Probabilities of Causation with Nonbinary Treatment and Effect

Ang Li[1], **Judea Pearl**[2]

[1]Department of Computer Science, Florida State University
[2]Cognitive Systems Laboratory, Department of Computer Science, University of California, Los Angeles
angli@cs.fsu.edu, judea@cs.ucla.edu

Abstract

Probabilities of causation are proven to be critical in modern decision-making. This paper deals with the problem of estimating the probabilities of causation when treatment and effect are not binary. Pearl defined the binary probabilities of causation, such as the probability of necessity and sufficiency (PNS), the probability of sufficiency (PS), and the probability of necessity (PN). Tian and Pearl then derived sharp bounds for these probabilities of causation using experimental and observational data. In this paper, we define and provide theoretical bounds for all types of probabilities of causation with multivalued treatments and effects. We further discuss examples where our bounds guide practical decisions and use simulation studies to evaluate how informative the bounds are for various data combinations.

Introduction

In many areas of industry, marketing, and health science, the probabilities of causation are widely used to solve decision-making problems. For example, Li and Pearl (Li and Pearl 2019) proposed the "benefit function", which is the payoff/cost associated with selecting an individual with given characteristics to identify a set of individuals who are most likely to exhibit a desired mode of behavior. In Li and Pearl's paper, the benefit function is a linear combination of the probabilities of causation with binary treatment and effect. For another example, Mueller and Pearl (Mueller and Pearl 2022) demonstrated that the probabilities of causation should be considered in personalized decision-making.

Consider the following motivating scenario: an elderly patient with cancer is faced with the choice of treatment to pursue. The options include surgery, chemotherapy, and radiation. The outcomes include ineffective, cured, and death. Given that the elderly patient has a high risk of death from cancer surgery, the patient wants to know the probability that he would be cured if he chose radiation, would die if he chose surgery, and nothing would change if he chose chemotherapy. Let X denotes the treatment, where x_1 denotes surgery, x_2 denotes chemotherapy, and x_3 denotes radiation. Let Y denotes the outcome, where y_1 denotes ineffective, y_2 denotes cured, and y_3 denotes death. The probability that the patient desires is the probability of causation, $P({y_3}_{x_1}, {y_1}_{x_2}, {y_2}_{x_3})$.

Pearl (Pearl 1999) first defined three binary probabilities of causation (i.e., PNS, PN, and PS) using structural causal model (SCM) (Galles and Pearl 1998; Halpern 2000; Pearl 2009). Tian and Pearl (Tian and Pearl 2000) then used observational and experimental data to bound those three probabilities of causation. Li and Pearl (Li and Pearl 2019, 2022b) provided formal proof of those bounds. Mueller, Li, and Pearl (Mueller, Li, and Pearl 2022) recently proposed using covariate information and the causal structure to narrow the bounds of the probability of necessity and sufficiency. Dawid et al. (Dawid, Musio, and Murtas 2017) also proposed using covariate information to narrow the bounds of the probability of necessity.

All the above-mentioned studies are restricted to binary treatment and effect, limiting the application of probabilities of causation. One can easily extend Tian and Pearl's bounds (Tian and Pearl 2000) to a non-binary situation by applying Balke's linear programming again. However, the vertex enumeration algorithm within Balke's linear programming (Balke 1995), involving an exponential number of variables, renders the task impractical.

Zhang, Tian, and Bareinboim (Zhang, Tian, and Bareinboim 2022), as well as Li and Pearl (Li and Pearl 2022a), proposed nonlinear programming-based solutions to compute the bounds of nonbinary probabilities of causation numerically. However, the theoretical foundation of nonbinary probabilities of causation is still required, not only because numerical methods are limited by computational power but also because people are interested in the theoretical foundation due to further development and analysis. In this paper, we will introduce the theoretical bounds of any probabilities of causation defined using SCM without restricting them to binary treatment and effect.

Preliminaries

In this section, we review the definitions for the three aspects of binary causation, as defined in (Pearl 1999). We use the language of counterfactuals in SCM, as defined in (Galles and Pearl 1998; Halpern 2000).

We use $Y_x = y$ to denote the counterfactual sentence "Variable Y would have the value y, had X been x". For the remainder of the paper, we use y_x to denote the event $Y_x = y$, $y_{x'}$ to denote the event $Y_{x'} = y$, y'_x to denote the event $Y_x = y'$, and $y'_{x'}$ to denote the event $Y_{x'} = y'$. We assume that experimental data will be summarized in the form of the causal effects such as $P(y_x)$ and observational data will be summarized in the form of the joint probability function such as $P(x, y)$. If not specified, the variable X stands for treatment and the variable Y stands for effect.

Three prominent probabilities of causation are 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 $X = true$ and $Y = true$, respectively, and x' and y' for their complements. The probability of necessity is defined as the expression (Pearl 1999)*

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

Definition 2 (Probability of sufficiency (PS)). *(Pearl 1999)*

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

Definition 3 (Probability of necessity and sufficiency (PNS)). *(Pearl 1999)*

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

PNS stands for the probability that y would respond to x both ways, and therefore measures both the sufficiency and necessity of x to produce y.

Tian and Pearl (Tian and Pearl 2000) provided tight bounds for PNS, PN, and PS using Balke's program (Balke 1995) (we will call them Tian-Pearl's bounds). Li and Pearl (Li and Pearl 2019, 2022b) provided theoretical proof of the tight bounds for PNS, PS, PN, and other binary probabilities of causation.

PNS, PN, and PS have the following tight bounds:

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

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

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

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

Note that we only consider PNS and PN here because the bounds of PS can easily be obtained by exchanging x with x' and y with y' in the bounds of PN. To obtain bounds for a specific population defined by a set C of characteristics, the expressions above should be modified by conditioning each term on $C = c$.

However, the above three probabilities of causation are unable to answer the query in our motivating example. In this paper, we demonstrate the bounds of any type of probability of causation. We illustrate the theorems by order of the number of hypothetical terms (i.e., the number of y_x terms in the probability of causation). For example, the number of hypothetical terms in $P(y_x, y'_{x'})$ is 2.

Probabilities of Causation with Single Hypothetical Term

We start with four simple probabilities of causation with a single hypothetical term. Let X denotes the treatment with potential values $x_1, ..., x_m$ and Y denotes the effect with potential values $y_1, ..., y_n$. The four probabilities of causation with a single hypothetical term are $P(y_{i_{x_j}}, y_i)$, $P(y_{i_{x_j}}, y_k), s.t., i \neq k$, $P(y_{i_{x_j}}, x_k), s.t., j \neq k$, and $P(y_{i_{x_j}}, y_k, x_m), s.t., m \neq j$. The following theorems define their bounds using observational and experimental data. These bounds are sharp in the sense that there exists an SCM for each point inside the bounds such that the probability of causation is equal to that point.

Theorem 4 (Probability of preservation(i, j) (PPre(i, j))). *Suppose variable X has m values $x_1, ..., x_m$ and Y has n values $y_1, ..., y_n$, then the probability of preservation(i, j) $P(y_{i_{x_j}}, y_i)$, where $1 \leq i \leq n, 1 \leq j \leq m$, has the following sharp bounds:*

$$\max \left\{ \begin{array}{c} P(x_j, y_i), \\ P(y_{i_{x_j}}) + P(y_i) - 1 \end{array} \right\} \leq P(y_{i_{x_j}}, y_i)$$

$$P(y_{i_{x_j}}, y_i) \leq \min \left\{ \begin{array}{c} P(y_{i_{x_j}}), \\ P(y_i) \end{array} \right\}$$

Note that the conditional probability $P(y_{i_{x_j}} | y_i)$ is simple to obtain by PPre$(i, j)/P(y_i)$, and it stands for the probability that y_i would have been preserved if X were set to x_j.

Theorem 5 (Probability of replacement(i, j, k) (PRep(i, j, k))). *Suppose variable X has m values $x_1, ..., x_m$ and Y has n values $y_1, ..., y_n$, then the probability of replacement(i, j, k) $P(y_{i_{x_j}}, y_k)$, where $1 \leq i, k \leq n, 1 \leq j \leq m, i \neq k$, has the following sharp bounds:*

$$\max \left\{ \begin{array}{c} 0, \\ P(y_{i_{x_j}}) + P(y_k) - 1, \\ \sum_{1 \leq p \leq m, p \neq j} \max \left\{ \begin{array}{c} 0, \\ P(y_{i_{x_j}}) \\ + P(x_p, y_k) \\ -1 + P(x_j) \\ -P(x_j, y_i) \end{array} \right\} \end{array} \right\} \leq P(y_{i_{x_j}}, y_k)$$

$$P(y_{i_{x_j}}, y_k) \leq \min \left\{ \begin{array}{c} P(y_{i_{x_j}}) - P(x_j, y_i), \\ P(y_k) - P(y_k, x_j) \end{array} \right\}$$

Note that the conditional probability $P(y_{i_{x_j}} | y_k)$ is simple to obtain by PRep$(i, j, k)/P(y_k)$, and it stands for the probability that y_k would have been replaced by y_i if X were set to x_j. In addition, one can interpret the probability of replacement(i, j, k) as the nonbinary version of the probability of disablement and the probability of enablement defined by Pearl (Pearl 1999).

Theorem 6 (Probability of substitute(i, j, k) (PSub(i, j, k)))**.** *Suppose variable X has m values $x_1, ..., x_m$ and Y has n values $y_1, ..., y_n$, then the probability of substitute(i, j, k) $P(y_{i_{x_j}}, x_k)$, where $1 \leq i \leq n, 1 \leq j, k \leq m, j \neq k$, has the following sharp bounds:*

$$\max \left\{ \begin{array}{c} 0, \\ P(y_{i_{x_j}}) - P(x_j, y_i) \\ -1 + P(x_j) + P(x_k) \end{array} \right\} \leq P(y_{i_{x_j}}, x_k)$$

$$P(y_{i_{x_j}}, x_k) \leq \min \left\{ \begin{array}{c} P(y_{i_{x_j}}) - P(x_j, y_i), \\ P(x_k) \end{array} \right\}$$

Note that the conditional probability $P(y_{i_{x_j}}|x_k)$ is simple to obtain by PSub$(i, j, k)/P(x_k)$, and it stands for the probability that y_i would have happened if x_k were substituted by x_j.

Theorem 7 (Probability of necessity(i, j, k, p) (PN(i,j,k,p)))**.** *Suppose variable X has m values $x_1, ..., x_m$ and Y has n values $y_1, ..., y_n$, then the probability of necessity(i, j, k, p) $P(y_{i_{x_j}}, y_k, x_p)$, where $1 \leq i, k \leq n, 1 \leq j, p \leq m, j \neq p$, has the following sharp bounds:*

$$\max \left\{ \begin{array}{c} 0, \\ P(y_{i_{x_j}}) + P(x_p, y_k) \\ -1 + P(x_j) - P(x_j, y_i) \end{array} \right\} \leq P(y_{i_{x_j}}, y_k, x_p)$$

$$P(y_{i_{x_j}}, y_k, x_p) \leq \min \left\{ \begin{array}{c} P(y_{i_{x_j}}) - P(x_j, y_i), \\ P(x_p, y_k) \end{array} \right\}$$

PN(i, j, k, p) is the nonbinary version of PN and PS defined by Pearl. Note that the conditional probability $P(y_{i_{x_j}}|x_p, y_k)$ is simple to obtain by PN$(i, j, k, p)/P(x_p, y_k)$, and it stands for the probability of that y_k would have been replaced by y_i if x_p were substituted by x_j.

Note that there is no theorem for the probability of causation $P(y_{i_{x_j}}, x_j)$ because $P(y_{i_{x_j}}, x_j)$ simply equals $P(y_i, x_j)$. The proofs for all the theorems, including those in the next section, are provided in the appendix. The key concept of the proof is to employ the Frechet inequalities and the consistency rule (i.e., $P(y_x, x) = P(x, y)$) to simplify the probabilities of causation into a combination of experimental distributions ($P(y_x)$) and observational distributions ($P(x, y)$). This process primarily involves mathematical inequality reasoning, without any exceptional elements.

Probabilities of Causation with Multi Hypothetical Terms

In this section, we deal with four complicated probabilities of causation with multi-hypothetical terms. They are $P(y_{i_{1_{x_{j_1}}}}, ..., y_{i_{k_{x_{j_k}}}})$, $P(y_{i_{1_{x_{j_1}}}}, ..., y_{i_{k_{x_{j_k}}}}, x_p)$, $P(y_{i_{1_{x_{j_1}}}}, ..., y_{i_{k_{x_{j_k}}}}, y_q)$, and $P(y_{i_{1_{x_{j_1}}}}, ..., y_{i_{k_{x_{j_k}}}}, x_p, y_q)$, s.t.,$j_s \neq p$ and $j_s \neq j_w$ if $s \neq w$ for $1 \leq s, w \leq k$. Unlike the bounds in single hypothetical term cases, the bounds in this section are bounded recursively with cases of a smaller number of hypothetical terms. Consequently, we lose the sharpness of the bounds because it is hard to satisfy the equality conditions of Frechet inequalities for all recursive steps.

Theorem 8 (Probability of necessity and sufficiency(k) (PNS(k)))**.** *Suppose variable X has m values $x_1, ..., x_m$ and Y has n values $y_1, ..., y_n$, then the probability of necessity and sufficiency(k) $P(y_{i_{1_{x_{j_1}}}}, ..., y_{i_{k_{x_{j_k}}}})$, where $1 \leq i_s \leq n, 1 \leq j_s \leq m, j_s \neq j_w$ if $s \neq w$ for $1 \leq s, w \leq k$, is bounded as following:*

$$\max \left\{ \begin{array}{c} 0, \\ \\ \sum_{1 \leq t \leq k} P(y_{i_{t_{x_{j_t}}}}) - k + 1, \\ \\ \max_{1 \leq t \leq k}(LB(P(y_{i_{1_{x_{j_1}}}}, ..., y_{i_{t-1_{x_{j_{t-1}}}}}, \\ y_{i_{t+1_{x_{j_{t+1}}}}}, ..., y_{i_{k_{x_{j_k}}}})) \\ +P(y_{i_{t_{x_{i_t}}}}) - 1), \\ \\ \sum_{1 \leq p \leq m, s.t., \exists r, 1 \leq r \leq k, p = j_r} \\ LB(P(y_{i_{1_{x_{j_1}}}}, ..., y_{i_{r-1_{x_{j_{r-1}}}}}, \\ y_{i_{r+1_{x_{j_{r+1}}}}}, ..., y_{i_{k_{x_{j_k}}}}, x_{j_r}, y_{i_r})) + \\ \sum_{1 \leq p \leq m, s.t., p \neq j_1 \neq ... \neq j_k} \\ LB(P(y_{i_{1_{x_{j_1}}}}, ..., y_{i_{k_{x_{j_k}}}}, x_p)) \end{array} \right\} \leq P(y_{i_{1_{x_{j_1}}}}, ..., y_{i_{k_{x_{j_k}}}})$$

$$P(y_{i_{1_{x_{j_1}}}}, ..., y_{i_{k_{x_{j_k}}}}) \leq \min \left\{ \begin{array}{c} \min_{1 \leq t \leq k} P(y_{i_{t_{x_{j_t}}}}), \\ \\ \min_{1 \leq t \leq k} UB(P(y_{i_{1_{x_{j_1}}}}, ..., y_{i_{t-1_{x_{j_{t-1}}}}}, \\ y_{i_{t+1_{x_{j_{t+1}}}}}, ..., y_{i_{k_{x_{j_k}}}})), \\ \\ \sum_{1 \leq p \leq m, s.t., \exists r, 1 \leq r \leq k, p = j_r} \\ UB(P(y_{i_{1_{x_{j_1}}}}, ..., y_{i_{r-1_{x_{j_{r-1}}}}}, \\ y_{i_{r+1_{x_{j_{r+1}}}}}, ..., y_{i_{k_{x_{j_k}}}}, x_{j_r}, y_{i_r})) + \\ \sum_{1 \leq p \leq m, s.t., p \neq j_1 \neq ... \neq j_k} \\ UB(P(y_{i_{1_{x_{j_1}}}}, ..., y_{i_{k_{x_{j_k}}}}, x_p)) \end{array} \right\}$$

where, LB(f) denotes the lower bound of a function f and UB(f) denotes the upper bound of a function f. The bounds of $P(y_{i_{1_{x_{j_1}}}}, ..., y_{i_{r-1_{x_{j_{r-1}}}}}, y_{i_{r+1_{x_{j_{r+1}}}}}, ..., y_{i_{k_{x_{j_k}}}}, x_{j_r}, y_{i_r})$ are given by Theorem 7 or 11, the bounds of $P(y_{i_{1_{x_{j_1}}}}, ..., y_{i_{k_{x_{j_k}}}}, x_p)$ are given by Theorem 6 or 9, and the bounds of $P(y_{i_{1_{x_{j_1}}}}, ..., y_{i_{t-1_{x_{j_{t-1}}}}}, y_{i_{t+1_{x_{j_{t+1}}}}}, ..., y_{i_{k_{x_{j_k}}}})$ are given by Theorem 8 or experimental data if $k = 2$ (i.e., $P(y_{i_{1_{x_{j_1}}}})$ or $P(y_{i_{2_{x_{j_2}}}})$).

Note that PNS(k) is the nonbinary higher-order version of PNS. For example, PNS(2) (nonbinary PNS) stands for the probability that y_{i_1} would respond to x_{j_1} and y_{i_2} would respond to x_{j_2}.

Theorem 9 (Probability of substitute(k,p) (PSub(k,p))). *Suppose variable X has m values $x_1, ..., x_m$ and Y has n values $y_1, ..., y_n$, then the probability of substitute(k,p) $P(y_{i_1 x_{j_1}}, ..., y_{i_k x_{j_k}}, x_p)$, where $1 \leq i_s \leq n, 1 \leq j_s, p \leq m, j_s \neq p, j_s \neq j_w$ if $s \neq w$ for $1 \leq s, w \leq k$, is bounded as following:*

$$\max \left\{ \begin{array}{c} 0, \\ \\ \sum_{1 \leq t \leq k} P(y_{i_t x_{j_t}}) + P(x_p) - k, \\ \\ \max_{1 \leq t \leq k}(LB(P(y_{i_1 x_{j_1}}, ..., y_{i_{t-1} x_{j_{t-1}}}, \\ y_{i_{t+1} x_{j_{t+1}}}, ..., y_{i_k x_{j_k}})) \\ +LB(P(y_{i_t x_{i_t}}, x_p)) - 1) \end{array} \right\} \leq P(y_{i_1 x_{j_1}}, ..., y_{i_k x_{j_k}}, x_p)$$

$$P(y_{i_1 x_{j_1}}, ..., y_{i_k x_{j_k}}, x_p) \leq \min \left\{ \begin{array}{c} \min_{1 \leq t \leq k} P(y_{i_t x_{j_t}}), \\ \\ P(x_p), \\ \\ \min_{1 \leq t \leq k} UB(P(y_{i_1 x_{j_1}}, ..., y_{i_{t-1} x_{j_{t-1}}}, \\ y_{i_{t+1} x_{j_{t+1}}}, ..., y_{i_k x_{j_k}})), \\ \\ \min_{1 \leq t \leq k} UB(P(y_{i_t x_{i_t}}, x_p)) \end{array} \right\}$$

where, LB(f) denotes the lower bound of a function f and UB(f) denotes the upper bound of a function f. The bounds of $P(y_{i_1 x_{j_1}}, ..., y_{i_{t-1} x_{j_{t-1}}}, y_{i_{t+1} x_{j_{t+1}}}, ..., y_{i_k x_{j_k}})$ are given by Theorem 8 or experimental data if $k = 2$ and the bounds of $P(y_{i_t x_{i_t}}, x_p)$ are given by Theorem 6.

Theorem 10 (Probability of replacement(k,q) (PRep(k,q))). *Suppose variable X has m values $x_1, ..., x_m$ and Y has n values $y_1, ..., y_n$, then the probability of replacement(k,q) $P(y_{i_1 x_{j_1}}, ..., y_{i_k x_{j_k}}, y_q)$, where $1 \leq i_s, q \leq n, 1 \leq j_s \leq m, j_s \neq j_w$ if $s \neq w$ for $1 \leq s, w \leq k$, is bounded as following:*

$$\max \left\{ \begin{array}{c} 0, \\ \\ \sum_{1 \leq t \leq k} P(y_{i_t x_{j_t}}) + P(y_q) - k, \\ \\ \max_{1 \leq t \leq k}(LB(P(y_{i_1 x_{j_1}}, ..., y_{i_{t-1} x_{j_{t-1}}}, \\ y_{i_{t+1} x_{j_{t+1}}}, ..., y_{i_k x_{j_k}})) \\ +LB(P(y_{i_t x_{i_t}}, y_q)) - 1), \\ \\ \sum_{1 \leq p \leq m, \exists r, 1 \leq r \leq k, p = j_r, q = i_r} \\ LB(P(y_{i_1 x_{j_1}}, ..., y_{i_{r-1} x_{j_{r-1}}}, \\ y_{i_{r+1} x_{j_{r+1}}}, ..., y_{i_k x_{j_k}}, x_{j_r}, y_{i_r})) + \\ \sum_{1 \leq p \leq m, s.t., p \neq j_1 \neq ... \neq j_k} \\ LB(P(y_{i_1 x_{j_1}}, ..., y_{i_k x_{j_k}}, x_p, y_q)) \end{array} \right\} \leq P(y_{i_1 x_{j_1}}, ..., y_{i_k x_{j_k}}, y_q)$$

$$P(y_{i_1 x_{j_1}}, ..., y_{i_k x_{j_k}}, y_q) \leq \min \left\{ \begin{array}{c} \min_{1 \leq t \leq k} P(y_{i_t x_{j_t}}), \\ \\ P(y_q), \\ \\ \min_{1 \leq t \leq k} UB(P(y_{i_1 x_{j_1}}, ..., y_{i_{t-1} x_{j_{t-1}}}, \\ y_{i_{t+1} x_{j_{t+1}}}, ..., y_{i_k x_{j_k}})), \\ \\ \min_{1 \leq t \leq k} UB(P(y_{i_t x_{i_t}}, y_q)), \\ \\ \sum_{1 \leq p \leq m, s.t., \exists r, 1 \leq r \leq k, p = j_r, q = i_r} \\ UB(P(y_{i_1 x_{j_1}}, ..., y_{i_{r-1} x_{j_{r-1}}}, \\ y_{i_{r+1} x_{j_{r+1}}}, ..., y_{i_k x_{j_k}}, x_{j_r}, y_{i_r})) + \\ \sum_{1 \leq p \leq m, s.t., p \neq j_1 \neq ... \neq j_k} \\ UB(P(y_{i_1 x_{j_1}}, ..., y_{i_k x_{j_k}}, x_p, y_q)) \end{array} \right\}$$

where, LB(f) denotes the lower bound of a function f and UB(f) denotes the upper bound of a function f. The bounds of $P(y_{i_1 x_{j_1}}, ..., y_{i_{r-1} x_{j_{r-1}}}, y_{i_{r+1} x_{j_{r+1}}}, ..., y_{i_k x_{j_k}}, x_{j_r}, y_{j_r})$, $P(y_{i_1 x_{j_1}}, ..., y_{i_k x_{j_k}}, x_p, y_q)$ are given by Theorem 7 or 11, the bounds of $P(y_{i_1 x_{j_1}}, ..., y_{i_{t-1} x_{j_{t-1}}}, y_{i_{t+1} x_{j_{t+1}}}, ..., y_{i_k x_{j_k}})$ are given by Theorem 8 or experimental data if $k = 2$, and the bounds of $P(y_{i_t x_{i_t}}, y_q)$ are given by Theorem 4 or 5.

Theorem 11 (Probability of necessity(k, p, q) (PN(k,p,q))). *Suppose variable X has m values $x_1, ..., x_m$ and Y has n values $y_1, ..., y_n$, then the probability of necessity(k, p, q) $P(y_{i_1 x_{j_1}}, ..., y_{i_k x_{j_k}}, x_p, y_q)$, where $1 \leq i_s, q \leq n, 1 \leq j_s, p \leq m, j_s \neq p, j_s \neq j_w$ if $s \neq w$ for $1 \leq s, w \leq k$, is bounded as following:*

$$\max \left\{ \begin{array}{c} 0, \\ \\ \sum_{1 \leq t \leq k} P(y_{i_t x_{j_t}}) + P(x_p, y_q) - k, \\ \\ \max_{1 \leq t \leq k}(LB(P(y_{i_1 x_{j_1}}, ..., y_{i_{t-1} x_{j_{t-1}}}, \\ y_{i_{t+1} x_{j_{t+1}}}, ..., y_{i_k x_{j_k}})) \\ +LB(P(y_{i_t x_{i_t}}, x_p, y_q)) - 1) \end{array} \right\} \leq P(y_{i_1 x_{j_1}}, ..., y_{i_k x_{j_k}}, x_p, y_q)$$

$$P(y_{i_1 x_{j_1}}, ..., y_{i_k x_{j_k}}, x_p, y_q) \leq \min \left\{ \begin{array}{c} \min_{1 \leq t \leq k} P(y_{i_t x_{j_t}}), \\ \\ P(x_p, y_q), \\ \\ \min_{1 \leq t \leq k} UB(P(y_{i_1 x_{j_1}}, ..., y_{i_{t-1} x_{j_{t-1}}}, \\ y_{i_{t+1} x_{j_{t+1}}}, ..., y_{i_k x_{j_k}})), \\ \\ \min_{1 \leq t \leq k} UB(P(y_{i_t x_{i_t}}, x_p, y_q)) \end{array} \right\}$$

where, LB(f) denotes the lower bound of a function f and UB(f) denotes the upper bound of a function f. The bounds of $P(y_{i_1 x_{j_1}}, ..., y_{i_{t-1} x_{j_{t-1}}}, y_{i_{t+1} x_{j_{t+1}}}, ..., y_{i_k x_{j_k}})$ are given *by Theorem 8 or experimental data if $k = 2$ and the bounds of $P(y_{i_t x_{i_t}}, x_p, y_q)$ are given by Theorem 7.*

Note that the PSub(k, p), PRep(k, q), PN(k, p, q) are the higher order of PSub(i, j, k), PRep(i, j, k), PN(i, j, k, p), respectively. We simply have that PSub$(1, p)$ =PSub(i_1, j_1, p), PRep$(1, q)$ =PRep(i_1, j_1, q), and PN$(1, p, q)$ =PN(i_1, j_1, p, q). Also, the recursive parts in theorems are guaranteed to reduce the number of hypothetical terms in the probabilities of causation by 1; therefore, the recursive parts can reach the single hypothetical term cases in Theorems 4 to 7.

Besides, as mentioned in the introduction section, when extending Balke's linear programming, the explicit form of the bounds for these probabilities of causation can be derived by solving a linear programming problem with mn^m variables, as pointed out by Tian (Tian and Pearl 2000). This is impractical due to the vertex enumeration algorithm with exponential number of variables. For example, if both m and n are set to 10, the vertex enumeration algorithm would require handling 10×10^{10} variables for a simple query like $P(y_{1_{x_1}}, y_{2_{x_2}})$.

Concerns about computational complexity might also arise for Theorems 8 to 11. However, if we consider the number of hypothetical terms as k, the maximum number of probabilities of causation considered in the recursion is $2^{(k+2)}$, with k typically being small.

Examples

In this section, we show how the presented theorems can be used in applications. We start with our motivating example.

Choice of Treatment

An elderly patient with cancer is faced with the choice of treatment. The options from the hospital include surgery, chemotherapy, and radiation. The outcomes include ineffective, cured, and death. Given the elderly patient's high risk of death from cancer surgery, the doctor of the hospital suggested radiation for the patient. So, the patient wants to know the probability that he would be cured if he chose radiation, that he would die if he chose surgery, and that nothing would change if he chose chemotherapy.

Let X denotes the treatment, where x_1 denotes surgery, x_2 denotes chemotherapy, and x_3 denotes radiation. Let Y denotes the outcome, where y_1 denotes ineffective, y_2 denotes cured, and y_3 denotes death. The probability that the patient desires is the PNS(3), $P(y_{3_{x_1}}, y_{1_{x_2}}, y_{2_{x_3}})$.

The doctor provided an experimental study of 900 elderly patients where all the patients were forced to take treatment. The results are shown in Table 1.

The doctor also provided an observational study of 900 elderly patients, where all the patients were open to all treatments and chose the treatment by themselves. The results are shown in Table 2.

	Surgery	Chemotherapy	Radiation
Ineffective	80	184	87
Cured	7	29	189
Death	213	87	24
Overall	300	300	300

Table 1: Experimental data collected by the hospital. Here, 300 patients were forced to receive surgery, 300 patients were forced to receive chemotherapy, and 300 patients were forced to receive radiation.

	Surgery	Chemotherapy	Radiation
Ineffective	238	10	147
Cured	20	77	72
Death	7	259	70
Overall	265	346	289

Table 2: Observational data collected by the hospital. Here, 900 patients were free to choose one of the three treatments by themselves; 265 patients chose surgery, 346 patients chose chemotherapy, and 289 patients chose radiation.

The experimental data provide the following estimates:

$$\begin{aligned}P(y_{1_{x_1}}) = 80/300, P(y_{2_{x_1}}) = 7/300,\\ P(y_{3_{x_1}}) = 213/300, P(y_{1_{x_2}}) = 184/300,\\ P(y_{2_{x_2}}) = 29/300, P(y_{3_{x_2}}) = 87/300,\\ P(y_{1_{x_3}}) = 87/300, P(y_{2_{x_3}}) = 189/300,\\ P(y_{3_{x_3}}) = 24/300.\end{aligned}$$

Here, all three experimental estimates, $P(y_{3_{x_1}})$, $P(y_{1_{x_2}})$, and $P(y_{2_{x_3}})$, in the target probability of causation are higher than 0.5, which may give us the sense that the target probability of causation, $P(y_{3_{x_1}}, y_{1_{x_2}}, y_{2_{x_3}})$, would be high.

The observational data provide the following estimates:

$$\begin{aligned}P(x_1, y_1) = 238/900, P(x_1, y_2) = 20/900,\\ P(x_1, y_3) = 7/900, P(x_2, y_1) = 10/900,\\ P(x_2, y_2) = 77/900, P(x_2, y_3) = 259/900,\\ P(x_3, y_1) = 147/900, P(x_3, y_2) = 72/900,\\ P(x_3, y_3) = 70/900.\end{aligned}$$

We then plug the estimates into Theorem 8 (see the appendix for the detailed calculations). We obtain the bounds of the target probability of causation as follows:

$$0 \leq P(y_{3_{x_1}}, y_{1_{x_2}}, y_{2_{x_3}}) \leq 0.099$$

In conclusion, the probability that the patient would be cured if he chose radiation, that he would die if he chose surgery, and that nothing would change if he chose chemotherapy is below 0.099, implying that the patient should not consider radiation as a treatment option.

Change of Institute

Bob is looking for a job in the job market. There are three institutes, say A, B, and C, that offer courses to help people prepare for job searches. Bob went to one of the institutes,

	Success	Failure	Overall
No institute	53	247	300
Institute A	269	31	300
Institute B	234	66	300
Institute C	151	149	300

Table 3: Experimental data collected by Bob. Here, 300 people were forced to take no course, 300 people were forced to take a course at institute A, 300 people were forced to take a course at institute B, and 300 people were forced to take a course at institute C.

	Success	Failure	Overall
No institute	92	58	150
Institute A	55	118	173
Institute B	24	231	255
Institute C	599	23	622

Table 4: Observational data collected by Bob. Here, 1200 people were open to all institutes, 150 people chose to take no course, 173 people chose to take a course at institute A, 255 people chose to take a course at institute B, and 622 people chose to take a course at institute C.

A, and took the course, but he still failed on the job market. Thus, Bob wonders if these courses improve his chance of getting a job. What would happen if he chose the other two institutes?

Let X denotes which institute a person is chosen, where x_1 denotes that no institute is chosen, x_2 denotes institute A, x_3 denotes institute B, and x_4 denotes institute C. Let Y denotes whether a person gets a job, where y_1 denotes success in job seeking, and y_2 denotes failure in job seeking. Therefore, Bob's questions become the following two probabilities of causation, $P(y_{1_{x_3}}|x_2, y_2)$ and $P(y_{1_{x_4}}|x_2, y_2)$.

All institutes provided experimental and observational studies to illustrate their effectiveness. Bob summarized the studies in Tables 3 and 4.

The experimental data provide the estimates:

$$P(y_{1_{x_1}}) = 53/300, P(y_{2_{x_1}}) = 247/300,$$
$$P(y_{1_{x_2}}) = 269/300, P(y_{2_{x_2}}) = 31/300,$$
$$P(y_{1_{x_3}}) = 234/300, P(y_{2_{x_3}}) = 66/300,$$
$$P(y_{1_{x_4}}) = 151/300, P(y_{2_{x_4}}) = 149/300.$$

The observational data provide the estimates:

$$P(x_1, y_1) = 92/1200, P(x_1, y_2) = 58/1200,$$
$$P(x_2, y_1) = 55/1200, P(x_2, y_2) = 118/1200,$$
$$P(x_3, y_1) = 24/1200, P(x_3, y_2) = 231/1200,$$
$$P(x_4, y_1) = 599/1200, P(x_4, y_2) = 23/1200.$$

Based on the experimental study, institute A claims that taking their course increased the success rate of finding a job from 0.177 to 0.897 and institute B claims that taking their course increased the success rate of finding a job from 0.177 to 0.780. Based on the observational study, institute C claims that taking their course increased the success rate of finding a job from 0.613 to 0.963. All of these seem useful to the job seeker, which is why Bob chose institute A previously. However, he still failed in the job market.

Now, consider the following two probabilities of causation,

$$P(y_{1_{x_3}}|x_2, y_2) = P(y_{1_{x_3}}, x_2, y_2)/P(x_2, y_2),$$
$$P(y_{1_{x_4}}|x_2, y_2) = P(y_{1_{x_4}}, x_2, y_2)/P(x_2, y_2),$$

What would be the probability of success if he had chosen the other two institutes?

We plug the experimental and observational estimates into Theorem 7 to obtain the following bounds:

$$0.720 \le P(y_{1_{x_3}}|x_2, y_2) \le 1,$$
$$0 \le P(y_{1_{x_4}}|x_2, y_2) \le 0.042.$$

Now Bob can see why he should change the institute to B.

Effectiveness of Vaccine

	Vaccinated	Unvaccinated
Uninfected	205	27
Asymptomatic	46	122
Mild Symptoms	343	87
Severe Condition	6	364
Overall	600	600

Table 5: Experimental data of the clinical study. Here, 600 people were forced to take the vaccine and 600 people were forced to take no vaccine.

A clinical study is conducted to test the effectiveness of the vaccine. The treatment includes vaccinated and unvaccinated. The outcomes include uninfected by the virus, asymptomatic infected, infected with mild symptoms, and infected in a severe condition.

The goal of the clinical study is to learn the probability that a patient would be infected in a severe condition if unvaccinated and would be uninfected if vaccinated, the probability that a patient would be infected in a severe condition if unvaccinated and would be asymptomatic infected if vaccinated, and the probability that a patient would be infected in a severe condition if unvaccinated and would be infected with mild symptoms if vaccinated.

Let X denotes vaccination with x_1 being vaccinated and x_2 being unvaccinated and Y denotes the outcome, where y_1 denotes uninfected by the virus, y_2 denotes asymptomatic infected, y_3 denotes infected with mild symptoms, and y_4 denotes infected in a severe condition. The probabilities of causation we want to evaluate are PNS(2): $P(y_{1_{x_1}}, y_{4_{x_2}})$, $P(y_{2_{x_1}}, y_{4_{x_2}})$, and $P(y_{3_{x_1}}, y_{4_{x_2}})$.

The experimental and observational data of the clinical study are summarized in Tables 5 and 6, respectively.

Based on the clinical study, the researcher of the vaccine claimed that the vaccine is effective in controlling the severe condition, and the number of patients with a severe conditions dropped from 364 to only 6. Besides, some of the patients would be even uninfected because the number of uninfected people increased from 27 to 205.

	Vaccinated	Unvaccinated
Uninfected	6	52
Asymptomatic	74	243
Mild Symptoms	632	147
Severe Condition	5	41
Overall	717	483

Table 6: Observational data of the clinical study. Here, 1200 people were free to access the vaccine. 717 people chose to take the vaccine and 483 people chose to take no vaccine.

Now, consider the probability that a patient would be in a severe condition if unvaccinated and would be uninfected by the virus if vaccinated, $P(y_{1_{x_1}}, y_{4_{x_2}})$, the probability that a patient would be in a severe condition if unvaccinated and would be asymptomatic infected if vaccinated, $P(y_{2_{x_1}}, y_{4_{x_2}})$, and the probability that a patient would be in a severe condition if unvaccinated and would be infected with mild symptoms if vaccinated, $P(y_{3_{x_1}}, y_{4_{x_2}})$.

The experimental data provide the following estimates:

$$P(y_{1_{x_1}}) = 205/600, P(y_{2_{x_1}}) = 46/600,$$
$$P(y_{3_{x_1}}) = 343/600, P(y_{4_{x_1}}) = 6/600,$$
$$P(y_{1_{x_2}}) = 27/600, P(y_{2_{x_2}}) = 122/600,$$
$$P(y_{3_{x_2}}) = 87/600, P(y_{4_{x_2}}) = 364/600.$$

The observational data provide the following estimates:

$$P(x_1, y_1) = 6/1200, P(x_1, y_2) = 74/1200,$$
$$P(x_1, y_3) = 632/1200, P(x_1, y_4) = 5/1200,$$
$$P(x_2, y_1) = 52/1200, P(x_2, y_2) = 243/1200,$$
$$P(x_2, y_3) = 147/1200, P(x_2, y_4) = 41/1200.$$

We plug the estimates into Theorem 8 to obtain the bounds:

$$0 \leq P(y_{1_{x_1}}, y_{4_{x_2}}) \leq 0.039$$
$$0.037 \leq P(y_{2_{x_1}}, y_{4_{x_2}}) \leq 0.077$$
$$0.502 \leq P(y_{3_{x_1}}, y_{4_{x_2}}) \leq 0.561.$$

Thus, the probability of causation that a patient would be in a severe condition if unvaccinated and would be uninfected if vaccinated is at most 0.039, the probability that a patient would be in a severe condition if unvaccinated and would be asymptomatic infected if vaccinated is at most 0.077, and the probability that a patient would be in a severe condition if unvaccinated and would be infected with mild symptoms if vaccinated is at least 0.502.

We conclude that the vaccine is effective in controlling the severe condition but can only make it infected with mild symptoms. The vaccine is ineffective for uninfected and asymptomatic infected if the patient is in a severe condition if unvaccinated.

Simulated Results

In this section, we show the quality of the proposed bounds of the probabilities of causation.

We set $m = 2$ (i.e., X has two values) and $n = 3$ (i.e., Y has three values). We focus on the probability of causation, $P(y_{1_{x_1}}, y_{1_{x_2}})$. We randomly generated 1000 samples of $P(y_{1_{x_1}}, y_{1_{x_2}})$. For each sample, we then generated sample distributions (observational data and experimental data) compatible with the $P(y_{1_{x_1}}, y_{1_{x_2}})$ (see the appendix for the generating algorithm). The advantage of this generating process is that we have the real value of the probability of causation for comparison. The generating algorithm ensures that the experimental data and observational data satisfy the general relation (i.e., $P(x, y|c) \leq P(y_x|c) \leq P(x, y|c) + 1 - P(x|c)$). For a sample i, let $[a_i, b_i]$ be the bounds of the $P(y_{1_{x_1}}, y_{1_{x_2}})$ obtained from the proposed theorems. We summarized the following criteria for each sample as illustrated in Figure 1:

- lower bound : a_i;
- upper bound : b_i;
- midpoint : $(a_i + b_i)/2$;
- real value;

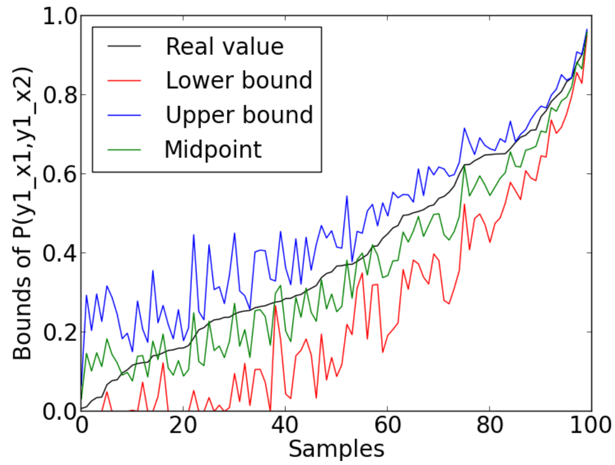

Figure 1: Bounds of the $P(y_{1_{x_1}}, y_{1_{x_2}})$ for 100 samples out of 1000.

From Figure 1, it is clear that the proposed bounds are a good estimation of the real probability of causation. The lower and upper bounds are closely around the real value, and the midpoints are almost identified with the real value. Besides, the average gap of the bounds, $\frac{\sum(b_i - a_i)}{1000}$, is 0.228, which make the bounds convincing.

Conclusion

We demonstrated how to obtain bounds for any probabilities of causation defined using SCM with nonbinary treatment and effect. We derived eight theorems for delivering reasonable bounds. Both examples and simulated studies are provided to support the proposed theorems. It is important to note that the objective of this paper is to establish the most comprehensive bounds applicable to various forms of probabilities of causation. Future research may concentrate on the tightness of the bounds presented in Theorems 8 to 11, or on narrowing the bounds of probabilities of causation under specific graphical conditions.

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

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

Dawid, P.; Musio, M.; and Murtas, R. 2017. The Probability of Causation. *Law, Probability and Risk*, (16): 163–179.

Galles, D.; and Pearl, J. 1998. An axiomatic characterization of causal counterfactuals. *Foundations of Science*, 3(1): 151–182.

Halpern, J. Y. 2000. Axiomatizing causal reasoning. *Journal of Artificial Intelligence Research*, 12: 317–337.

Li, A.; and Pearl, J. 2019. Unit Selection Based on Counterfactual Logic. In *Proceedings of the Twenty-Eighth International Joint Conference on Artificial Intelligence, IJCAI-19*, 1793–1799. International Joint Conferences on Artificial Intelligence Organization.

Li, A.; and Pearl, J. 2022a. Bounds on causal effects and application to high dimensional data. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 36, 5773–5780.

Li, A.; and Pearl, J. 2022b. Unit selection with causal diagram. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 36, 5765–5772.

Mueller, S.; Li, A.; and Pearl, J. 2022. Causes of effects: Learning individual responses from population data. In *Proceedings of the Thirty-First International Joint Conference on Artificial Intelligence (IJCAI-22)*, 2712–2718.

Mueller, S.; and Pearl, J. 2022. Personalized Decision Making – A Conceptual Introduction. Technical Report R-513, Department of Computer Science, University of California, Los Angeles, CA.

Pearl, J. 1999. Probabilities of Causation: Three Counterfactual Interpretations and Their Identification. *Synthese*, 93–149.

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

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

Zhang, J.; Tian, J.; and Bareinboim, E. 2022. Partial counterfactual identification from observational and experimental data. In *International Conference on Machine Learning*, 26548–26558. PMLR.