'} \min\{P(y|z, x, c), P(y'|z', x', c)\} \\ &\quad \times \min\{P(z_x|c), P(z'_{x'}|c)\}. \end{aligned}$$

Combined with the bounds in Lemma 4, Lemma 8 holds. Note that Equation 13 is due to $Y_x \perp\!\!\!\perp Z_{x'} \mid Z_x, C$ and $Y_{x'} \perp\!\!\!\perp Z_x \mid Z_{x'}, C$. Equation 14 is due to $\forall x \in X, Y_x \perp\!\!\!\perp X \mid Z_x, C$. □

Theorem 2. Given a causal diagram G and distribution compatible with G , let Z be a set of variables such that $\forall x, x' \in X : x \neq x', (Y_x \perp\!\!\!\perp X \cup Z_{x'} \mid Z_x, C)$ in G , and C does not contain any descendant of X in G , then the benefit function $f(c) = \beta P(y_x, y'_{x'}|c) + \gamma P(y_x, y_{x'}|c) + \theta P(y'_{x'}, y'_{x'}|c) + \delta P(y'_{x'}, y_{x'}|c)$ is bounded as follows:

$$\begin{aligned} W + \sigma U &\leq f \leq W + \sigma L & \text{if } \sigma < 0, \\ W + \sigma L &\leq f \leq W + \sigma U & \text{if } \sigma > 0, \end{aligned}$$

where σ, W, L, U are given by,

$$\begin{aligned} \sigma &= \beta - \gamma - \theta + \delta, \\ W &= (\gamma - \delta)P(y_x|c) + \delta P(y_{x'}|c) + \theta P(y'_{x'}|c), \\ L &= \max \left\{ \begin{array}{l} 0, \\ P(y_x|c) - P(y_{x'}|c), \\ P(y|c) - P(y_{x'}|c), \\ P(y_x|c) - P(y|c) \end{array} \right\}, \\ U &= \min \left\{ \begin{array}{l} P(y_x|c), \\ P(y'_{x'}|c), \\ P(y, x|c) + P(y', x'|c), \\ P(y_x|c) - P(y_{x'}|c) + \\ + P(y, x'|c) + P(y', x|c), \\ \sum_z \sum_{z'} \min\{P(y|z, x, c), \\ P(y'|z', x', c)\} \\ \times \min\{P(z_x|c), P(z'_{x'}|c)\} \end{array} \right\}. \end{aligned}$$

Proof. By Lemmas 8 and 7, substituting Equations 11 and 12 into Equation 8, Theorem 2 holds.

Note that, if we substituting Lemma 4 into Lemma 7, we have the same results as in Li-Pearl's Theorem. From the proof of Lemma 8, we know that the lower bound in Lemma 8 is the same as in Lemma 4 and the upper bound in Lemma 8 is no worse than the upper bound in Lemma 4. Therefore, the lower bound in Theorem 2 is the same as in Li-Pearl's Theorem, and the upper bound in Theorem 2 is guaranteed to be no worse than the upper bound in Li-Pearl's Theorem. \square

Lemma 9. *Given a causal diagram G in Figure 9 and distribution that compatible with G , and C is not a descendant of X , then c -PNS $P(y_x, y'_{x'}|c)$ is bounded as follow:*

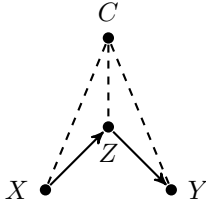


Figure 9: Mediator Z with no direct effects of X on Y .

$$\max \left\{ \begin{array}{l} 0, \\ P(y_x|c) - P(y_{x'}|c), \\ P(y|c) - P(y_{x'}|c), \\ P(y_x|c) - P(y|c) \end{array} \right\} \leq c\text{-PNS}, \quad (15)$$

$$\min \left\{ \begin{array}{l} P(y_x|c), \\ P(y'_{x'}|c), \\ P(y, x|c) + P(y', x'|c), \\ P(y_x|c) - P(y_{x'}|c) + \\ + P(y, x'|c) + P(y', x|c), \\ \sum_z \sum_{z'} \min\{P(y|z, c), \\ P(y'|z', c)\} \\ \times \min\{P(z|x, c), P(z'|x', c)\} \end{array} \right\} \geq c\text{-PNS}. \quad (16)$$

Proof. First we show that in graph G , if an individual is a c -complier from X to Y , then $Z_x|c$ and $Z_{x'}|c$ must have the different values. This is because the structural equations for Y and Z are $f_y(z, u_y, c)$ and $f_z(x, u_z, c)$, respectively. If an individual has the same $Z_x|c$ and $Z_{x'}|c$ value, then $f_z(x, u_z, c) = f_z(x', u_z, c)$. This means $f_y(f_z(x, u_z, c), u_y, c) = f_y(f_z(x', u_z, c), u_y, c)$, i.e., $Y_x|c$ and $Y_{x'}|c$ must have the same value. Thus this individual is not a c -complier. Therefore,

$$\begin{aligned} &c\text{-PNS} \\ &= P(y_x, y'_{x'}|c) \\ &= \sum_z \sum_{z' \neq z} P(y_z, y'_{z'}|c) \times P(z_x, z'_{x'}|c) \\ &\leq \sum_z \sum_{z' \neq z} \min\{P(y_z|c), P(y'_{z'}|c)\} \\ &\quad \times \min\{P(z_x|c), P(z'_{x'}|c)\} \\ &= \sum_z \sum_{z' \neq z} \min\{P(y|z, c), P(y'|z', c)\} \\ &\quad \times \min\{P(z|x, c), P(z'|x', c)\}. \end{aligned}$$

Combined with the bounds in Lemma 4, Lemma 9 holds. \square

Theorem 3. *Given a causal diagram G in Figure 9 and distribution compatible with G , and C does not contain any descendant of X , then the benefit function $f(c) = \beta P(y_x, y'_{x'}|c) + \gamma P(y_x, y_{x'}|c) + \theta P(y'_{x'}, y'_{x'}|c) + \delta P(y_{x'}, y'_{x'}|c)$ is bounded as follows:*

$$\begin{aligned} W + \sigma U &\leq f \leq W + \sigma L && \text{if } \sigma < 0, \\ W + \sigma L &\leq f \leq W + \sigma U && \text{if } \sigma > 0, \end{aligned}$$

where σ, W, L, U are given by,

$$\begin{aligned} \sigma &= \beta - \gamma - \theta + \delta, \\ W &= (\gamma - \delta)P(y_x|c) + \delta P(y_{x'}|c) + \theta P(y'_{x'}|c), \\ L &= \max \left\{ \begin{array}{l} 0, \\ P(y_x|c) - P(y_{x'}|c), \\ P(y|c) - P(y_{x'}|c), \\ P(y_x|c) - P(y|c) \end{array} \right\}, \\ U &= \min \left\{ \begin{array}{l} P(y_x|c), \\ P(y'_{x'}|c), \\ P(y, x|c) + P(y', x'|c), \\ P(y_x|c) - P(y_{x'}|c) + \\ + P(y, x'|c) + P(y', x|c), \\ \sum_z \sum_{z' \neq z} \min\{P(y|z, c), \\ P(y'|z', c)\} \\ \times \min\{P(z|x, c), P(z'|x', c)\} \end{array} \right\}. \end{aligned}$$

Proof. By Lemmas 9 and 7, substituting Equations 15 and 16 into Equation 8, Theorem 3 holds.

Note that, if we substituting Lemma 4 into Lemma 7, we have the same results as in Li-Pearl's Theorem. From the proof of Lemma 9, we know that the lower bound in Lemma 9 is the same as in Lemma 4 and the upper bound in Lemma 9 is no worse than the upper bound in Lemma 4. Therefore, the lower bound in Theorem 3 is the same as in Li-Pearl's Theorem, and the upper bound in Theorem 3 is guaranteed to be no worse than the upper bound in Li-Pearl's Theorem. \square

Calculation in the Examples

In order to clearly see the calculation steps, we list an equivalent form of Li-Pearl's Theorem as following (see the proof in the previous section for the equivalence):

Theorem 10. *Given a causal diagram G and distribution compatible with G , let C be a set of variables that does not contain any descendant of X in G , then the benefit function $f(c) = \beta P(y_x, y'_{x'}|c) + \gamma P(y_x, y_{x'}|c) + \theta P(y'_{x'}, y'_{x'}|c) + \delta P(y_{x'}, y'_x|c)$ is bounded as follows:*

$$\begin{aligned} W + \sigma U &\leq f(c) \leq W + \sigma L && \text{if } \sigma < 0, \\ W + \sigma L &\leq f(c) \leq W + \sigma U && \text{if } \sigma > 0, \end{aligned}$$

where σ, W, L, U are given by,

$$\begin{aligned} \sigma &= \beta - \gamma - \theta + \delta, \\ W &= (\gamma - \delta)P(y_x|c) + \delta P(y_{x'}|c) + \theta P(y'_{x'}|c), \\ L &= \max \left\{ \begin{array}{l} 0, \\ P(y_x|c) - P(y_{x'}|c), \\ P(y|c) - P(y_{x'}|c), \\ P(y_x|c) - P(y|c) \end{array} \right\}, \\ U &= \min \left\{ \begin{array}{l} P(y_x|c), \\ P(y'_{x'}|c), \\ P(y, x|c) + P(y', x'|c), \\ P(y_x|c) - P(y_{x'}|c) + \\ + P(y, x'|c) + P(y', x|c) \end{array} \right\}. \end{aligned}$$

Company Selection First, we apply Li-Pearl's Theorem (Theorem 10) to the data in Tables 1 and 2. The benefit vector is $(100, -60, 0, -140)$.

We have,

$$\begin{aligned} \sigma &= \beta - \gamma - \theta + \delta \\ &= 100 - (-60) - 0 + (-140) \\ &= 20 \end{aligned}$$

$$\begin{aligned} W &= (\gamma - \delta)P(r_a|c) + \delta P(r_{a'}|c) + \theta P(r'_{a'}|c) \\ &= (-60 - (-140)) \times 0.83729 + 0 \times 0.47405 + \\ &\quad + (-140) \times 0.52595 \\ &= -6.64980 \end{aligned}$$

$$\begin{aligned} L &= \max \left\{ \begin{array}{l} 0, \\ P(r_a|c) - P(r_{a'}|c), \\ P(r|c) - P(r_{a'}|c), \\ P(r_a|c) - P(r|c) \end{array} \right\} \\ &= \max \left\{ \begin{array}{l} 0, \\ 0.83729 - 0.52595, \\ 0.70714 - 0.52595, \\ 0.83729 - 0.70714 \end{array} \right\} \\ &= 0.31134 \end{aligned}$$

$$\begin{aligned} U &= \min \left\{ \begin{array}{l} P(r_a|c), \\ P(r'_{a'}|c), \\ P(r, a|c) + P(r', a'|c), \\ P(r_a|c) - P(r_{a'}|c) + \\ + P(r, a'|c) + P(r', a|c) \end{array} \right\} \\ &= \min \left\{ \begin{array}{l} 0.83729, \\ 1 - 0.52595, \\ 0.35286 + 0.20143, \\ 0.83729 - 0.52595 + \\ + 0.35428 + 0.09143 \end{array} \right\} \\ &= 0.47405 \end{aligned}$$

Therefore,

$$\begin{aligned} W + \sigma L &\leq f(c) \leq W + \sigma U, \\ -6.64980 + 20 \times 0.31134 &\leq f(c) \\ &\leq -6.64980 + 20 \times 0.47405, \\ -0.423 &\leq f(c) \leq 2.832. \end{aligned}$$

Then, we apply Theorem 1 to the data in Tables 1 and 2. σ and W are the same as above.

And we have,

$$\begin{aligned} L &= \sum_z \max \left\{ \begin{array}{l} 0, \\ P(r_a|z, c) - P(r_{a'}|z, c), \\ P(r|z, c) - P(r_{a'}|z, c), \\ P(r_a|z, c) - P(r|z, c) \end{array} \right\} \\ &\quad \times P(z|c) \\ &= \max \left\{ \begin{array}{l} 0, \\ 0.44600 - 0.05000, \\ 0.49010 - 0.05000, \\ 0.44600 - 0.49010 \end{array} \right\} \times 0.28857 \\ &\quad + \max \left\{ \begin{array}{l} 0, \\ 0.99600 - 0.71900, \\ 0.79518 - 0.71900, \\ 0.99600 - 0.79518 \end{array} \right\} \times 0.71143 \\ &= 0.44010 \times 0.28857 + 0.27700 \times 0.71143 \\ &= 0.32407 \end{aligned}$$

$$\begin{aligned}
U &= \sum_z \min \left\{ \begin{array}{c} P(r_a|z, c), \\ P(r_{a'}|z, c), \\ P(r, a|z, c) + P(r', a'|z, c), \\ P(r_a|z, c) - P(r_{a'}|z, c) + \\ + P(r, a'|z, c) + P(r', a|z, c) \end{array} \right\} \\
&\quad \times P(z|c) \\
&= \min \left\{ \begin{array}{c} 0.44600, \\ 1 - 0.05000, \\ 0.44555 + 0.20297, \\ 0.44600 - 0.05000 + \\ + 0.04455 + 0.30693 \end{array} \right\} \times 0.28857 \\
&\quad + \min \left\{ \begin{array}{c} 0.99600, \\ 1 - 0.71900, \\ 0.31526 + 0.20080, \\ 0.99600 - 0.71900 + \\ + 0.47992 + 0.00402 \end{array} \right\} \times 0.71143 \\
&= 0.44600 \times 0.28857 + 0.28100 \times 0.71143 \\
&= 0.32862
\end{aligned}$$

Therefore,

$$\begin{aligned}
W + \sigma L &\leq f(c) \leq W + \sigma U, \\
-6.64980 + 20 \times 0.32407 &\leq f(c) \\
\leq -6.64980 + 20 \times 0.32862, \\
-0.168 &\leq f(c) \leq -0.077.
\end{aligned}$$

Effective Patients of a Drug First, the set $\{C\}$ satisfied the back-door criterion for both (A, Z) and (A, R) . By Pearl's adjustment formula, the experimental data needed are:

$$\begin{aligned}
P(r_a|c) &= P(r|a, c) = 0.66666, \\
P(r_{a'}|c) &= P(r|a', c) = 0.33265, \\
P(z_a|c) &= P(z|a, c) = 0.68878, \\
P(z_{a'}|c) &= P(z'|a', c) = 0.01232.
\end{aligned}$$

Then, we apply Li-Pearl's Theorem (Theorem 10) to the data in Table 3 and the above experimental data. The benefit vector is $(1, -1, -1, -1)$.

We have,

$$\begin{aligned}
\sigma &= \beta - \gamma - \theta + \delta \\
&= 1 - (-1) - (-1) + (-1) \\
&= 2
\end{aligned}$$

$$\begin{aligned}
W &= (\gamma - \delta)P(r_a|c) + \delta P(r_{a'}|c) + \theta P(r'_{a'}|c) \\
&= (-1 + 1)P(r_a|c) - P(r_{a'}|c) - P(r'_{a'}|c) \\
&= -1
\end{aligned}$$

$$\begin{aligned}
L &= \max \left\{ \begin{array}{c} 0, \\ P(r_a|c) - P(r_{a'}|c), \\ P(r|c) - P(y_{a'}|c), \\ P(r_a|c) - P(r|c) \end{array} \right\} \\
&= \max \left\{ \begin{array}{c} 0, \\ 0.66666 - 0.33265, \\ 0.51535 - 0.33265, \\ 0.66666 - 0.51535 \end{array} \right\} \\
&= 0.33401
\end{aligned}$$

$$\begin{aligned}
U &= \min \left\{ \begin{array}{c} P(r_a|c), \\ P(r_{a'}|c), \\ P(r, a|c) + P(r', a'|c), \\ P(r_a|c) - P(r_{a'}|c) + \\ + P(r, a'|c) + P(r', a|c) \end{array} \right\} \\
&= \min \left\{ \begin{array}{c} 0.66666, \\ 1 - 0.33265, \\ 0.36465 + 0.30233, \\ 0.66666 - 0.33265 + \\ + 0.15070 + 0.18232 \end{array} \right\} \\
&= 0.66666
\end{aligned}$$

Therefore,

$$\begin{aligned}
W + \sigma L &\leq f(c) \leq W + \sigma U, \\
-1 + 2 \times 0.33401 &\leq f(c) \\
\leq -1 + 2 \times 0.66666, \\
-0.3320 &\leq f(c) \leq 0.3333.
\end{aligned}$$

Then, we apply Theorem 2 to the data in Table 3 and the above experimental data. σ , W , and L are the same as above. And we have,

$$\begin{aligned}
U &= \min \left\{ \begin{array}{c} P(r_a|c), \\ P(r_{a'}|c), \\ P(r, a|c) + P(r', a'|c), \\ P(r_a|c) - P(r_{a'}|c) + \\ + P(r, a'|c) + P(r', a|c), \\ \sum_z \sum_{z'} \min\{P(r|z, a, c), \\ P(r'|z', a', c)\} \\ \times \min\{P(z_a|c), P(z_{a'}|c)\} \end{array} \right\} \\
&= \min \left\{ \begin{array}{c} 0.66666, \\ 1 - 0.33265, \\ 0.36465 + 0.30233, \\ 0.66666 - 0.33265 + \\ + 0.15070 + 0.18232, \\ \min\{0.92593, 0.66944\} \times \\ \min\{0.68878, 0.98768\} + \\ \min\{0.92593, 0.50000\} \times \\ \min\{0.68878, 0.01232\} + \\ \min\{0.09290, 0.66944\} \times \\ \min\{0.31122, 0.98768\} + \\ \min\{0.09290, 0.50000\} \times \\ \min\{0.31122, 0.01232\} \end{array} \right\} \\
&= 0.49731
\end{aligned}$$

Therefore,

$$\begin{aligned}
W + \sigma L &\leq f(c) \leq W + \sigma U, \\
-1 + 2 \times 0.33401 &\leq f(c) \\
\leq -1 + 2 \times 0.49731, \\
-0.3320 &\leq f(c) \leq -0.0054.
\end{aligned}$$

Algorithm 1: Generate sample distributions for non-descendant covariates

Input: n , number of sample distributions needed.

Output: n sample distributions (observational data and experimental data).

```

1: for  $i = 1$  to  $n$  do
2:    $\text{\textit{rand}}(0, 1)$  is the function that random uniformly
   generate a number from 0 to 1.
3:    $t_1, t_2, t_3$ , and  $t_4$  can be interpreted as the number of
   individuals such that  $x \wedge z$ ,  $x' \wedge z$ ,  $x \wedge z'$ , and  $x' \wedge z'$ 
   respectively.
4:    $t_1 = \text{\textit{rand}}(0, 1) \times 1000$ ;
5:    $t_2 = \text{\textit{rand}}(0, 1) \times (1000 - t_1)$ ;
6:    $t_3 = \text{\textit{rand}}(0, 1) \times (1000 - t_1 - t_2)$ ;
7:    $t_4 = 1000 - t_1 - t_2 - t_3$ ;
8:    $o_1, o_2, o_3$ , and  $o_4$  can be interpreted as the number
   of individuals such that  $x \wedge z \wedge y$ ,  $x' \wedge z \wedge y$ ,  $x \wedge z' \wedge y$ ,
   and  $x' \wedge z' \wedge y$  respectively.
9:    $o_1 = \text{\textit{rand}}(0, 1) \times t_1$ ;
10:   $o_2 = \text{\textit{rand}}(0, 1) \times t_2$ ;
11:   $o_3 = \text{\textit{rand}}(0, 1) \times t_3$ ;
12:   $o_4 = \text{\textit{rand}}(0, 1) \times t_4$ ;
13:  // Each  $c_i$  corresponding to a sample distribution.
14:  // The following are experimental data that satisfied
   the general bounds provided by Tian and Pearl.
15:   $P(y|do(x), z, c_i) = \text{\textit{rand}}(0, 1) \times \frac{t_2}{t_1+t_2} + \frac{o_1}{t_1+t_2}$ ;
16:   $P(y|do(x'), z, c_i) = \text{\textit{rand}}(0, 1) \times \frac{t_1}{t_1+t_2} + \frac{o_2}{t_1+t_2}$ ;
17:   $P(y|do(x), z', c_i) = \text{\textit{rand}}(0, 1) \times \frac{t_4}{t_3+t_4} + \frac{o_3}{t_3+t_4}$ ;
18:   $P(y|do(x'), z', c_i) = \text{\textit{rand}}(0, 1) \times \frac{t_3}{t_3+t_4} + \frac{o_4}{t_3+t_4}$ ;
19:  // The following are observational data.
20:   $P(x, y, z|c_i) = o_1/1000$ ;
21:   $P(x, y, z'|c_i) = o_3/1000$ ;
22:   $P(x, y', z|c_i) = (t_1 - o_1)/1000$ ;
23:   $P(x, y', z'|c_i) = (t_3 - o_3)/1000$ ;
24:   $P(x', y, z|c_i) = o_2/1000$ ;
25:   $P(x', y, z'|c_i) = o_4/1000$ ;
26:   $P(x', y', z|c_i) = (t_2 - o_2)/1000$ ;
27:   $P(x', y', z'|c_i) = (t_4 - o_4)/1000$ ;
28: end for

```

Distribution Generating Algorithms

Here, the sample distribution generating algorithms in simulated studies are presented.

Non-descendant Covariates The Algorithm 1 is the sample distribution generating algorithm in the simulated study of non-descendant covariates case. It generated both experimental and observational data compatible with Figure 5 (X, Y, Z are binary) that satisfy the general relation provided by Tian and Pearl (i.e., the general relation between experimental and observational data).

Partial Mediators The observational data compatible with Figure 1 (X, Y, Z are binary) in the simulated study of partial mediators case was generated by Algorithm 2. The experimental data needed was computed via adjustment formula because the set $\{C\}$ satisfied the back-door criterion for both

Algorithm 2: Generate sample distributions for partial mediators

Input: n , number of sample distributions needed.

Output: n sample distributions (observational data in conditional probability tables).

```

1: for  $i = 1$  to  $n$  do
2:    $\text{\textit{rand}}(0, 1)$  is the function that random uniformly
   generate a number from 0 to 1.
3:   // Each  $c_i$  corresponding to a sample distribution.
4:    $P(x|c_i) = \text{\textit{rand}}(0, 1)$ ;
5:    $P(z|x, c_i) = \text{\textit{rand}}(0, 1)$ ;
6:    $P(z|x', c_i) = \text{\textit{rand}}(0, 1)$ ;
7:    $P(y|x, z, c_i) = \text{\textit{rand}}(0, 1)$ ;
8:    $P(y|x', z, c_i) = \text{\textit{rand}}(0, 1)$ ;
9:    $P(y|x, z', c_i) = \text{\textit{rand}}(0, 1)$ ;
10:   $P(y|x', z', c_i) = \text{\textit{rand}}(0, 1)$ ;
11: end for

```

Algorithm 3: Generate sample distributions for pure mediators

Input: n , number of sample distributions needed.

Output: n sample distributions (observational data in conditional probability tables).

```

1: for  $i = 1$  to  $n$  do
2:    $\text{\textit{rand}}(0, 1)$  is the function that random uniformly
   generate a number from 0 to 1.
3:   // Each  $c_i$  corresponding to a sample distribution.
4:    $P(x|c_i) = \text{\textit{rand}}(0, 1)$ ;
5:    $P(z|x, c_i) = \text{\textit{rand}}(0, 1)$ ;
6:    $P(z|x', c_i) = \text{\textit{rand}}(0, 1)$ ;
7:    $P(y|z, c_i) = \text{\textit{rand}}(0, 1)$ ;
8:    $P(y|z', c_i) = \text{\textit{rand}}(0, 1)$ ;
9: end for

```

(X, Z) and (X, Y).

Pure Mediators The observational data compatible with Figure 2 (X, Y, Z are binary) in the simulated study of pure mediators case was generated by Algorithm 3. The experimental data needed was computed via adjustment formula because the set $\{C\}$ satisfied the back-door criterion for (X, Y).