

---

# Probabilities of Causation: Role of Observational Data

---

**Ang Li**

Cognitive Systems Laboratory,  
Department of Computer Science,  
University of California Los Angeles  
Los Angeles, CA 90095  
angli@cs.ucla.edu

**Judea Pearl**

Cognitive Systems Laboratory,  
Department of Computer Science,  
University of California Los Angeles  
Los Angeles, CA 90095  
judea@cs.ucla.edu

## Abstract

Probabilities of causation play a crucial role in modern decision-making. Pearl defined three binary probabilities of causation, the probability of necessity and sufficiency (PNS), the probability of sufficiency (PS), and the probability of necessity (PN). These probabilities were then bounded by Tian and Pearl using a combination of experimental and observational data. However, observational data are not always available in practice; in such a case, Tian and Pearl’s Theorem provided valid but less effective bounds using pure experimental data. In this paper, we discuss the conditions that observational data are worth considering to improve the quality of the bounds. More specifically, we defined the expected improvement of the bounds by assuming the observational distributions are uniformly distributed on their feasible interval. We further applied the proposed theorems to the unit selection problem defined by Li and Pearl.

## 1 INTRODUCTION

Probabilities of causation are widely used in social science, political science, health science, economics, etc. For example, Li and Pearl used a linear combination of probabilities of causation to solve the unit selection problem (10; 11; 13; 15), showing the advantage of their model over the A/B test heuristic. Also, Scott and Pearl showed that the probabilities of causation should be considered in personal decision-making (16). In addition, probabilities of causation have improved the accuracy of machine learning models by including probabilities of causation terms in the label of training data (12). Machine learning models can also be used to learn the bounds of Probabilities of causation if the proper training data are provided (6; 7).

The probabilities of causation have been studied for decades.

Pearl first defined three binary probabilities of causation (i.e., PNS, PN, and PS) (19) using the structural causal model (SCM) (3; 4). The sharp bounds of these probabilities of causation are then derived by Tian and Pearl (referred to as Tian-Pearl’s Theorem) (20) utilizing a combination of experimental and observational data, and Balke’s linear programming (1). These bounds are then improved with additional covariate information and causal structures (2; 14; 17). Besides, the theoretical foundation of non-binary probabilities of causation was studied by Li and Pearl (9).

The probabilities of causation are not generally identifiable (i.e., no point estimation exists if there is no additional assumption); therefore, all the works mentioned above focused on bounding the probabilities of causation using a combination of experimental and observational data. However, observational data are not always available in the real world, and observational studies usually take longer time to conduct. For example, a new drug is generally proven to the market by only a randomized controlled clinical study (i.e., experimental study), lacking observational study (i.e., patients are free to choose whether or not to take the drug). In such a situation, Tian-Pearl’s Theorem can still provide valid bounds of the probabilities of causation using pure experimental data. However, the bounds might be less informative (e.g.,  $0.1 \leq PNS \leq 0.9$ ). We then wonder whether we should collect observational data to improve the bounds and what the expected improvement could be. In this paper, we formally defined the expected improvement of lower and upper bounds assuming the observational distributions are uniformly distributed on their feasible interval.

## 2 PRELIMINARIES

Here, we review the three binary probabilities of causation defined by Pearl (19) and their sharp bounds derived by Tian and Pearl (20). Readers who are familiar with the model may skip this section.

The language of SCM (3; 4) will be used in which the probabilities of causation are well-defined. The basic coun-

terfactual sentence “Variable  $Y$  would have the value  $y$ , had  $X$  been  $x$ ” is denoted by  $Y_{X=x} = y$  and shorted as  $y_x$ . The experimental data are in the form of the causal effects  $P(y_x)$ , and the observational data are in the form of a joint probability function  $P(x, y)$ .

The three binary probabilities are then defined as follows:

**Definition 1** (Probability of necessity (PN)). *Let  $X$  and  $Y$  be two binary variables in a causal model  $M$ , let  $x$  and  $y$  stand for the propositions  $X = \text{true}$  and  $Y = \text{true}$ , respectively, and  $x'$  and  $y'$  for their complements. The probability of necessity is defined as the expression (19)*

$$\begin{aligned} PN &\triangleq P(Y_{x'} = \text{false} | X = \text{true}, Y = \text{true}) \\ &\triangleq P(y'_{x'} | x, y) \end{aligned}$$

**Definition 2** (Probability of sufficiency (PS)). *Let  $X$  and  $Y$  be two binary variables in a causal model  $M$ , let  $x$  and  $y$  stand for the propositions  $X = \text{true}$  and  $Y = \text{true}$ , respectively, and  $x'$  and  $y'$  for their complements. The probability of sufficiency is defined as the expression (19)*

$$PS \triangleq P(y_x | y', x')$$

**Definition 3** (Probability of necessity and sufficiency (PNS)). *Let  $X$  and  $Y$  be two binary variables in a causal model  $M$ , let  $x$  and  $y$  stand for the propositions  $X = \text{true}$  and  $Y = \text{true}$ , respectively, and  $x'$  and  $y'$  for their complements. The probability of necessity and sufficiency is defined as the expression (19)*

$$PNS \triangleq P(y_x, y'_{x'})$$

The bounds of the above probabilities of causation are as follows:

$$PNS \geq \max \left\{ \begin{array}{c} 0, \\ P(y_x) - P(y_{x'}), \\ P(y) - P(y_{x'}), \\ P(y_x) - P(y) \end{array} \right\} \quad (1)$$

$$PNS \leq \min \left\{ \begin{array}{c} P(y_x), \\ P(y'_{x'}), \\ P(x, y) + P(x', y'), \\ P(y_x) - P(y_{x'}) + \\ + P(x, y') + P(x', y) \end{array} \right\} \quad (2)$$

$$PN \geq \max \left\{ \begin{array}{c} 0, \\ \frac{P(y) - P(y_{x'})}{P(x, y)} \end{array} \right\}$$

$$PN \leq \min \left\{ \begin{array}{c} 1, \\ \frac{P(y'_{x'}) - P(x', y')}{P(x, y)} \end{array} \right\}$$

$$PS \geq \max \left\{ \begin{array}{c} 0, \\ \frac{P(y') - P(y'_x)}{P(x', y')} \end{array} \right\}$$

$$PS \leq \min \left\{ \begin{array}{c} 1, \\ \frac{P(y_x) - P(x, y)}{P(x', y')} \end{array} \right\}$$

The theoretical proof of these bounds can be found in (13). Note that in these bounds, both experimental data (e.g.,  $P(y_x)$  in Equation 1) and observational data (e.g.,  $P(y)$  in Equation 1) are used. If the observational data are not available, the bounds of PNS become the following:

$$PNS \geq \max \left\{ \begin{array}{c} 0, \\ P(y_x) - P(y_{x'}) \end{array} \right\} \quad (3)$$

$$PNS \leq \min \left\{ \begin{array}{c} P(y_x), \\ P(y'_{x'}) \end{array} \right\} \quad (4)$$

In this paper, we will focus on the expected improvement of the lower (i.e., Equation 1 – Equation 3) and upper (i.e., Equation 4 – Equation 2) bounds of PNS. The results can be extended to other probabilities of causation.

### 3 MAIN RESULTS

Suppose a decision maker has the experimental data  $P(y_x)$  and  $P(y_{x'})$ , and obtained the bounds of PNS by Equations 3 and 4. To improve the obtained bounds, the decision maker wondered whether he should conduct an observational study to obtain observational data (i.e., apply Equations 1 and 2). We then have the following theorems (the proof of all theorems in this paper is provided in the appendix):

**Theorem 4.** *Given experimental data  $P(y_x)$  and  $P(y_{x'})$  and let  $D = P(y)$  be a random variable. Let  $L$  be the lower bound of PNS using pure experimental data and  $L'$  be the lower bound of PNS using a combination of experimental and observational data. If  $D$  is uniformly distributed on its feasible interval  $[\max(0, P(y_x) - P(y'_{x'})), \min(1, P(y_x) + P(y_{x'}))]$ , and  $P(y_x) + P(y_{x'}) \neq 0$  and  $P(y'_x) + P(y'_{x'}) \neq 0$ , then we have the expectation of the increased lower bound  $E(L' - L)$  as follows:*

$$\begin{aligned} &E(L' - L) \\ &= \frac{\min\{P^2(y_x), P^2(y'_{x'}), P^2(y_{x'}), P^2(y'_{x'})\}}{\min\{P(y_x) + P(y_{x'}), P(y'_x) + P(y'_{x'})\}} \end{aligned}$$

where,

$$\begin{aligned} P(y'_x) &= 1 - P(y_x), \\ P(y'_{x'}) &= 1 - P(y_{x'}), \end{aligned}$$

$$L' = \max \left\{ \begin{array}{c} 0, \\ P(y_x) - P(y_{x'}), \\ D - P(y_{x'}), \\ P(y_x) - D \end{array} \right\},$$

$$L = \max \left\{ \begin{array}{c} 0, \\ P(y_x) - P(y_{x'}) \end{array} \right\}.$$

**Theorem 5.** Given experimental data  $P(y_x)$  and  $P(y_{x'})$  and let  $D = P(x, y) + P(x', y')$  be a random variable. Let  $U$  be the upper bound of PNS using pure experimental data and  $U'$  be the upper bound of PNS using a combination of experimental and observational data. If  $D$  is uniformly distributed on its feasible interval  $[\max(0, P(y_x) - P(y_{x'})), \min(1, P(y_x) + P(y'_{x'}))]$ , and  $P(y_x) + P(y'_{x'}) \neq 0$  and  $P(y'_{x'}) + P(y_{x'}) \neq 0$ , then we have the expectation of the decreased upper bound  $E(U - U')$  as follows:

$$\begin{aligned} & E(U - U') \\ &= \frac{\min\{P^2(y_x), P^2(y'_{x'}), P^2(y_{x'}), P^2(y'_{x'})\}}{\min\{P(y_x) + P(y'_{x'}), P(y'_{x'}) + P(y_{x'})\}} \end{aligned}$$

where,

$$\begin{aligned} P(y'_{x'}) &= 1 - P(y_x), \\ P(y'_{x'}) &= 1 - P(y_{x'}), \end{aligned}$$

$$U' = \min \left\{ \begin{array}{c} P(y_x), \\ P(y'_{x'}), \\ D, \\ P(y_x) + P(y'_{x'}) - D \end{array} \right\},$$

$$U = \min \left\{ \begin{array}{c} P(y_x), \\ P(y'_{x'}) \end{array} \right\}.$$

The expected increased lower bound and decreased upper bound represent the improvement of the bounds of PNS when considering observational data. Note that Theorem 4 requires  $P(y_x) + P(y_{x'}) \neq 0$  and  $P(y'_{x'}) + P(y'_{x'}) \neq 0$ , and Theorem 5 requires  $P(y_x) + P(y'_{x'}) \neq 0$  and  $P(y'_{x'}) + P(y_{x'}) \neq 0$ . In fact, if  $P(y_x) + P(y_{x'}) = 0$  or  $P(y_x) + P(y'_{x'}) = 0$  or  $P(y'_{x'}) + P(y'_{x'}) = 0$ , PNS is reduced to the point estimation 0 (i.e.,  $PNS = 0$ ), and if  $P(y'_{x'}) + P(y_{x'}) = 0$ , PNS is reduced to the point estimation 1 (i.e.,  $PNS = 1$ ). Besides, the feasible intervals of  $D$  in both theorems come from the general relationship between experimental and observational data (i.e.,  $P(x, y) \leq P(y_x) \leq 1 - P(x, y')$ ) proposed by Tian and Pearl (20).

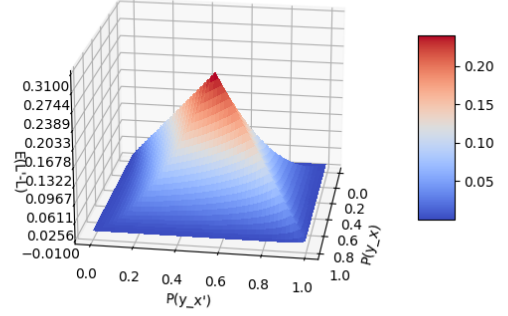


Figure 1:  $(P(y_x), P(y_{x'}))$  v.s. Expected increased lower bound  $E(L' - L)$ .

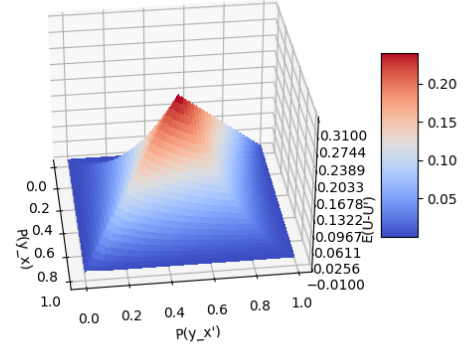


Figure 2:  $(P(y_x), P(y_{x'}))$  v.s. Expected decreased upper bound  $E(U - U')$ .

### 3.1 Visualization of Theorems

In order to have a better understanding of Theorems 4 and 5. We plotted  $(P(y_x), P(y_{x'}))$  v.s.  $E$  graphs as shown in Figures 1 and 2. Both expected increased lower bounds and decreased upper bounds are high when both  $P(y_x)$  and  $P(y_{x'})$  are close to 0.5.

To have a better visualization, we then fixed the value of  $P(y_{x'})$  and plotted the  $P(y_x)$  v.s.  $E$  graphs as shown in Figures 3 and 4. For each of the fixed  $P(y_{x'})$ , the maximum values of expected increased lower bound come from the situation that  $P(y_x) = P(y_{x'})$ , and the maximum values of expected decreased upper bound come from the situation that  $P(y_x) + P(y_{x'}) = 1$ . In addition, same as we concluded from Figures 1 and 2, the overall maximum values of both expected increased lower bound and decreased upper bound reached when  $P(y_x) = P(y_{x'}) = 0.5$ .

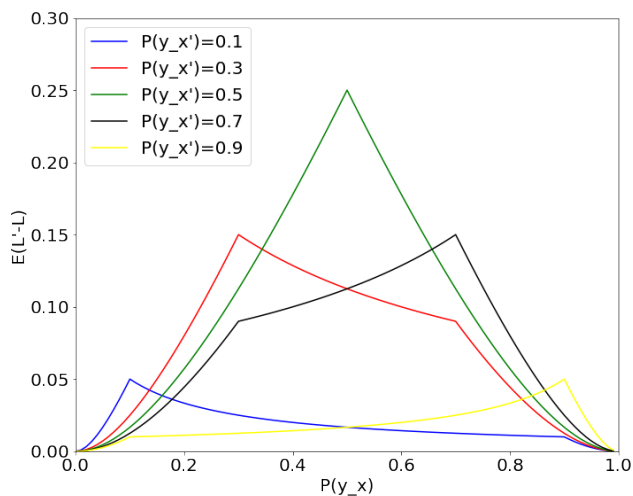


Figure 3:  $P(y_x)$  v.s. Expected increased lower bound  $E(L' - L)$ .

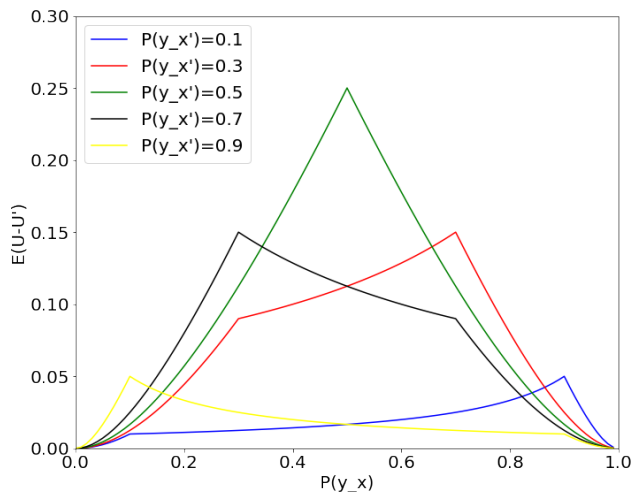


Figure 4:  $P(y_x)$  v.s. Expected decreased upper bound  $E(U - U')$ .

Table 1: Results of an experimental study where 1500 individuals were forced to receive the vaccine and 1500 individuals were forced not to receive the vaccine.

	Vaccinated	Unvaccinated
Unaffected	795	720
Affected	705	780

## 4 EXAMPLES

Here, in this section, we provided two simulated examples of how to apply the proposed theorems.

### 4.1 PNS of Vaccine

Consider a pharmaceutical company that invented a new vaccine for a virus. The company wants to claim the effectiveness of the vaccine by showing that the PNS of the vaccine (i.e., the percentage of individuals who would not be affected by the virus if vaccinated and would be affected if unvaccinated) is high.

Let  $X = x$  denote that an individual received the vaccine, and  $X = x'$  denote that an individual received no vaccine. Let  $Y = y$  denote that an individual is not affected by the virus, and  $Y = y'$  denote that an individual is affected by the virus.

The pharmaceutical company has conducted an experimental study that 1500 individuals were forced to receive the vaccine and 1500 individuals were forced to not (Note that 1500 is the suggested sample size by Li and Pearl (8) to have accuracy estimation of PNS). The results are shown in Table 1.

If frequentist is used for experimental data, we have  $P(y_x) = 0.53$  and  $P(y_{x'}) = 0.48$ . We then plugged the experimental data into Equations 3 and 4, and obtained that  $0.05 \leq PNS \leq 0.52$ . It is hard for the pharmaceutical company to claim that the vaccine is effective because PNS can be as low as 0.05 though the upper bound is 0.52. Thus, they were considering obtaining observational data.

Before that, they plugged the experimental data in to Theorems 4 and 5, and obtained that  $E(L' - L)$  (i.e., Expected increased lower bound) is 0.2231 and  $E(U - U')$  (i.e., Expected decreased upper bound) is 0.2325. The theorem indicates that both lower and upper bounds have non-minor expected improvement. The pharmaceutical company then conducted an observational study with 1500 individuals who have access to the vaccine, where 660 individuals chose to receive the vaccine and 840 individuals chose not to. The results of the observational study are shown in Table 2.

Again, if frequentist is used for observational data, we have  $P(x, y) = 0.14$ ,  $P(x', y) = 0.06$ ,  $P(x, y') = 0.3$ , and  $P(x', y') = 0.5$ . We then plugged both experimental and

Table 2: Results of an observational study with 1500 individuals who have access to the vaccine, where 660 individuals chose to receive the vaccine and 840 individuals chose not to.

	Vaccinated	Unvaccinated
Unaffected	210	90
Affected	450	750

Table 3: Results of an experimental study where 1500 customers were forced to receive the discount and 1500 customers were forced to not.

	Received discount	No discount
Purchased	150	1350
Not purchased	1350	150

observational data into Equations 1 and 2, and obtained that  $0.33 \leq PNS \leq 0.41$ . These bounds are tight enough to conclude that the vaccine is effective because at least 33% of individuals are the ones who would be unaffected by the virus if vaccinated and would be affected if unvaccinated.

## 4.2 PNS of Enticement

A car dealer wants to send an enticement (i.e., discount) to all customers to encourage them to buy a hybrid car. The manager of the car dealer wants to know how many customers are the ones who would buy the hybrid car if they received the discount and would not otherwise. The manager then conducted an experimental study by offering the discount to 1500 of its customers and offering no discount to another 1500 of its customers. The experimental results are shown in Table 3.

Let  $X = x$  denote that a customer received the discount and  $X = x'$  denote that a customer received no discount. Let  $Y = y$  denote that a customer purchased the hybrid car and  $Y = y'$  denote that a customer did not purchase the hybrid car.

The experimental data from Table 3 are  $P(y_x) = 0.1$  and  $P(y_{x'}) = 0.9$ . We can then obtained the bounds of PNS from Equations 3 and 4, where  $0 \leq PNS \leq 0.1$ . In order to convince that there do exist customers who would buy the hybrid car if they received the discount and would not otherwise, the lower bound of PNS should be improved by investigating observational data.

The car manager then applied Theorem 4 and obtained that the expectation of the increased lower bound is  $E(L' - L) = 0.01$ . This means that it is hard to improve the lower bound even combining with observational data; therefore, we suggested that it is no need to conduct an observational study, and the manager should consider other approaches (e.g., obtaining the covariate information (17)) to show there

exist the desired customers.

## 5 APPLICATION TO UNIT SELECTION PROBLEM

The unit selection problem defined by Li and Pearl (13) is to identify individuals who have the desired behavior, for example, the individuals who would have a positive effect if treated and would have a negative effect otherwise.

Let  $X = x$  denote that an individual received the treatment and  $X = x'$  denote that an individual received no treatment. Let  $Y = y$  denote that an individual has a positive effect and  $Y = y'$  denote that an individual has a negative effect.

According to Li and Pearl, individuals are divided into four response types: complier (i.e.,  $(y_x, y'_{x'})$ ), always-taker (i.e.,  $(y_x, y_{x'})$ ), never-taker (i.e.,  $(y'_{x'}, y'_{x'})$ ), and defier (i.e.,  $(y'_{x'}, y_{x'})$ ). Suppose the payoff of selecting a complier, always-taker, never-taker, and defier is  $(\beta, \gamma, \theta, \delta)$ , respectively (i.e., the benefit vector). The objective function defined by Li and Pearl (i.e., the benefit function) is then (13)

$$f(c) = \beta P(y_x, y'_{x'}|c) + \gamma P(y_x, y_{x'}|c) + \theta P(y'_{x'}, y'_{x'}|c) + \delta P(y'_{x'}, y_{x'}|c).$$

Note that  $c$  is the population-specific characteristics and the benefit function is a linear combination of the probabilities of causation; therefore, Li and Pearl derived the tight bounds of the benefit function as follows:

**Theorem 6.** *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 (13):*

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

where  $\sigma, W, L, U$  are given by,

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

The above bounds of the benefit function used a combination of experimental and observational data. If the observational

data is unavailable, then the bounds of the benefit function become

**Theorem 7.** *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 (5):*

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

where  $\sigma, W, L, U$  are given by,

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

### 5.1 Role of Observational Data in Unit Selection Problem

The benefit function is a linear combination of probabilities of causation; thus, our Theorems 4 and 5 could be easily extended to the benefit function as follows:

**Corollary 8.** *Given a causal diagram  $G$  and distribution compatible with  $G$  with experimental data  $P(y_x|c)$  and  $P(y_{x'}|c)$ , let  $C$  be a set of variables that does not contain any descendant of  $X$  in  $G$ . let  $D = P(y|c)$  be a random variable, and let  $D' = P(x, y|c) + P(x', y'|c)$  be another random variable. Let  $LB, UB$  be the lower and upper bound of the benefit function using pure experimental data, respectively. Let  $LB', UB'$  be the lower and upper bound of the benefit function using a combination of experimental and observational data, respectively. If  $D$  is uniformly distributed on its feasible interval  $[\max(0, P(y_x|c) - P(y'_{x'}|c)), \min(1, P(y_x|c) + P(y_{x'}|c))]$  and  $D'$  is uniformly distributed on its feasible interval  $[\max(0, P(y_x|c) - P(y_{x'}|c)), \min(1, P(y_x) + P(y'_{x'}|c))]$ , and  $P(y_x|c) + P(y'_{x'}|c) \neq 0$  and  $P(y'_{x'}|c) + P(y_{x'}|c) \neq 0$  and  $P(y_x|c) + P(y_{x'}|c) \neq 0$  and  $P(y'_{x'}|c) + P(y'_{x'}|c) \neq 0$ , then we have the expectation of the increased lower bound  $E(LB' - LB)$  and the decreased*

upper bound  $E(UB - UB')$  as follows:

$$\begin{aligned} &E(LB' - LB) \\ &= \sigma \frac{\min\{P^2(y_x|c), P^2(y'_{x'}|c), P^2(y_{x'}|c), P^2(y'_{x'}|c)\}}{\min\{P(y_x|c) + P(y_{x'}|c), P(y'_{x'}|c) + P(y'_{x'}|c)\}} \\ &\quad \text{if } \sigma > 0, \\ &E(LB' - LB) \\ &= -\sigma \frac{\min\{P^2(y_x|c), P^2(y'_{x'}|c), P^2(y_{x'}|c), P^2(y'_{x'}|c)\}}{\min\{P(y_x|c) + P(y'_{x'}|c), P(y'_{x'}|c) + P(y_{x'}|c)\}} \\ &\quad \text{if } \sigma < 0, \\ &E(UB - UB') \\ &= \sigma \frac{\min\{P^2(y_x|c), P^2(y'_{x'}|c), P^2(y_{x'}|c), P^2(y'_{x'}|c)\}}{\min\{P(y_x|c) + P(y'_{x'}|c), P(y'_{x'}|c) + P(y_{x'}|c)\}} \\ &\quad \text{if } \sigma > 0, \\ &E(UB - UB') \\ &= -\sigma \frac{\min\{P^2(y_x|c), P^2(y'_{x'}|c), P^2(y_{x'}|c), P^2(y'_{x'}|c)\}}{\min\{P(y_x|c) + P(y_{x'}|c), P(y'_{x'}|c) + P(y'_{x'}|c)\}} \\ &\quad \text{if } \sigma < 0, \end{aligned}$$

where,

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

$$W = (\gamma - \delta)P(y_x|c) + \delta P(y_{x'}|c) + \theta P(y'_{x'}|c),$$

$$P(y'_{x'}|c) = 1 - P(y_x|c),$$

$$P(y'_{x'}|c) = 1 - P(y_{x'}|c),$$

$$LB = W + \sigma U \quad \text{if } \sigma < 0,$$

$$LB = W + \sigma L \quad \text{if } \sigma > 0,$$

$$LB' = W + \sigma U' \quad \text{if } \sigma < 0,$$

$$LB' = W + \sigma L' \quad \text{if } \sigma > 0,$$

$$UB = W + \sigma L \quad \text{if } \sigma < 0,$$

$$UB = W + \sigma U \quad \text{if } \sigma > 0,$$

$$UB' = W + \sigma L' \quad \text{if } \sigma < 0,$$

$$UB' = W + \sigma U' \quad \text{if } \sigma > 0,$$

$$L' = \max \left\{ \begin{array}{l} 0, \\ P(y_x|c) - P(y_{x'}|c), \\ D - P(y_{x'}|c), \\ P(y_x|c) - D \end{array} \right\},$$

$$L = \max \left\{ \begin{array}{l} 0, \\ P(y_x|c) - P(y_{x'}|c) \end{array} \right\},$$

$$U' = \min \left\{ \begin{array}{l} P(y_x|c), \\ P(y'_{x'}|c), \\ D', \\ P(y_x|c) + P(y'_{x'}|c) - D' \end{array} \right\},$$

$$U = \min \left\{ \begin{array}{l} P(y_x|c), \\ P(y'_{x'}|c) \end{array} \right\}.$$

Note that  $\sigma = 0$  is the case called Gain Equality (13) where

Table 4: Results of an experimental study where 1500 customers were forced to receive the discount and 1500 customers were forced to not.

	Received discount	No discount
Purchased	825	600
Not purchased	675	900

the benefit function is reduced to a point estimation with purely experimental data. The  $\sigma$  here can be interpreted as the improvement amplifier. This corollary looks more complicated than Theorems 4 and 5, but the usage is exactly the same as the Theorems 4 and 5. We will illustrate how to apply this corollary in the next section.

## 5.2 Nonimmediate Profit

Consider a car dealer again who wants to send an enticement (i.e., discount) to all customers to increase its total nonimmediate profit. The manager acknowledged that the payoff of sending the discount to a complier (i.e., the customer who would buy the hybrid car if they received the discount and would not otherwise) is \$1500 as the profit of selling a hybrid car is \$2000, but the discount is \$500. The payoff of sending the discount to an always-taker (i.e., the customer who would buy the hybrid car no matter whether or not they received the discount) is  $-\$800$ . First, the dealer loses the cost of the discount, \$500. Second, the customer may require a discount in the future (i.e., the offer changed the customer’s response type, and the manager assessed that this nonimmediate cost is \$300). The payoff of sending the discount to a never-taker (i.e., the customer who would not buy the hybrid car no matter whether or not they received the discount) is \$0, as the cost of sending the discount is negligible. The payoff of sending the discount to a defier (i.e., the customer who would buy the hybrid car if they received no discount and would not otherwise) is  $-\$2000$  as the car dealer loses one customer due to the discount.

Again, let  $X = x$  denote that a customer received the discount and  $X = x'$  denote that a customer received no discount. Let  $Y = y$  denote that a customer purchased the hybrid car and  $Y = y'$  denote that a customer did not purchase the hybrid car. Note that the population-specific characteristics  $c$  here simply mean the entire customer.

The benefit vector is then  $(1500, -800, 0, -2000)$ . The manager conducted an experimental study with 3000 randomly selected customers. They sent the discount to 1500 of the selected customers and sent no discount to the rest 1500 of the selected customers. The results are shown in Table 4.

The experimental data are then  $P(y_x|c) = 0.55$  and  $P(y_{x'}|c) = 0.4$ . We then obtained that  $-95 \leq f(c) \leq 20$  by Theorem 7 using the experimental data. These results

Table 5: Results of an observational study with 1500 customers who have access to the discount, where 795 customers chose to receive the discount and 705 customers chose not to.

	Chose discount	No discount
Purchased	450	30
Not purchased	345	675

are hard for the decision-maker because the average gain profit of sending the discount to a customer is across the 0 (i.e., either gain profit or lose profit).

An observational study was then considered by the manager; therefore, the manager applied the Corollary 8 and obtained that  $E(LB' - LB) = 50.53$  and  $E(UB - UB') = 56.47$ , where the expected new upper bound of the benefit function may smaller than 0.

The manager then conducted an observational study where 1500 randomly selected customers had access to the discount and were free to choose to receive it or not. 795 of the selected customers chose to receive the discount, and 705 of the selected customers decided not to. The results are shown in Table 5.

The observational data are then  $P(x, y) = 0.3$ ,  $P(x', y) = 0.02$ ,  $P(x, y') = 0.23$ , and  $P(x', y') = 0.45$ . The bounds of the benefit function obtained from Theorem 6 is then  $-71 \leq f(c) \leq -20$ , where the upper bound is smaller than 0. The conclusion is then not to send the discount because the average profit gained per customer by sending the discount is negative (i.e., the car dealer will lose the profit if sending the discount to all customers).

## 6 DISCUSSION

The paper provides theorems that quantify the degree of improvement one can expect from observational data, when considering the bounds of PNS. Here, we further discuss two additional topics related to the theorems provided.

First, while this paper answers the question of whether the observational studies are worth conducting, the symmetrical problem can also be answered. If experimental data are unavailable, the bounds of PNS are  $0 \leq PNS \leq P(x, y) + P(x', y')$  which are less informative as discussed in (5). In such cases, experimental data should be investigated either by random controlled trial or by adjustment formula (18).

Second, the assumptions in Theorems 4 and 5 are  $P(y)$  and  $P(x, y) + P(x', y')$  are uniformly distributed on their feasible interval. We know that given large enough observational samples of  $(X, Y)$ , the sample estimation of a fixed  $P(y)$  (or  $P(x, y) + P(x', y')$ ) is normally distributed with the center at  $P(y)$  (or  $P(x, y) + P(x', y')$ ). However, the distribution of  $P(y)$  (or  $P(x, y) + P(x', y')$ ) is unknown; one

can interpret that this distribution is among all compatible models in the world (e.g.,  $P(y) = 0.1$  in one model and  $P(y) = 0.5$  in another model). If there is any progress on these distributions, the uniform distribution assumptions in both theorems can be extended.

## 7 CONCLUSION

In this work, we formally defined the expectation of improvement of the bounds of PNS when considering observational data. The proposed theorems are extendable to other probabilities of causation. Simulated examples showed that the proposed theorems should be applied before conducting any observational study. We further expanded the proposed theorems to the bounds of the benefit function in the unit selection problem.

### Acknowledgements

This research was supported in parts by grants from the National Science Foundation [#IIS-2106908 and #IIS-2231798], Office of Naval Research [#N00014-21-1-2351], and Toyota Research Institute of North America [#PO-000897].

### References

- [1] Alexander A Balke and Judea Pearl. Probabilistic counterfactuals: Semantics, computation, and applications. Technical report, UCLA Dept. of Computer Science, 1997.
- [2] Philip Dawid, Monica Musio, and Rossella Murtas. The probability of causation. *Law, Probability and Risk*, (16):163–179, 2017.
- [3] David Galles and Judea Pearl. An axiomatic characterization of causal counterfactuals. *Foundations of Science*, 3(1):151–182, 1998.
- [4] Joseph Y Halpern. Axiomatizing causal reasoning. *Journal of Artificial Intelligence Research*, 12:317–337, 2000.
- [5] A. Li. Unit selection based on counterfactual logic. Technical Report R-507, <[http://ftp.cs.ucla.edu/pub/stat\\_ser/r507.pdf](http://ftp.cs.ucla.edu/pub/stat_ser/r507.pdf)>, Ph.D. Thesis, Department of Computer Science, University of California, Los Angeles, CA, 2021.
- [6] A. Li, S. Jiang, Y. Sun, and J. Pearl. Learning probabilities of causation from finite population data. Technical Report R-519, <[http://ftp.cs.ucla.edu/pub/stat\\_ser/r519.pdf](http://ftp.cs.ucla.edu/pub/stat_ser/r519.pdf)>, Department of Computer Science, University of California, Los Angeles, CA, 2022.
- [7] A. Li, S. Jiang, Y. Sun, and J. Pearl. Unit selection: Learning benefit function from finite population data. Technical Report R-520, <[http://ftp.cs.ucla.edu/pub/stat\\_ser/r520.pdf](http://ftp.cs.ucla.edu/pub/stat_ser/r520.pdf)>, Department of Computer Science, University of California, Los Angeles, CA, 2022.
- [8] A. Li, R. Mao, and J. Pearl. Probabilities of causation: Adequate size of experimental and observational samples. Technical Report R-518, <[http://ftp.cs.ucla.edu/pub/stat\\_ser/r518.pdf](http://ftp.cs.ucla.edu/pub/stat_ser/r518.pdf)>, Department of Computer Science, University of California, Los Angeles, CA, 2022.
- [9] A. Li and J. Pearl. Probabilities of causation with non-binary treatment and effect. Technical Report R-516, Department of Computer Science, University of California, Los Angeles, CA, 2022.
- [10] A. Li and J. Pearl. Unit selection: Case study and comparison with a/b test heuristic. Technical Report R-521, <[http://ftp.cs.ucla.edu/pub/stat\\_ser/r521.pdf](http://ftp.cs.ucla.edu/pub/stat_ser/r521.pdf)>, Department of Computer Science, University of California, Los Angeles, CA, 2022.
- [11] A. Li and J. Pearl. Unit selection with nonbinary treatment and effect. Technical Report R-517, <[http://ftp.cs.ucla.edu/pub/stat\\_ser/r517.pdf](http://ftp.cs.ucla.edu/pub/stat_ser/r517.pdf)>, Department of Computer Science, University of California, Los Angeles, CA, 2022.
- [12] Ang Li, Suming J. Chen, Jingzheng Qin, and Zhen Qin. Training machine learning models with causal logic. In *Companion Proceedings of the Web Conference 2020*, pages 557–561, 2020.
- [13] Ang Li and Judea Pearl. Unit selection based on counterfactual logic. In *Proceedings of the Twenty-Eighth International Joint Conference on Artificial Intelligence, IJCAI-19*, pages 1793–1799. International Joint Conferences on Artificial Intelligence Organization, 7 2019.
- [14] Ang Li and Judea Pearl. Bounds on causal effects and application to high dimensional data. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 36, pages 5773–5780, 2022.
- [15] Ang Li and Judea Pearl. Unit selection with causal diagram. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 36, pages 5765–5772, 2022.
- [16] Mueller and Pearl. Personalized decision making – a conceptual introduction. Technical Report R-513, Department of Computer Science, University of California, Los Angeles, CA, 2022.
- [17] S. Mueller, A. Li, and J. Pearl. Causes of effects: Learning individual responses from population data. Technical Report R-505, <[http://ftp.cs.ucla.edu/pub/stat\\_ser/r505.pdf](http://ftp.cs.ucla.edu/pub/stat_ser/r505.pdf)>, Department of Computer Science, University of California, Los Angeles, CA, 2021. Forthcoming, Proceedings of IJCAI-2022.



- [18] J Pearl. Aspects of graphical models connected with causality. *Proceedings of the 49th Session of the international Statistical Institute, Italy*, pages 399–401, 1993.
- [19] Judea Pearl. Probabilities of causation: Three counterfactual interpretations and their identification. *Synthese*, pages 93–149, 1999.
- [20] Jin Tian and Judea Pearl. Probabilities of causation: Bounds and identification. *Annals of Mathematics and Artificial Intelligence*, 28(1-4):287–313, 2000.

## 8 APPENDIX

### 8.1 Proof of Theorem 4

**Theorem 4.** *Given experimental data  $P(y_x)$  and  $P(y_{x'})$  and let  $D = P(y)$  be a random variable. Let  $L$  be the lower bound of PNS using pure experimental data and  $L'$  be the lower bound of PNS using a combination of experimental and observational data. If  $D$  is uniformly distributed on its feasible interval  $[\max(0, P(y_x) - P(y'_{x'})), \min(1, P(y_x) + P(y_{x'}))]$ , and  $P(y_x) + P(y_{x'}) \neq 0$  and  $P(y'_x) + P(y'_{x'}) \neq 0$ , then we have the expectation of the increased lower bound  $E(L' - L)$  as follows:*

$$E(L' - L) = \frac{\min\{P^2(y_x), P^2(y'_x), P^2(y_{x'}), P^2(y'_{x'})\}}{\min\{P(y_x) + P(y_{x'}), P(y'_x) + P(y'_{x'})\}}$$

where,

$$P(y'_x) = 1 - P(y_x),$$

$$P(y'_{x'}) = 1 - P(y_{x'}),$$

$$L' = \max \left\{ \begin{array}{c} 0, \\ P(y_x) - P(y_{x'}), \\ D - P(y_{x'}), \\ P(y_x) - D \end{array} \right\},$$

$$L = \max \left\{ \begin{array}{c} 0, \\ P(y_x) - P(y_{x'}) \end{array} \right\}.$$

*Proof.* First, by Tian and Pearl (20), we have,

$$P(x, y) \leq P(y_x) \leq 1 - P(x, y'),$$

$$P(x', y) \leq P(y_{x'}) \leq 1 - P(x', y').$$

Therefore,

$$P(y_x) + P(y_{x'}) \geq P(x, y) + P(x', y) = P(y),$$

and,

$$P(y_x) + P(y_{x'}) \leq 2 - P(x, y') - P(x', y')$$

$$= 2 - P(y')$$

$$= 1 + P(y),$$

$$P(y) \geq P(y_x) + P(y_{x'}) - 1$$

$$= P(y_x) - P(y'_{x'}).$$

Thus, the feasible space of  $D = P(y)$  is

$$D \geq \max\{0, P(y_x) - P(y'_{x'})\},$$

$$D \leq \min\{1, P(y_x) + P(y_{x'})\}.$$

Case 1:  $P(y_x) \geq P(y_{x'})$  and  $P(y_x) + P(y_{x'}) \geq 1$ .  
Thus,

$$P(y_x) + P(y_{x'}) - 1 \geq 0,$$

$$P(y_x) - P(y'_{x'}) \geq 0.$$

Therefore, the feasible space of  $D$  is

$$P(y_x) - P(y'_{x'}) \leq D \leq 1.$$

and  $D$  uniformly distributed on  $[P(y_x) - P(y'_{x'}), 1]$ . Also,  $L = \max\{0, P(y_x) - P(y_{x'})\} = P(y_x) - P(y_{x'})$ , and  $L' \geq P(y_x) - P(y_{x'})$ . In order to make  $L' > L$ ,  $D$  has to be in interval  $[P(y_x) - P(y'_{x'}), P(y_{x'})]$  or  $(P(y_x), 1]$ . We also know that,

$$1 - P(y_x) = P(y_{x'}) - P(y_x) + P(y'_{x'}) = P(y'_{x'}),$$

therefore, the maximum value of  $L' - L$  is  $P(y'_{x'})$ , and  $L' - L \leq z \leq P(y'_{x'})$ , when  $D$  in interval  $[P(y_{x'}) - z, P(y_{x'})]$  or  $[P(y_x), P(y_x) + z]$ .

Thus, we have,

$$P(L' - L \leq z) = \begin{cases} 0, & \text{if } z < 0, \\ \frac{P(y_x) - P(y_{x'}) + 2z}{P(y'_{x'}) + P(y'_{x'})}, & \text{if } 0 \leq z < P(y'_{x'}), \\ 1, & \text{if } z \geq P(y'_{x'}). \end{cases}$$

Then, the probability density function  $f$  is

$$f(z) = \begin{cases} 0, & \text{if } z < 0, \\ \frac{2}{P(y'_{x'}) + P(y'_{x'})}, & \text{if } 0 \leq z < P(y'_{x'}), \\ 0, & \text{if } z \geq P(y'_{x'}). \end{cases}$$

Then, we have,

$$E(L' - L) = \int_0^{P(y'_{x'})} \frac{2z}{P(y'_{x'}) + P(y'_{x'})} dz$$

$$= \frac{P^2(y'_{x'})}{P(y'_{x'}) + P(y'_{x'})}.$$

Case 2:  $P(y_x) \geq P(y_{x'})$  and  $P(y_x) + P(y_{x'}) < 1$ .

Thus,

$$P(y_x) + P(y_{x'}) - 1 < 0,$$

$$P(y_x) - P(y'_{x'}) < 0.$$

Therefore, the feasible space of  $D$  is

$$0 \leq D \leq P(y_x) + P(y_{x'}).$$

and  $D$  uniformly distributed on  $[0, P(y_x) + P(y_{x'})]$ . Also,  $L = \max\{0, P(y_x) - P(y_{x'})\} = P(y_x) - P(y_{x'})$ , and  $L' \geq P(y_x) - P(y_{x'})$ . In order to make  $L' > L$ ,  $D$  has

to be in interval  $[0, P(y_{x'})]$  or  $(P(y_x), P(y_x) + P(y_{x'}))$ .  
We also know that,

$$P(y_{x'}) - 0 = P(y_x) + P(y_{x'}) - P(y_x) = P(y_{x'}),$$

therefore, the maximum value of  $L' - L$  is  $P(y_{x'})$ ,  
and  $L' - L \leq z \leq P(y_{x'})$ , when  $D$  in interval  $[P(y_{x'}) - z, P(y_{x'})]$  or  $[P(y_x), P(y_x) + z]$ .

Thus, we have,

$$P(L' - L \leq z) = \begin{cases} 0, & \text{if } z < 0, \\ \frac{P(y_x) - P(y_{x'}) + 2z}{P(y_x) + P(y_{x'})}, & \text{if } 0 \leq z < P(y_{x'}), \\ 1, & \text{if } z \geq P(y_{x'}). \end{cases}$$

Then, the probability density function  $f$  is

$$f(z) = \begin{cases} 0, & \text{if } z < 0, \\ \frac{2}{P(y_x) + P(y_{x'})}, & \text{if } 0 \leq z < P(y_{x'}), \\ 0, & \text{if } z \geq P(y_{x'}). \end{cases}$$

Then, we have,

$$\begin{aligned} E(L' - L) &= \int_0^{P(y_{x'})} \frac{2z}{P(y_x) + P(y_{x'})} dz \\ &= \frac{P^2(y_{x'})}{P(y_x) + P(y_{x'})}. \end{aligned}$$

Case 3:  $P(y_x) < P(y_{x'})$  and  $P(y_x) + P(y_{x'}) \geq 1$ .

Thus,

$$\begin{aligned} P(y_x) + P(y_{x'}) - 1 &\geq 0, \\ P(y_x) - P(y'_{x'}) &\geq 0. \end{aligned}$$

Therefore, the feasible space of  $D$  is

$$P(y_x) - P(y'_{x'}) \leq D \leq 1.$$

and  $D$  uniformly distributed on  $[P(y_x) - P(y'_{x'}), 1]$ .

Also,  $L = \max\{0, P(y_x) - P(y_{x'})\} = 0$ ,

and  $L' \geq 0$ . In order to make  $L' > L$ ,  $D$  has to be in interval  $[P(y_x) - P(y'_{x'}), P(y_x)]$  or  $(P(y_{x'}), 1]$ .

We also know that,

$$1 - P(y_{x'}) = P(y_x) - P(y_x) + P(y'_{x'}) = P(y'_{x'}),$$

therefore, the maximum value of  $L' - L$  is  $P(y'_{x'})$ ,  
and  $L' - L \leq z \leq P(y'_{x'})$ , when  $D$  in interval  $[P(y_x) - z, P(y_x)]$  or  $[P(y_{x'}), P(y_{x'}) + z]$ .

Thus, we have,

$$P(L' - L \leq z) = \begin{cases} 0, & \text{if } z < 0, \\ \frac{P(y_{x'}) - P(y_x) + 2z}{P(y'_{x'}) + P(y'_{x'})}, & \text{if } 0 \leq z < P(y'_{x'}), \\ 1, & \text{if } z \geq P(y'_{x'}). \end{cases}$$

Then, the probability density function  $f$  is

$$f(z) = \begin{cases} 0, & \text{if } z < 0, \\ \frac{2}{P(y'_{x'}) + P(y'_{x'})}, & \text{if } 0 \leq z < P(y'_{x'}), \\ 0, & \text{if } z \geq P(y'_{x'}). \end{cases}$$

Then, we have,

$$\begin{aligned} E(L' - L) &= \int_0^{P(y'_{x'})} \frac{2z}{P(y'_{x'}) + P(y'_{x'})} dz \\ &= \frac{P^2(y'_{x'})}{P(y'_{x'}) + P(y'_{x'})}. \end{aligned}$$

Case 4:  $P(y_x) < P(y_{x'})$  and  $P(y_x) + P(y_{x'}) < 1$ .

Thus,

$$\begin{aligned} P(y_x) + P(y_{x'}) - 1 &< 0, \\ P(y_x) - P(y'_{x'}) &< 0. \end{aligned}$$

Therefore, the feasible space of  $D$  is

$$0 \leq D \leq P(y_x) + P(y_{x'}).$$

and  $D$  uniformly distributed on  $[0, P(y_x) + P(y_{x'})]$ .

Also,  $L = \max\{0, P(y_x) - P(y_{x'})\} = 0$ ,

and  $L' \geq 0$ . In order to make  $L' > L$ ,  $D$  has to be in interval  $[0, P(y_x)]$  or  $(P(y_{x'}), P(y_x) + P(y_{x'}))$ .

We also know that,

$$P(y_x) - 0 = P(y_x) + P(y_{x'}) - P(y_{x'}) = P(y_x),$$

therefore, the maximum value of  $L' - L$  is  $P(y_x)$ ,  
and  $L' - L \leq z \leq P(y_x)$ , when  $D$  in interval  $[P(y_x) - z, P(y_x)]$  or  $[P(y_{x'}), P(y_{x'}) + z]$ .

Thus, we have,

$$P(L' - L \leq z) = \begin{cases} 0, & \text{if } z < 0, \\ \frac{P(y_{x'}) - P(y_x) + 2z}{P(y_x) + P(y_{x'})}, & \text{if } 0 \leq z < P(y_x), \\ 1, & \text{if } z \geq P(y_x). \end{cases}$$

Then, the probability density function  $f$  is

$$f(z) = \begin{cases} 0, & \text{if } z < 0, \\ \frac{2}{P(y_x) + P(y_{x'})}, & \text{if } 0 \leq z < P(y_x), \\ 0, & \text{if } z \geq P(y_x). \end{cases}$$

Then, we have,

$$\begin{aligned} E(L' - L) &= \int_0^{P(y_x)} \frac{2z}{P(y_x) + P(y_{x'})} dz \\ &= \frac{P^2(y_x)}{P(y_x) + P(y_{x'})}. \end{aligned}$$

We then combine the results of cases 1 and 3.  
 Note that when  $P(y_x) \geq P(y_{x'})$ , we have  $P(y'_{x'}) \leq P(y'_{x'})$ ,  
 and when  $P(y_x) < P(y_{x'})$ , we have  $P(y'_{x'}) > P(y'_{x'})$ ,  
 therefore,  
 when  $P(y_x) + P(y_{x'}) \geq 1$ , we have,

$$\begin{aligned} & E(L' - L) \\ &= \frac{\min\{P^2(y'_{x'}), P^2(y'_{x'})\}}{P(y'_{x'}) + P(y'_{x'})}. \end{aligned}$$

Similarly, after combine the results of cases 2 and 4, we have,  
 when  $P(y_x) + P(y_{x'}) < 1$ , we have,

$$\begin{aligned} & E(L' - L) \\ &= \frac{\min\{P^2(y_x), P^2(y_{x'})\}}{P(y_x) + P(y_{x'})}. \end{aligned}$$

Now, note that when  $P(y_x) + P(y_{x'}) \geq 1$ , we have,  
 $P(y_x) + P(y_{x'}) \geq P(y'_{x'}) + P(y'_{x'})$ .  
 Also,  $P(y_x) \geq P(y'_{x'})$  and  $P(y_{x'}) \geq P(y'_{x'})$ ,  
 thus,  $\min\{P^2(y_x), P^2(y_{x'})\} \geq \min\{P^2(y'_{x'}), P^2(y'_{x'})\}$ .  
 Similarly, when  $P(y_x) + P(y_{x'}) < 1$ , we have,  
 $P(y_x) + P(y_{x'}) < P(y'_{x'}) + P(y'_{x'})$ ,  
 and,  $\min\{P^2(y_x), P^2(y_{x'})\} < \min\{P^2(y'_{x'}), P^2(y'_{x'})\}$ .  
 Thus, finally, we have,

$$\begin{aligned} & E(L' - L) \\ &= \frac{\min\{P^2(y_x), P^2(y'_{x'}), P^2(y_{x'}), P^2(y'_{x'})\}}{\min\{P(y_x) + P(y_{x'}), P(y'_{x'}) + P(y'_{x'})\}}. \end{aligned}$$

□

## 8.2 Proof of Theorem 5

**Theorem 5.** *Given experimental data  $P(y_x)$  and  $P(y_{x'})$  and let  $D = P(x, y) + P(x', y')$  be a random variable. Let  $U$  be the upper bound of PNS using pure experimental data and  $U'$  be the upper bound of PNS using a combination of experimental and observational data. If  $D$  is uniformly distributed on its feasible interval  $[\max(0, P(y_x) - P(y_{x'})), \min(1, P(y_x) + P(y'_{x'}))]$ , and  $P(y_x) + P(y'_{x'}) \neq 0$  and  $P(y'_{x'}) + P(y_{x'}) \neq 0$ , then we have the expectation of the decreased upper bound  $E(U - U')$  as follows:*

$$\begin{aligned} & E(U - U') \\ &= \frac{\min\{P^2(y_x), P^2(y'_{x'}), P^2(y_{x'}), P^2(y'_{x'})\}}{\min\{P(y_x) + P(y'_{x'}), P(y'_{x'}) + P(y_{x'})\}} \end{aligned}$$

where,

$$\begin{aligned} P(y'_{x'}) &= 1 - P(y_x), \\ P(y'_{x'}) &= 1 - P(y_{x'}), \end{aligned}$$

$$U' = \min \left\{ \begin{array}{c} P(y_x), \\ P(y'_{x'}), \\ D, \\ P(y_x) + P(y'_{x'}) - D \end{array} \right\},$$

$$U = \min \left\{ \begin{array}{c} P(y_x), \\ P(y'_{x'}) \end{array} \right\}.$$

*Proof.* First, by Tian and Pearl (20), we have,

$$\begin{aligned} P(x, y) &\leq P(y_x) \leq 1 - P(x, y'), \\ P(x', y') &\leq P(y'_{x'}) \leq 1 - P(x', y). \end{aligned}$$

Therefore,

$$P(y_x) + P(y'_{x'}) \geq P(x, y) + P(x', y'),$$

and,

$$\begin{aligned} P(y_x) + P(y'_{x'}) &\leq 2 - P(x, y') - P(x', y) \\ &= 1 + P(x, y) + P(x', y'), \\ P(x, y) + P(x', y') &\geq P(y_x) + P(y'_{x'}) - 1 \\ &= P(y_x) - P(y_{x'}). \end{aligned}$$

Thus, the feasible space of  $D = P(x, y) + P(x', y')$  is

$$\begin{aligned} D &\geq \max\{0, P(y_x) - P(y_{x'})\}, \\ D &\leq \min\{1, P(y_x) + P(y'_{x'})\}. \end{aligned}$$

Case 1:  $P(y_x) \geq P(y'_{x'})$  and  $P(y_x) + P(y'_{x'}) \geq 1$ .  
 Thus,

$$\begin{aligned} P(y_x) + P(y'_{x'}) - 1 &\geq 0, \\ P(y_x) - P(y_{x'}) &\geq 0. \end{aligned}$$

Therefore, the feasible space of  $D$  is

$$P(y_x) - P(y_{x'}) \leq D \leq 1.$$

and  $D$  uniformly distributed on  $[P(y_x) - P(y_{x'}), 1]$ .

Also,  $U = \min\{P(y_x), P(y'_{x'})\} = P(y'_{x'})$ ,  
 and  $U' \leq P(y'_{x'})$ . In order to make  $U' < U$ ,  $D$  has to be in interval  $[P(y_x) - P(y_{x'}), P(y'_{x'})]$  or  $(P(y_x), 1]$ .

We also know that,

$$1 - P(y_x) = P(y'_{x'}) - P(y_x) + P(y_{x'}) = P(y'_{x'}),$$

therefore, the maximum value of  $U - U'$  is  $P(y'_{x'})$ ,  
 and  $U - U' \leq z \leq P(y'_{x'})$ , when  $D$  in interval  $[P(y'_{x'}), z, P(y'_{x'})]$  or  $[P(y_x), P(y_x) + z]$ .

Thus, we have,

$$\begin{aligned} & P(U - U' \leq z) \\ &= \left\{ \begin{array}{ll} 0, & \text{if } z < 0, \\ \frac{P(y_x) - P(y'_{x'}) + 2z}{P(y'_{x'}) + P(y_{x'})}, & \text{if } 0 \leq z < P(y'_{x'}), \\ 1, & \text{if } z \geq P(y'_{x'}). \end{array} \right\}, \end{aligned}$$

Then, the probability density function  $f$  is

$$f(z) = \begin{cases} 0, & \text{if } z < 0, \\ \frac{2}{P(y'_x) + P(y_x)}, & \text{if } 0 \leq z < P(y'_x), \\ 0, & \text{if } z \geq P(y'_x). \end{cases}$$

Then, we have,

$$\begin{aligned} & E(U - U') \\ &= \int_0^{P(y'_x)} \frac{2z}{P(y'_x) + P(y_x)} dz \\ &= \frac{P^2(y'_x)}{P(y'_x) + P(y_x)}. \end{aligned}$$

Case 2:  $P(y_x) \geq P(y'_x)$  and  $P(y_x) + P(y'_x) < 1$ .

Thus,

$$\begin{aligned} P(y_x) + P(y'_x) - 1 &< 0, \\ P(y_x) - P(y'_x) &< 0. \end{aligned}$$

Therefore, the feasible space of  $D$  is

$$0 \leq D \leq P(y_x) + P(y'_x).$$

and  $D$  uniformly distributed on  $[0, P(y_x) + P(y'_x)]$ .

Also,  $U = \min\{P(y_x), P(y'_x)\} = P(y'_x)$ , and  $U' < P(y'_x)$ . In order to make  $U' < U$ ,  $D$  has to be in interval  $[0, P(y'_x)]$  or  $(P(y_x), P(y_x) + P(y'_x)]$ .

We also know that,

$$P(y'_x) - 0 = P(y_x) + P(y'_x) - P(y_x) = P(y'_x),$$

therefore, the maximum value of  $U - U'$  is  $P(y'_x)$ , and  $U - U' \leq z \leq P(y'_x)$ , when  $D$  in interval  $[P(y'_x), P(y'_x) - z, P(y'_x)]$  or  $[P(y_x), P(y_x) + z]$ .

Thus, we have,

$$P(U - U' \leq z) = \begin{cases} 0, & \text{if } z < 0, \\ \frac{P(y_x) - P(y'_x) + 2z}{P(y_x) + P(y'_x)}, & \text{if } 0 \leq z < P(y'_x), \\ 1, & \text{if } z \geq P(y'_x). \end{cases}$$

Then, the probability density function  $f$  is

$$f(z) = \begin{cases} 0, & \text{if } z < 0, \\ \frac{2}{P(y_x) + P(y'_x)}, & \text{if } 0 \leq z < P(y'_x), \\ 0, & \text{if } z \geq P(y'_x). \end{cases}$$

Then, we have,

$$\begin{aligned} & E(U - U') \\ &= \int_0^{P(y'_x)} \frac{2z}{P(y_x) + P(y'_x)} dz \\ &= \frac{P^2(y'_x)}{P(y_x) + P(y'_x)}. \end{aligned}$$

Case 3:  $P(y_x) < P(y'_x)$  and  $P(y_x) + P(y'_x) \geq 1$ . Thus,

$$\begin{aligned} P(y_x) + P(y'_x) - 1 &\geq 0, \\ P(y_x) - P(y'_x) &\geq 0. \end{aligned}$$

Therefore, the feasible space of  $D$  is

$$P(y_x) - P(y'_x) \leq D \leq 1.$$

and  $D$  uniformly distributed on  $[P(y_x) - P(y'_x), 1]$ .

Also,  $U = \min\{P(y_x), P(y'_x)\} = P(y_x)$ , and  $U' \leq P(y_x)$ . In order to make  $U' < U$ ,  $D$  has to be in interval  $[P(y_x) - P(y'_x), P(y_x)]$  or  $(P(y'_x), 1]$ .

We also know that,

$$1 - P(y'_x) = P(y_x) - P(y_x) + P(y'_x) = P(y'_x),$$

therefore, the maximum value of  $U - U'$  is  $P(y'_x)$ ,

and  $U - U' \leq z \leq P(y'_x)$ , when  $D$  in interval  $[P(y_x) - z, P(y_x)]$  or  $[P(y'_x), P(y'_x) + z]$ .

Thus, we have,

$$P(U - U' \leq z) = \begin{cases} 0, & \text{if } z < 0, \\ \frac{P(y'_x) - P(y_x) + 2z}{P(y'_x) + P(y_x)}, & \text{if } 0 \leq z < P(y'_x), \\ 1, & \text{if } z \geq P(y'_x). \end{cases}$$

Then, the probability density function  $f$  is

$$f(z) = \begin{cases} 0, & \text{if } z < 0, \\ \frac{2}{P(y'_x) + P(y_x)}, & \text{if } 0 \leq z < P(y'_x), \\ 0, & \text{if } z \geq P(y'_x). \end{cases}$$

Then, we have,

$$\begin{aligned} & E(U - U') \\ &= \int_0^{P(y'_x)} \frac{2z}{P(y'_x) + P(y_x)} dz \\ &= \frac{P^2(y'_x)}{P(y'_x) + P(y_x)}. \end{aligned}$$

Case 4:  $P(y_x) < P(y'_x)$  and  $P(y_x) + P(y'_x) < 1$ .

Thus,

$$\begin{aligned} P(y_x) + P(y'_x) - 1 &< 0, \\ P(y_x) - P(y'_x) &< 0. \end{aligned}$$

Therefore, the feasible space of  $D$  is

$$0 \leq D \leq P(y_x) + P(y'_x).$$

and  $D$  uniformly distributed on  $[0, P(y_x) + P(y'_x)]$ .

Also,  $U = \min\{P(y_x), P(y'_x)\} = P(y_x)$ ,

and  $U' < P(y_x)$ . In order to make  $U' < U$ ,  $D$  has to be in interval  $[0, P(y_x)]$  or  $(P(y'_{x'}), P(y_x) + P(y'_{x'}))$ . We also know that,

$$P(y_x) - 0 = P(y_x) + P(y'_{x'}) - P(y'_{x'}) = P(y_x),$$

therefore, the maximum value of  $U - U'$  is  $P(y_x)$ , and  $U - U' \leq z \leq P(y_x)$ , when  $D$  in interval  $[P(y_x) - z, P(y_x)]$  or  $[P(y'_{x'}), P(y'_{x'}) + z]$ . Thus, we have,

$$P(U - U' \leq z) = \left\{ \begin{array}{ll} 0, & \text{if } z < 0, \\ \frac{P(y'_{x'}) - P(y_x) + 2z}{P(y_x) + P(y'_{x'})}, & \text{if } 0 \leq z < P(y_x), \\ 1, & \text{if } z \geq P(y_x). \end{array} \right\},$$

Then, the probability density function  $f$  is

$$f(z) = \left\{ \begin{array}{ll} 0, & \text{if } z < 0, \\ \frac{2}{P(y_x) + P(y'_{x'})}, & \text{if } 0 \leq z < P(y_x), \\ 0, & \text{if } z \geq P(y_x). \end{array} \right\},$$

Then, we have,

$$\begin{aligned} E(U - U') &= \int_0^{P(y_x)} \frac{2z}{P(y_x) + P(y'_{x'})} dz \\ &= \frac{P^2(y_x)}{P(y_x) + P(y'_{x'})}. \end{aligned}$$

We then combine the results of cases 1 and 3.

Note that when  $P(y_x) \geq P(y'_{x'})$ , we have  $P(y'_x) \leq P(y_{x'})$ , and when  $P(y_x) < P(y'_{x'})$ , we have  $P(y'_x) > P(y_{x'})$ , therefore,

when  $P(y_x) + P(y'_{x'}) \geq 1$ , we have,

$$\begin{aligned} E(U - U') &= \frac{\min\{P^2(y'_x), P^2(y_{x'})\}}{P(y'_x) + P(y_{x'})}. \end{aligned}$$

Similarly, after combine the results of cases 2 and 4, we have,

when  $P(y_x) + P(y'_{x'}) < 1$ , we have,

$$\begin{aligned} E(U - U') &= \frac{\min\{P^2(y_x), P^2(y'_{x'})\}}{P(y_x) + P(y'_{x'})}. \end{aligned}$$

Now, note that when  $P(y_x) + P(y'_{x'}) \geq 1$ , we have,

$P(y_x) + P(y'_{x'}) \geq P(y'_x) + P(y_{x'})$ .

Also,  $P(y_x) \geq P(y_{x'})$  and  $P(y'_{x'}) \geq P(y'_x)$ ,

thus,  $\min\{P^2(y_x), P^2(y'_{x'})\} \geq \min\{P^2(y'_x), P^2(y_{x'})\}$ .

Similarly, when  $P(y_x) + P(y'_{x'}) < 1$ , we have,

$P(y_x) + P(y'_{x'}) < P(y'_x) + P(y_{x'})$ ,

and,  $\min\{P^2(y_x), P^2(y'_{x'})\} < \min\{P^2(y'_x), P^2(y_{x'})\}$ .

Thus, finally, we have,

$$\begin{aligned} E(U - U') &= \frac{\min\{P^2(y_x), P^2(y'_x), P^2(y_{x'}), P^2(y'_{x'})\}}{\min\{P(y_x) + P(y'_{x'}), P(y'_x) + P(y_{x'})\}}. \end{aligned}$$

□

### 8.3 Proof of Corollary 8

First, when  $C$  is a set of variables that does not contain any descendant of  $X$ . We have the conditional on  $c$  version of Theorems 4 and 5 for c-PNS  $P(y_x, y'_{x'}|c)$  as following. The proof is exactly the same as above, but with every probability conditioned on  $c$ .

**Corollary 9.** *Given a causal diagram  $G$  and distribution compatible with  $G$  with experimental data  $P(y_x|c)$  and  $P(y_{x'}|c)$ , let  $C$  be a set of variables that does not contain any descendant of  $X$  in  $G$ . let  $D = P(y|c)$  be a random variable. Let  $L$  be the lower bound of c-PNS using pure experimental data and  $L'$  be the lower bound of c-PNS using a combination of experimental and observational data. If  $D$  is uniformly distributed on its feasible interval  $[\max(0, P(y_x|c) - P(y'_{x'}|c)), \min(1, P(y_x|c) + P(y_{x'}|c))]$ , and  $P(y_x|c) + P(y_{x'}|c) \neq 0$  and  $P(y'_x|c) + P(y'_{x'}|c) \neq 0$ , then we have the expectation of the increased lower bound  $E(L' - L)$  as follows:*

$$\begin{aligned} E(L' - L) &= \frac{\min\{P^2(y_x|c), P^2(y'_x|c), P^2(y_{x'}|c), P^2(y'_{x'}|c)\}}{\min\{P(y_x|c) + P(y_{x'}|c), P(y'_x|c) + P(y'_{x'}|c)\}} \end{aligned}$$

where,

$$\begin{aligned} P(y'_x|c) &= 1 - P(y_x|c), \\ P(y'_{x'}|c) &= 1 - P(y_{x'}|c), \end{aligned}$$

$$L' = \max \left\{ \begin{array}{l} 0, \\ P(y_x|c) - P(y_{x'}|c), \\ D - P(y_{x'}|c), \\ P(y_x|c) - D \end{array} \right\},$$

$$L = \max \left\{ \begin{array}{l} 0, \\ P(y_x|c) - P(y_{x'}|c) \end{array} \right\}.$$

*Proof.* The proof follows exactly the same as Theorem 4, but with every probability conditioned on  $c$ . □

**Corollary 10.** *Given a causal diagram  $G$  and distribution compatible with  $G$  with experimental data  $P(y_x|c)$  and  $P(y_{x'}|c)$ , let  $C$  be a set of variables that does not contain*

any descendant of  $X$  in  $G$ . let  $D = P(x, y|c) + P(x', y'|c)$  be a random variable. Let  $U$  be the upper bound of  $c$ -PNS using pure experimental data and  $U'$  be the upper bound of  $c$ -PNS using a combination of experimental and observational data. If  $D$  is uniformly distributed on its feasible interval  $[\max(0, P(y_x|c) - P(y_{x'}|c)), \min(1, P(y_x|c) + P(y'_{x'}|c))]$ , and  $P(y_x|c) + P(y'_{x'}|c) \neq 0$  and  $P(y'_{x'}|c) + P(y_{x'}|c) \neq 0$ , then we have the expectation of the decreased upper bound  $E(U - U')$  as follows:

$$\begin{aligned} & E(U - U') \\ &= \frac{\min\{P^2(y_x|c), P^2(y'_{x'}|c), P^2(y_{x'}|c), P^2(y'_{x'}|c)\}}{\min\{P(y_x|c) + P(y'_{x'}|c), P(y'_{x'}|c) + P(y_{x'}|c)\}} \end{aligned}$$

where,

$$\begin{aligned} P(y'_{x'}|c) &= 1 - P(y_x|c), \\ P(y'_{x'}|c) &= 1 - P(y_{x'}|c), \end{aligned}$$

$$U' = \min \left\{ \begin{array}{c} P(y_x|c), \\ P(y'_{x'}|c), \\ D, \\ P(y_x|c) + P(y'_{x'}|c) - D \end{array} \right\},$$

$$U = \min \left\{ \begin{array}{c} P(y_x|c), \\ P(y'_{x'}|c) \end{array} \right\}.$$

*Proof.* The proof follows exactly the same as Theorem 5, but with every probability conditioned on  $c$ .  $\square$

Now, let us proof Corollary 8.

**Corollary 8.** Given a causal diagram  $G$  and distribution compatible with  $G$  with experimental data  $P(y_x|c)$  and  $P(y_{x'}|c)$ , let  $C$  be a set of variables that does not contain any descendant of  $X$  in  $G$ . let  $D = P(y|c)$  be a random variable, and let  $D' = P(x, y|c) + P(x', y'|c)$  be another random variable. Let  $LB, UB$  be the lower and upper bound of the benefit function using pure experimental data, respectively. Let  $LB', UB'$  be the lower and upper bound of the benefit function using a combination of experimental and observational data, respectively. If  $D$  is uniformly distributed on its feasible interval  $[\max(0, P(y_x|c) - P(y'_{x'}|c)), \min(1, P(y_x|c) + P(y_{x'}|c))]$  and  $D'$  is uniformly distributed on its feasible interval  $[\max(0, P(y_x|c) - P(y_{x'}|c)), \min(1, P(y_x|c) + P(y'_{x'}|c))]$ , and  $P(y_x|c) + P(y'_{x'}|c) \neq 0$  and  $P(y'_{x'}|c) + P(y_{x'}|c) \neq 0$  and  $P(y_x|c) + P(y_{x'}|c) \neq 0$  and  $P(y'_{x'}|c) + P(y_{x'}|c) \neq 0$ , then we have the expectation of the increased lower bound  $E(LB' - LB)$  and the decreased

upper bound  $E(UB - UB')$  as follows:

$$\begin{aligned} & E(LB' - LB) \\ &= \sigma \frac{\min\{P^2(y_x|c), P^2(y'_{x'}|c), P^2(y_{x'}|c), P^2(y'_{x'}|c)\}}{\min\{P(y_x|c) + P(y_{x'}|c), P(y'_{x'}|c) + P(y_{x'}|c)\}} \\ & \quad \text{if } \sigma > 0, \\ & E(LB' - LB) \\ &= -\sigma \frac{\min\{P^2(y_x|c), P^2(y'_{x'}|c), P^2(y_{x'}|c), P^2(y'_{x'}|c)\}}{\min\{P(y_x|c) + P(y'_{x'}|c), P(y'_{x'}|c) + P(y_{x'}|c)\}} \\ & \quad \text{if } \sigma < 0, \\ & E(UB - UB') \\ &= \sigma \frac{\min\{P^2(y_x|c), P^2(y'_{x'}|c), P^2(y_{x'}|c), P^2(y'_{x'}|c)\}}{\min\{P(y_x|c) + P(y'_{x'}|c), P(y'_{x'}|c) + P(y_{x'}|c)\}} \\ & \quad \text{if } \sigma > 0, \\ & E(UB - UB') \\ &= -\sigma \frac{\min\{P^2(y_x|c), P^2(y'_{x'}|c), P^2(y_{x'}|c), P^2(y'_{x'}|c)\}}{\min\{P(y_x|c) + P(y_{x'}|c), P(y'_{x'}|c) + P(y'_{x'}|c)\}} \\ & \quad \text{if } \sigma < 0, \end{aligned}$$

where,

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

$$W = (\gamma - \delta)P(y_x|c) + \delta P(y_{x'}|c) + \theta P(y'_{x'}|c),$$

$$P(y'_{x'}|c) = 1 - P(y_x|c),$$

$$P(y'_{x'}|c) = 1 - P(y_{x'}|c),$$

$$LB = W + \sigma U \quad \text{if } \sigma < 0,$$

$$LB = W + \sigma L \quad \text{if } \sigma > 0,$$

$$LB' = W + \sigma U' \quad \text{if } \sigma < 0,$$

$$LB' = W + \sigma L' \quad \text{if } \sigma > 0,$$

$$UB = W + \sigma L \quad \text{if } \sigma < 0,$$

$$UB = W + \sigma U \quad \text{if } \sigma > 0,$$

$$UB' = W + \sigma L' \quad \text{if } \sigma < 0,$$

$$UB' = W + \sigma U' \quad \text{if } \sigma > 0,$$

$$L' = \max \left\{ \begin{array}{c} 0, \\ P(y_x|c) - P(y_{x'}|c), \\ D - P(y_{x'}|c), \\ P(y_x|c) - D \end{array} \right\},$$

$$L = \max \left\{ \begin{array}{c} 0, \\ P(y_x|c) - P(y_{x'}|c) \end{array} \right\},$$

$$U' = \min \left\{ \begin{array}{c} P(y_x|c), \\ P(y'_{x'}|c), \\ D', \\ P(y_x|c) + P(y'_{x'}|c) - D' \end{array} \right\},$$

$$U = \min \left\{ \begin{array}{c} P(y_x|c), \\ P(y'_{x'}|c) \end{array} \right\}.$$

*Proof.* By Corollary 9, we have,

$$\begin{aligned} & E(L' - L) \\ = & \frac{\min\{P^2(y_x|c), P^2(y'_x|c), P^2(y_{x'}|c), P^2(y'_{x'}|c)\}}{\min\{P(y_x|c) + P(y_{x'}|c), P(y'_x|c) + P(y'_{x'}|c)\}}, \end{aligned}$$

and by Corollary 10, we have,

$$\begin{aligned} & E(U - U') \\ = & \frac{\min\{P^2(y_x|c), P^2(y'_x|c), P^2(y_{x'}|c), P^2(y'_{x'}|c)\}}{\min\{P(y_x|c) + P(y'_{x'}|c), P(y'_x|c) + P(y_{x'}|c)\}}. \end{aligned}$$

Case 1:  $\sigma > 0$ .

$$\begin{aligned} & E(LB' - LB) \\ = & E(W + \sigma L' - W - \sigma L) \\ = & \sigma E(L' - L) \\ = & \sigma \frac{\min\{P^2(y_x|c), P^2(y'_x|c), P^2(y_{x'}|c), P^2(y'_{x'}|c)\}}{\min\{P(y_x|c) + P(y_{x'}|c), P(y'_x|c) + P(y'_{x'}|c)\}}, \end{aligned}$$

$$\begin{aligned} & E(UB - UB') \\ = & E(W + \sigma U - W - \sigma U') \\ = & \sigma E(U - U') \\ = & \sigma \frac{\min\{P^2(y_x|c), P^2(y'_x|c), P^2(y_{x'}|c), P^2(y'_{x'}|c)\}}{\min\{P(y_x|c) + P(y'_{x'}|c), P(y'_x|c) + P(y_{x'}|c)\}}. \end{aligned}$$

Case 2:  $\sigma < 0$ .

$$\begin{aligned} & E(LB' - LB) \\ = & E(W + \sigma U' - W - \sigma U) \\ = & -\sigma E(U - U') \\ = & -\sigma \frac{\min\{P^2(y_x|c), P^2(y'_x|c), P^2(y_{x'}|c), P^2(y'_{x'}|c)\}}{\min\{P(y_x|c) + P(y'_{x'}|c), P(y'_x|c) + P(y_{x'}|c)\}}, \end{aligned}$$

$$\begin{aligned} & E(UB - UB') \\ = & E(W + \sigma L - W - \sigma L') \\ = & -\sigma E(L' - L) \\ = & -\sigma \frac{\min\{P^2(y_x|c), P^2(y'_x|c), P^2(y_{x'}|c), P^2(y'_{x'}|c)\}}{\min\{P(y_x|c) + P(y_{x'}|c), P(y'_x|c) + P(y'_{x'}|c)\}}. \end{aligned}$$

□