On the Tightness of Bounds for Probabilities of Causation

Yuxuan Xie[1] **Ang Li**[2]

[1]Algonquin College, Ottawa, Ontario, Canada
[2]Dept. of Computer Science, Florida State University, Tallahassee, Florida, USA

Abstract

Probabilities of causation (PoCs) within Structural Causal Models (SCMs) provide a principled framework for counterfactual reasoning in explainable and personalized decision-making. Since PoCs are generally not point identifiable, existing work has focused on deriving informative bounds. Although the bounds for binary probability of necessity (PN), probability of sufficiency (PS), and probability of necessity and sufficiency (PNS) are widely regarded as tight, no formal definition or theoretical proof of tightness has been established. In this paper, we formally define the notion of tightness for PoCs and provide a constructive proof of the tightness of the binary PNS bounds. The key idea is to characterize the existence of SCMs through feasible counterfactual distributions and partition the observed distribution space according to active bound terms. We further discuss extensions to PN, PS, and non-binary PoCs. This work provides a general framework for formally establishing tightness results in PoCs.

1 INTRODUCTION

In today's increasingly complex decision-making environment, it is important not only to make accurate decisions, but also to explain them in a transparent and understandable way so that users and decision-makers can trust and interpret the reasoning behind them. At the same time, decision-making methods must be adapted to the specific requirements and standards of different application domains, since solutions that perform well in one setting may not be appropriate for another. Balancing explainability with domain-specific requirements has therefore become a central challenge in modern decision-making systems. Personalized decision-making further increases this challenge by incorporating individual preferences, characteristics, and contextual information, leading to decisions that are more relevant and effective for each user [Lu and Boutilier, 2011, Li and Pearl, 2019, Mueller and Pearl, 2023]. By integrating explainability, domain-specific adaptation, and personalization, decision-making systems can become more effective, trustworthy, and broadly applicable across many real-world domains.

A fundamental difficulty in the above settings stems from their counterfactual nature, since the questions of interest often require reasoning about what would have happened under alternative actions, rather than relying solely on directly observed outcomes. For example, during the early stages of a pandemic, decision-makers may seek to identify individuals who would avoid infection if vaccinated and would become infected if unvaccinated, in order to maximize the effective use of limited vaccine supplies.

Probabilities of causation (PoCs) [Pearl, 1999, Tian and Pearl, 2000, Li and Pearl, 2024, Shu et al., 2025] within Structural Causal Models (SCMs) [Galles and Pearl, 1998, Halpern, 2000, Pearl, 2009] enable such decision-making tasks by properly representing counterfactual relationships. Researchers have also demonstrated the capability of PoCs in decision-making applications [Tian and Pearl, 2000, Li and Pearl, 2019, 2022, Mueller and Pearl, 2023]. However, the estimation of PoCs still faces several fundamental challenges. Tian and Pearl [2000] showed that, without additional assumptions, PoCs are generally not point identifiable. Instead, only informative bounds can be obtained. In particular, Tian and Pearl [2000] derived tight bounds for three binary PoCs: the probability of necessity (PN), the probability of sufficiency (PS), and the probability of necessity and sufficiency (PNS). Here, "tight" ("sharp" in Tian and Pearl [2000]) means that the bounds cannot be further improved without introducing stronger assumptions. Later, Li and Pearl [2024] and Shu et al. [2025] defined and bounded more fundamental non-binary PoCs, which are important for modern decision-making tasks. However, the tightness of these bounds remains unknown.

The notion of tightness plays a critical role in decision-making. If the obtained bounds are too loose to support a reliable conclusion, and the bounds are already known to be tight, then the decision-maker must seek additional assumptions or data to further tighten them, as explored in [Mueller et al., 2022]. Otherwise, it may still be possible to derive better bounds through more sophisticated mathematical analysis under the current setting. The tightness results for binary PN, PS, and PNS [Tian and Pearl, 2000] were established through the global optimization guarantees of Balke's linear programming framework [Balke, 1995]. However, while Balke's framework provides optimization-based guarantees of tightness, a formal constructive proof showing that the bounds can be realized by valid counterfactual distributions induced by SCMs has remained unclear. More importantly, due to the computational complexity of Balke's method, directly extending the tightness analysis to the non-binary settings considered in [Li and Pearl, 2024, Shu et al., 2025] becomes computationally intractable.

In this paper, we formally define the notion of tightness for probabilities of causation (PoCs) and theoretically prove the tightness of the bounds for binary PNS. The framework naturally extends to PS and PN, and may provide a path toward non-binary settings (although the non-binary case is more technically involved, it remains feasible). This paper provides a general framework for how tightness results for PoCs can be formally established and proved.

2 PRELIMINARIES

We briefly introduce the three binary Probabilities of Causation (PoCs), namely PS, PN, and PNS [Pearl, 1999], using the language of Structural Causal Models (SCMs) [Galles and Pearl, 1998, Halpern, 2000, Pearl, 2009], together with the notation used throughout this paper.

Let $Y_x = y$ denote the counterfactual statement "Y would take the value y if X were set to x." For notational convenience, throughout the paper we write y_x for the event $Y_x = y$, $y_{x'}$ for $Y_{x'} = y$, y'_x for $Y_x = y'$, and $y'_{x'}$ for $Y_{x'} = y'$.

Pearl [1999] defined the following three probabilities of causation.

Definition 2.1 (Probability of Necessity (PN))**.** Let X and Y be two binary variables in a causal model M, where x and y denote the propositions $X = \text{true}$ and $Y = \text{true}$, respectively, and x' and y' denote their complements. The probability of necessity is defined as [Pearl, 1999]

$$\text{PN} \quad \triangleq \quad P(y'_{x'} \mid x, y) \tag{1}$$

Definition 2.2 (Probability of Sufficiency (PS))**.** [Pearl, 1999]

$$\text{PS} \triangleq P(y_x \mid y', x') \tag{2}$$

Definition 2.3 (Probability of Necessity and Sufficiency (PNS))**.** [Pearl, 1999]

$$\text{PNS} \triangleq P(y_x, y'_{x'}) \tag{3}$$

Tian and Pearl [2000] derived bounds for PNS, PN, and PS using Balke's linear programming framework [Balke, 1995]. These bounds are generally considered tight because they are obtained through global optimization within Balke's framework. Later, Li and Pearl [2019] provided a theoretical proof of these bounds. However, the proof of tightness was established only in the sense that there exist particular combinations of $P(X, Y)$ and $P(Y_X)$ for which the lower and upper bounds can be attained.

The tight bounds for PNS and PN are given by

$$\text{PNS} \geq \max \left\{ \begin{array}{c} 0, \\ P(y_x) - P(y_{x'}), \\ P(y) - P(y_{x'}), \\ P(y_x) - P(y) \end{array} \right\} \tag{4}$$

$$\text{PNS} \leq \min \left\{ \begin{array}{c} P(y_x), \\ P(y'_{x'}), \\ P(x, y) + P(x', y'), \\ P(y_x) - P(y_{x'}) + \\ P(x, y') + P(x', y) \end{array} \right\} \tag{5}$$

$$\text{PN} \geq \max \left\{ \begin{array}{c} 0, \\ \frac{P(y) - P(y_{x'})}{P(x,y)} \end{array} \right\} \tag{6}$$

$$\text{PN} \leq \min \left\{ \begin{array}{c} 1, \\ \frac{P(y'_{x'}) - P(x',y')}{P(x,y)} \end{array} \right\} \tag{7}$$

The bounds for PS follow directly by symmetry, by exchanging x with x' and y with y'. However, despite the widespread use of the term "tight bounds" in the literature, there has been no formal definition of tightness for PoC bounds, nor a general theoretical framework for proving such tightness.

3 MAIN RESULTS

We first formally define the notion of tightness.

Definition 3.1 (Tightness of the Bounds for a Probability of Causation)**.** Let PC be a Probability of Causation defined over variables X and Y, and suppose it is bounded by $L \leq PC \leq U$, where L and U are functions of the experimental and observational distributions $P(Y_X)$ and $P(X, Y)$. The bounds are said to be *tight* if, for any given combination of $P(X, Y)$ and $P(Y_X)$, there exists a Structural Causal Model (SCM) compatible with $P(X, Y)$ and $P(Y_X)$, such that $PC = L$, and there exists an SCM compatible with $P(X, Y)$ and $P(Y_X)$ such that $PC = U$.

Within this definition, we can now prove the following theorem on the tightness of the bounds for the binary PNS defined in Equations 4 and 5. The main idea is as follows. Since Li and Pearl [2019] already provided a theoretical proof for the bounds in Equations 4 and 5, for any given $P(X, Y)$ and $P(Y_X)$, if there exists a compatible SCM such that PNS is equal to one term in the $\max$ (or $\min$) bracket, then that term must be the active $\max$ (or $\min$) term. Therefore, to establish tightness (e.g., for the lower bound in Equation 4), it is sufficient to show that, for any given $P(X, Y)$ and $P(Y_X)$, there exists a compatible SCM such that PNS is equal to one of the terms in the $\max$ bracket.

Directly constructing such SCMs is challenging. However, by setting PNS equal to a specific term, for example, the value 0 in Equation 4, we can derive a corresponding condition on $P(X, Y)$ and $P(Y_X)$ under which there exists a compatible SCM satisfying $\text{PNS} = 0$. If the conditions obtained by setting PNS equal to each of the four terms in the $\max$ bracket form a partition of the entire $P(X, Y)$ and $P(Y_X)$ space, then the tightness is established.

There is one more issue that remains unclear before going into the formal proof: what is meant by the statement "there exists a compatible SCM." In this paper, a compatible SCM means an SCM whose induced counterfactual distribution is consistent with the given observational and experimental distributions. More specifically, an SCM uniquely determines a joint distribution over the counterfactual variables, namely $P(Y_x, Y_{x'}, X, Y)$ (for binary X), which uniquely determines $P(X, Y)$ and $P(Y_X)$. Therefore, throughout the proof, when we say that "there exists a compatible SCM," it is sufficient to construct a valid nonnegative joint distribution $P(Y_x, Y_{x'}, X, Y)$ satisfying the required constraints, since any such distribution can be realized by an SCM.

Theorem 3.2. *The bounds of PNS given by Equations 4 and 5 are tight.*

Proof. We first focus on the lower bound in Equation 4. Given the experimental distribution $P(Y_X)$ and the observational distribution $P(X, Y)$, and similarly to the formalization in [Tian and Pearl, 2000], to show that there exists a compatible SCM, it is sufficient to construct the following eight nonnegative parameters (note that Y is no longer needed here, since it is uniquely determined by $Y_x, Y_{x'}, X$):

$$\begin{aligned} p_{111} &= P(y_x, y_{x'}, x) & p_{110} &= P(y_x, y_{x'}, x') \\ p_{101} &= P(y_x, y'_{x'}, x) & p_{100} &= P(y_x, y'_{x'}, x') \\ p_{011} &= P(y'_x, y_{x'}, x) & p_{010} &= P(y'_x, y_{x'}, x') \\ p_{001} &= P(y'_x, y'_{x'}, x) & p_{000} &= P(y'_x, y'_{x'}, x') \end{aligned}$$

These parameters must satisfy the following linear system:

$$\begin{aligned} \sum_{u=0}^{1}\sum_{j=0}^{1}\sum_{k=0}^{1} p_{ijk} = 1, p_{111} + p_{101} &= P(x, y) \\ p_{011} + p_{001} = P(x, y'), p_{110} + p_{010} &= P(x', y) \\ p_{111} + p_{110} + p_{101} + p_{100} &= P(y_x) \\ p_{111} + p_{110} + p_{011} + p_{010} &= P(y_{x'}) \end{aligned}$$

Now let $\text{PNS} = p_{101} + p_{100} = 0$. We then obtain the following solution to the above linear system:

$$\begin{aligned} p_{111} &= P(x, y) & p_{110} &= P(y_x) - P(x, y) \\ p_{101} &= 0 & p_{100} &= 0 \\ p_{011} &= P(y_{x'}) - P(y) & p_{010} &= P(y) - P(y_x) \\ p_{001} &= P(y'_{x'}) - P(x', y') & p_{000} &= P(x', y') \end{aligned}$$

Note that we naturally have $P(y_x) \geq P(x, y)$ and $P(y'_{x'}) \geq P(x', y')$ [Tian and Pearl, 2000]. Therefore, the non-negativity implies that $P(y_x) \leq P(y) \leq P(y_{x'})$. That is, under the conditions $P(y_x) \leq P(y) \leq P(y_{x'})$, there exists a compatible SCM such that $\text{PNS} = 0$. In other words, if $P(X, Y)$ and $P(Y_X)$ satisfy $P(y_x) \leq P(y) \leq P(y_{x'})$, then the lower bound in Equation 4 is tight. In this case, 0 must be the tight lower bound, since we have explicitly constructed an SCM achieving $\text{PNS} = 0$. Therefore, if any other term inside the $\max$ operator were greater than 0, it would contradict both the attainability of 0 and the fact that the result of the $\max$ operator is the lower bound. Conversely, terms smaller than 0 cannot be selected by the $\max$ operator because 0 is larger.

Now let $\text{PNS} = p_{101} + p_{100} = P(y_x) - P(y'_x)$, We then obtain the following solution to the linear system:

$$\begin{aligned} p_{111} &= P(y_{x'}) - P(x', y) & p_{110} &= P(x', y) \\ p_{101} &= P(y) - P(y_{x'}) & p_{100} &= P(y_x) - P(y) \\ p_{011} &= 0 & p_{010} &= 0 \\ p_{001} &= P(x, y') & p_{000} &= P(y'_x) - P(x, y') \end{aligned}$$

Again note that we naturally have $P(y_{x'}) \geq P(x', y)$ and $P(y'_x) \geq P(x, y')$ [Tian and Pearl, 2000]. Therefore, the non-negativity implies that $P(y_{x'}) \leq P(y) \leq P(y_x)$. That is, under the conditions $P(y_{x'}) \leq P(y) \leq P(y_x)$, there exists a compatible SCM such that $\text{PNS} = P(y_x) - P(y_{x'})$. Now let $\text{PNS} = p_{101} + p_{100} = P(y) - P(y'_x)$, We then obtain the following solution to the linear system:

$$\begin{aligned} p_{111} &= P(y_{x'}) - P(x', y) & p_{110} &= P(y_x) - P(x, y) \\ p_{101} &= P(y) - P(y_{x'}) & p_{100} &= 0 \\ p_{011} &= 0 & p_{010} &= P(y) - P(y_x) \\ p_{001} &= P(x, y') & p_{000} &= P(x', y') \end{aligned}$$

Again note that we naturally have $P(y_{x'}) \geq P(x', y)$ and $P(y_x) \geq P(x, y)$ [Tian and Pearl, 2000]. Therefore, the non-negativity implies that $P(y_{x'}) \leq P(y)$ and $P(y) \geq P(y_x)$. That is, under the conditions $P(y_{x'}) \leq P(y)$ and $P(y) \geq P(y_x)$, there exists a compatible SCM such that $\text{PNS} = P(y) - P(y_{x'})$.

Lastly, let $\text{PNS} = p_{101} + p_{100} = P(y_x) - P(y)$, We then obtain the following solution to the linear system:

$$p_{111} = P(x, y) \qquad p_{110} = P(x', y)$$

$$p_{101} = 0 \qquad p_{100} = P(y_x) - P(y)$$
$$p_{011} = P(y_{x'}) - P(y) \qquad p_{010} = 0$$
$$p_{001} = P(y'_{x'}) - P(x', y') \qquad p_{000} = P(y'_x) - P(x, y')$$

Again note that we naturally have $P(y'_{x'}) \geq P(x', y')$ and $P(y'_x) \geq P(x, y')$ [Tian and Pearl, 2000]. Therefore, the non-negativity implies that $P(y_{x'}) \geq P(y)$ and $P(y) \leq P(y_x)$. That is, under the conditions $P(y_{x'}) \geq P(y)$ and $P(y) \leq P(y_x)$, there exists a compatible SCM such that $\text{PNS} = P(y_x) - P(y)$.
Notice that the above four conditions

- $P(y_x) \leq P(y) \leq P(y_{x'})$
- $P(y_{x'}) \leq P(y) \leq P(y_x)$
- $P(y_{x'}) \leq P(y)$ and $P(y) \geq P(y_x)$
- $P(y_{x'}) \geq P(y)$ and $P(y) \leq P(y_x)$

form a partition of the space of all possible combinations of $P(X, Y)$ and $P(Y_X)$. Therefore, the lower bound in Equation 4 is tight.
We now turn to the upper bound, let $\text{PNS} = P(y_x)$, we then obtain the following solution to the linear system:

$$p_{110} = 0 \qquad p_{011} = P(y_{x'}) - P(x', y)$$
$$p_{111} = 0 \qquad p_{100} = P(y_x) - P(x, y)$$
$$p_{101} = P(x, y) \qquad p_{000} = P(y'_x) - P(x, y') - P(x', y)$$
$$p_{010} = P(x', y) \qquad p_{001} = P(x, y') + P(x', y) - P(y_{x'})$$

Again note that we naturally have $P(y_x) \geq P(x, y)$ and $P(y_{x'}) \geq P(x', y)$ [Tian and Pearl, 2000]. Therefore, the non-negativity implies that $P(y_{x'}) \leq P(x, y') + P(x', y) \leq P(y'_x)$. That is, under the conditions $P(y_{x'}) \leq P(x, y') + P(x', y) \leq P(y'_x)$, there exists a compatible SCM such that $\text{PNS} = P(y_x)$.
Let $\text{PNS} = P(y'_{x'})$, we then obtain the following solution to the linear system:

$$p_{001} = 0 \qquad p_{101} = P(y'_{x'}) - P(x', y')$$
$$p_{000} = 0 \qquad p_{110} = P(x, y') + P(x', y) - P(y'_x)$$
$$p_{100} = P(x', y') \qquad p_{111} = P(y_{x'}) - P(x, y') - P(x', y)$$
$$p_{011} = P(x, y') \qquad p_{010} = P(y'_x) - P(x, y')$$

Again note that we naturally have $P(y'_{x'}) \geq P(x', y')$ and $P(y'_x) \geq P(x, y')$ [Tian and Pearl, 2000]. Therefore, the non-negativity implies that $P(y'_x) \leq P(x, y') + P(x', y) \leq P(y_{x'})$. That is, under the conditions $P(y'_x) \leq P(x, y') + P(x', y) \leq P(y_{x'})$ there exists a compatible SCM such that $\text{PNS} = P(y'_{x'})$.
Now let $\text{PNS} = p_{101} + p_{100} = P(x, y) + P(x', y')$, We then obtain the following solution to the linear system:

$$p_{111} = 0 \qquad p_{110} = P(x, y') + P(x', y) - P(y'_x)$$
$$p_{101} = P(x, y) \qquad p_{001} = P(x, y') + P(x', y) - P(y_{x'})$$
$$p_{100} = P(x', y') \qquad p_{011} = P(y_{x'}) - P(x', y)$$
$$p_{000} = 0 \qquad p_{010} = P(y'_x) - P(x, y')$$

Again note that we naturally have $P(y_{x'}) \geq P(x', y)$ and $P(y'_x) \geq P(x, y')$ [Tian and Pearl, 2000]. Therefore, the non-negativity implies that $P(y'_x) \leq P(x, y') + P(x', y)$ and $P(y_{x'}) \leq P(x, y') + P(x', y)$. That is, under the conditions $P(y'_x) \leq P(x, y') + P(x', y)$ and $P(y_{x'}) \leq P(x, y') + P(x', y)$, there exists a compatible SCM such that $\text{PNS} = P(x, y) + P(x', y')$.
Lastly, let $\text{PNS} = p_{101} + p_{100} = P(y_x) - P(y_{x'}) + P(x, y') + P(x', y)$, We then obtain the following solution to the linear system:

$$p_{110} = 0 \qquad p_{101} = P(y'_{x'}) - P(x', y')$$
$$p_{001} = 0 \qquad p_{100} = P(y_x) - P(x, y)$$
$$p_{011} = P(x, y') \qquad p_{111} = P(y_{x'}) - P(x, y') - P(x', y)$$
$$p_{010} = P(x', y) \qquad p_{000} = P(y'_x) - P(x, y') - P(x', y)$$

Again note that we naturally have $P(y'_{x'}) \geq P(x', y')$ and $P(y_x) \geq P(x, y)$ [Tian and Pearl, 2000]. Therefore, the non-negativity implies that $P(y'_x) \geq P(x, y') + P(x', y)$ and $P(y_{x'}) \geq P(x, y') + P(x', y)$. That is, under the conditions $P(y'_x) \geq P(x, y') + P(x', y)$ and $P(y_{x'}) \geq P(x, y') + P(x', y)$, there exists a compatible SCM such that $\text{PNS} = P(y_x) - P(y_{x'}) + P(x, y') + P(x', y)$.
Similar to the lower bound, the above four conditions

- $P(y_{x'}) \leq P(x, y') + P(x', y) \leq P(y'_x)$
- $P(y'_x) \leq P(x, y') + P(x', y) \leq P(y_{x'})$
- $P(y'_x) \leq P(x, y') + P(x', y)$ and $P(y_{x'}) \leq P(x, y') + P(x', y)$
- $P(y'_x) \geq P(x, y') + P(x', y)$ and $P(y_{x'}) \geq P(x, y') + P(x', y)$

form a partition of the space of all possible combinations of $P(X, Y)$ and $P(Y_X)$. Therefore, the upper bound in Equation 5 is tight. □

Similarly, the above theorem can be directly extended to the binary cases of PS and PN.

4 CONCLUSION

In this paper, we formally defined the notion of tightness for probabilities of causation (PoCs) within Structural Causal Models. We further provided a theoretical proof of the tightness for binary PNS, which can be naturally extended to PS and PN, thereby completing the mathematical foundation underlying the binary PoC bounds established in [Tian and Pearl, 2000, Li and Pearl, 2019].

The goal of this paper is to provide a general framework for how the tightness of PoCs can be theoretically proved. Extending these tightness results to non-binary PoCs and the complete class of PoCs introduced in [Li and Pearl, 2024, Shu et al., 2025] is a natural and promising direction for future work. The main challenge lies in characterizing when the corresponding large linear systems admit nonnegative solutions and when they do not.

References

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

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

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

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. doi: 10.24963/ijcai.2019/248.

Ang Li and Judea Pearl. Unit selection: Case study and comparison with a/b test heuristic. *arXiv preprint arXiv:2210.05030*, 2022.

Ang Li and Judea Pearl. Probabilities of causation with non-binary treatment and effect. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 38(18), pages 20465–20472, 2024.

Tyler Lu and Craig Boutilier. Budgeted social choice: From consensus to personalized decision making. In *IJCAI*, volume 11, pages 280–286, 2011.

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

Scott Mueller and Judea Pearl. Personalized decision making–a conceptual introduction. *Journal of Causal Inference*, 11(1):20220050, 2023.

Judea Pearl. Probabilities of causation: Three counterfactual interpretations and their identification. *Synthese*, pages 93–149, 1999.

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

Xin Shu, Shuai Wang, and Ang Li. Identification of probabilities of causation: from recursive to closed-form bounds. *arXiv preprint arXiv:2505.15274*, 2025.

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