Proceedings of the Thirty-Eighth AAAI Conference on Artificial Intelligence (AAAI-24),
Volume 38, Technical Track 18, 20465-20472, 2024.

TECHNICAL REPORT
R-517
February 2024

Unit Selection with Nonbinary Treatment and Effect

Ang Li, Judea Pearl

Cognitive Systems Laboratory, Department of Computer Science,
University of California, Los Angeles,
Los Angeles, California, USA.
{angli, judea}@cs.ucla.edu

Abstract

The unit selection problem aims to identify a set of individuals who are most likely to exhibit a desired mode of behavior, for example, selecting individuals who would respond one way if encouraged and a different way if not encouraged. Using a combination of experimental and observational data, Li and Pearl derived tight bounds on the "benefit function", which is the payoff/cost associated with selecting an individual with given characteristics. This paper extends the benefit function to the general form such that the treatment and effect are not restricted to binary. We propose an algorithm to test the identifiability of the nonbinary benefit function and an algorithm to compute the bounds of the nonbinary benefit function using experimental and observational data.

Introduction

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

The binary benefit function for the unit selection problem was defined by Li and Pearl (Li and Pearl 2019) (we will call this Li-Pearl's model), and it properly captures the nature of the desired behavior. Using a combination of experimental and observational data, Li and Pearl derived tight bounds of the benefit function. The only assumption is that the treatment has no effect on the population-specific characteristics. Inspired by the idea of Mueller, Li, and Pearl (Mueller, Li, and Pearl 2021) and Dawid et al. (Dawid, Musio, and Murtas 2017) that the bounds of probabilities of causation could be narrowed using covariates information, Li and Pearl (Li and Pearl 2022c) narrowed the bounds of the benefit function using covariates information and their causal structure. However, the abovementioned studies are based on binary treatment and effect. Recently, researchers have shown interest in developing bounds for probabilities of causation with nonbinary treatment and effect. 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. Li and Pearl (Li and Pearl 2022b) provided the theoretical bounds of nonbinary probabilities of causation. The benefit function is a linear combination of probabilities of causation; therefore, in this paper, we focus on discovering the bounds of any benefit function without restricting them to binary treatment and effect.

Consider the following motivating scenario: a clinical study is conducted to test the effectiveness of a vaccine. The treatments include vaccinated and unvaccinated. The outcomes include uninfected, asymptomatic infected, and infected in a severe condition. The benefited individuals include the following: the individual who would be infected in a severe condition if unvaccinated and would be asymptomatic infected if vaccinated, the individual who would be infected in a severe condition if unvaccinated and would be uninfected if vaccinated, and the individual who would be asymptomatic infected if unvaccinated and would be uninfected if vaccinated. The harmed individuals include the following: the individual who would be asymptomatic infected if unvaccinated and would be infected in a severe condition if vaccinated, the individual who would be uninfected if unvaccinated and would be infected in a severe condition if vaccinated, and the individual who would be uninfected if unvaccinated and would be asymptomatic infected if vaccinated. All others are unaffected individuals. The researcher performing the clinical study has collected both experimental and observational data. The researcher then wants to know the expected difference between benefited and harmed individuals to emphasize the effectiveness of the vaccine.

We cannot apply Li-Pearl's model because we have two treatments and three outcomes. In this paper, we extend Li-Pearl's benefit function to general form without restricting them to binary treatment and effect. We will provide an algorithm to test the identifiability of the nonbinary benefit function and an algorithm to compute the bounds of the nonbinary benefit function using experimental and observational data.

Preliminaries

In this section, we review Li and Pearl's binary benefit function of the unit selection problem (Li and Pearl 2019), and the theoretical bounds of the probabilities of causation recently proposed by Li and Pearl (Li and Pearl 2022b).

In this paper we use the language of counterfactuals in structural model semantics, as given 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 simplicity purposes, in the rest 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 efects 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.

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

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

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

Li and Pearl (Li and Pearl 2022b) provided eight theorems to compute bounds for any type of probabilities of causation with nonbinary treatment and effect. Suppose variable X has m values and Y has n values, the following probabilities of causation are bounded. Besides, if the probabilities of causation are conditioned on a population-specific variable c that is not affected by X, then all the theorems still hold (we provided the extended theorems from Li and Pearl in the appendix).

$$
\begin{aligned}
& P(y_{i_{x_j}}, y_i), \\
& s.t., 1 \le i \le n, 1 \le j \le m, \\
& P(y_{i_{x_j}}, y_k), \\
& s.t., 1 \le i, k \le n, 1 \le j \le m, i \ne k \\
& P(y_{i_{x_j}}, x_k), \\
& s.t., 1 \le i \le n, 1 \le j, k \le m, j \ne k \\
& P(y_{i_{x_j}}, y_k, x_p), \\
& s.t., 1 \le i, k \le n, 1 \le j, p \le m, j \ne p \\
& P(y_{i_1 x_{j_1}}, ..., y_{i_k x_{j_k}}), \\
& s.t., 1 \le i_1, ..., i_k \le n, 1 \le j_1, ..., j_k \le m, j_1 \ne ... \ne j_k \\
& P(y_{i_1 x_{j_1}}, ..., y_{i_k x_{j_k}}, x_p), \\
& s.t., 1 \le i_1, ..., i_k \le n, 1 \le j_1, ..., j_k, p \le m, \\
& j_1 \ne ... \ne j_k \ne p \\
& P(y_{i_1 x_{j_1}}, ..., y_{i_k x_{j_k}}, y_q), \\
& s.t., 1 \le i_1, ..., i_k, q \le n, 1 \le j_1, ..., j_k \le m, \\
& j_1 \ne ... \ne j_k \\
& P(y_{i_1 x_{j_1}}, ..., y_{i_k x_{j_k}}, x_p, y_q), \\
& s.t., 1 \le i_1, ..., i_k, q \le n, 1 \le j_1, ..., j_k, p \le m, \\
& j_1 \ne ... \ne j_k \ne p.
\end{aligned}
$$

The benefit function is a linear combination of the probabilities of causation; therefore, we define the general benefit function for the unit selection problem based on Li and Pearl's results.

Counterfactual Formulation of the Unit Selection Problem

Based on Li and Pearl (Li and Pearl 2019), the objective is to find a set of characteristics c that maximizes the benefit associated with the resulting mixture of different response types of individuals. Let X denotes the treatment with m values and Y denotes the effect with n values. Therefore, we have n^m different response types (i.e., one response type means assigning one effect to each of the treatments). Suppose the benefit of selecting an individual are $(\alpha_1, ..., \alpha_{n^m})$ (we call $(\alpha_1, ..., \alpha_{n^m})$ as benefit vector). Our objective, then, should be to find c that maximizes the following expression (If the goal is to evaluate the average benefit gained per individual for a specific population c, $argmax_c$ can be dropped):

$$
\begin{aligned}
argmax_c \quad & \alpha_1 P(y_{1_{x_1}}, y_{1_{x_2}}, ..., y_{1_{x_m}}|c) + \\
& \alpha_2 P(y_{1_{x_1}}, y_{1_{x_2}}, ..., y_{2_{x_m}}|c) + ... \\
& \alpha_n P(y_{1_{x_1}}, y_{1_{x_2}}, ..., y_{n_{x_m}}|c) + ... \\
& \alpha_{n^{m-1}+1} P(y_{2_{x_1}}, y_{1_{x_2}}, ..., y_{1_{x_m}}|c) + ... \\
& \alpha_{n^m} P(y_{n_{x_1}}, y_{n_{x_2}}, ..., y_{n_{x_m}}|c).
\end{aligned}
$$

Note that c can be interpreted as the population-specific variable, the only assumption is that the treatment X has no effect on the population-specific variable. Recall from Li and Pearl's paper (Li and Pearl 2019), the benefit vector is provided by the decision-maker who uses the model.

In the next section, we will provide an algorithm that could check whether a given benefit function with the benefit vector is identifiable with purely experimental data (i.e., we can find the exact value of the benefit function rather than bounds). If it is not identifiable we will then provide an algorithm that computes the bounds of the benefit function given the benefit vector using experimental and observational data.

Main Results

Identifiability of Benefit Function

Recall that in binary case, the conditions of identifiable are gain equality (i.e., $\beta + \delta = \gamma + \theta$) or monotonicity (i.e., $P(y_{x'}, y'_x) = 0$) (Li and Pearl 2019). Here, it is complicated in nonbinary cases, therefore; we provide an algorithm to test whether a given benefit function with the benefit vector is identifiable with purely experimental data.

Theorem 1. *Suppose variables X has m values $x_1, ..., x_m$ and Y has n values $y_1, ..., y_n$. Then the benefit function $f(c)$ is identifiable if Algorithm 1 returns (True, res), and res is the value of the benefit function.*

$$
\begin{aligned}
f(c) = \quad & \alpha_1 P(y_{1_{x_1}}, y_{1_{x_2}}, ..., y_{1_{x_m}}|c) + \\
& \alpha_2 P(y_{1_{x_1}}, y_{1_{x_2}}, ..., y_{2_{x_m}}|c) + ... \\
& \alpha_n P(y_{1_{x_1}}, y_{1_{x_2}}, ..., y_{n_{x_m}}|c) + ... \\
& \alpha_{n^{m-1}+1} P(y_{2_{x_1}}, y_{1_{x_2}}, ..., y_{1_{x_m}}|c) + ... \\
& \alpha_{n^m} P(y_{n_{x_1}}, y_{n_{x_2}}, ..., y_{n_{x_m}}|c).
\end{aligned}
$$

The correctness of the algorithm simply follow the fact that $\sum_{n^{m-1}\text{terms}} P(..., y_{i_{x_j}}, ...|c) = P(y_{i_{x_j}}|c)$. Therefore, if there exist such n^{m-1} terms in the benefit function, then we can obtain an equivalent benefit function by replacing one of the n^{m-1} terms with experimental data $P(y_{i_{x_j}}|c)$. We exhausted all equivalent benefit functions to check if we could replace all the counterfactual terms with experimental data (i.e., identifiable).

For example, consider $m = n = 2$ and the benefit function:

$$
\begin{aligned}
& 7P(y_{1_{x_1}}, y_{1_{x_2}}|c) + 2P(y_{1_{x_1}}, y_{2_{x_2}}|c) + \\
& 4P(y_{2_{x_1}}, y_{1_{x_2}}|c) - P(y_{2_{x_1}}, y_{2_{x_2}}|c) \\
= \; & 7P(y_{1_{x_1}}|c) - 5P(y_{1_{x_1}}, y_{2_{x_2}}|c) + \\
& 4P(y_{2_{x_1}}, y_{1_{x_2}}|c) - P(y_{2_{x_1}}, y_{2_{x_2}}|c) \\
= \; & 7P(y_{1_{x_1}}|c) - 5P(y_{1_{x_1}}, y_{2_{x_2}}|c) + \\
& 4P(y_{2_{x_1}}|c) - 5P(y_{2_{x_1}}, y_{2_{x_2}}|c) \\
= \; & 7P(y_{1_{x_1}}|c) - 5P(y_{2_{x_2}}) + 4P(y_{2_{x_1}}|c).
\end{aligned}
$$

Bounds of Benefit Function

If Algorithm 1 returns false, we then need to compute the bounds of the benefit function using experimental and observational data. We first obtain the bounds of the probabilities of causation, $P(y_{1_{x_1}}, y_{1_{x_2}}, ..., y_{1_{x_m}}|c)$, ...,

Algorithm 1: Check identifiability of the benefit function

Input: a, the benefit function,where $a[i]$ is a $m+1$ tuple that stands for ith term in the benefit function. If the ith term is $\alpha_i P(y_{i_1 x_1}, y_{i_2 x_2}, ..., y_{i_m x_m}|c)$, then $a[i] = (\alpha_i, i_1, i_2, ..., i_m)$.

$d[1, ..., m][1, ..., n]$, the experimental data, where $d[i][j] = P(y_{j_{x_i}}|c)$.

e, the adjusted value of the benefit function.

The initial call of the algorithm is $IBF(a[1, ..., n^m], d[1, ..., m][1, ..., n], 0)$, where $a[1, ..., n^m]$ corresponding to the original benefit function.

All lists in this algorithm start with index 1.

Output: (identifiable, value), a tuple, where identifiable = True if the given benefit function is identifiable and value is the value of the benefit function.

```
Function $IBF(a, d, e)$:
$m = True$;
$l = length(a)$;
// Base case, if all benefit vector equals to $(0, ..., 0)$, then
   the input benefit function is identifiable, and its value
   equals to the adjusted value.
for $i = 1$ to $l$ do
   if $a[i][1] \neq 0$ then
      $m = False$;
      break;
   end if
end for
if $m == True$ then
   Return$(True, e)$;
end if
// Build an equivalent benefit function by the fact that
   if $\exists 2 \leq r \leq (m+1) s.t., a[j_1][r] = ... = a[j_{n^{m-1}}][r]$,
   then the sum of these $n^{m-1}$ terms without coefficients
   is equal to $P(y_{a[j_1][r]_{x_r}})$, we then recursively solve the
   equivalent benefit function.
for every $n^{m-1}$ pair in $a$, $(a[j_1], ..., a[j_{n^{m-1}}])$, s.t., there
   $\exists 2 \leq r \leq (m+1) s.t., a[j_1][r] = ... = a[j_{n^{m-1}}][r]$ do
   for $k = 1$ to $n^{m-1}$ do
      $na = a$;
      $nc = e + a[j_k][1] * d[r-1][a[j_1][r]]$;
      for $t = 1$ to $n^{m-1}$ do
         if $t \neq k$ then
            $na[j_t][1] = na[j_t][1] - na[j_k][1]$;
         end if
      end for
      $Remove(na[j_k])$;
      $res = IBF(na, d, nc)$;
      if $res[0] == True$ then
         Return $res$
      end if
   end for
end for
Return $(False, e)$;
```

	Vaccinated	Unvaccinated
Uninfected	52 People	329 People
Asymptomatic	512 People	58 People
Severe Condition	36 People	213 People
Overall	600 People	600 People

Table 1: Experimental data of the clinical study. Here, 600 people were forced to take the vaccine and 600 people were forced to take no vaccine.

$P(y_{1_{x_1}}, y_{1_{x_2}}, ..., y_{n_{x_m}}|c)$, ..., $P(y_{2_{x_1}}, y_{1_{x_2}}, ..., y_{1_{x_m}}|c)$, ..., $P(y_{n_{x_1}}, y_{n_{x_2}}, ..., y_{n_{x_m}}|c)$, by Li and Pearl's theorems (Li and Pearl 2022b). We then have the following theorem.

Theorem 2. *Suppose variables X has m values $x_1, ..., x_m$ and Y has n values $y_1, ..., y_n$. Then the bounds of the benefit function $f(c)$ is obtained by Algorithm 2.*

$$\begin{aligned} f(c) = \quad & \alpha_1 P(y_{1_{x_1}}, y_{1_{x_2}}, ..., y_{1_{x_m}}|c) + \\ & \alpha_2 P(y_{1_{x_1}}, y_{1_{x_2}}, ..., y_{2_{x_m}}|c) + ... \\ & \alpha_n P(y_{1_{x_1}}, y_{1_{x_2}}, ..., y_{n_{x_m}}|c) + ... \\ & \alpha_{n^{m-1}+1} P(y_{2_{x_1}}, y_{1_{x_2}}, ..., y_{1_{x_m}}|c) + ... \\ & \alpha_{n^m} P(y_{n_{x_1}}, y_{n_{x_2}}, ..., y_{n_{x_m}}|c). \end{aligned}$$

Again, the correctness of the algorithm simply follow the fact that $\sum_{n^{m-1}\text{terms}} P(..., y_{i_{x_j}}, ...|c) = P(y_{i_{x_j}}|c)$. We exhausted all equivalent benefit functions and take the maximum of all the lower bounds and take the minimum of all the upper bounds of equivalent benefit functions.

Example: Effectiveness of a Vaccine

Recall the motivating example at the beginning, a clinical study is conducted to test the effectiveness of a vaccine. The treatments include vaccinated and unvaccinated. The outcomes include uninfected, asymptomatic infected, and infected in a severe condition. The researcher of the clinical study has collected both experimental and observational data.

Task 1

The researcher wants to know the expected difference between benefited and harmed individuals to emphasize the effectiveness of the vaccine.

Let X denotes vaccination with x_1 being vaccinated and x_2 being unvaccinated and Y denotes outcome, where y_1 denotes uninfected, y_2 denotes asymptomatic infected, and y_3 denotes infected in a severe condition. The experimental and observational data of the clinical study are summarized in Tables 1 and 2.

Based on the clinical study, the researcher of the vaccine claimed that the vaccine is effective in controlling the severe condition, the number of severe condition patients dropped from 213 to only 36.

Algorithm 2: Compute the bounds of the benefit function

Input: a, the benefit function,where $a[i]$ is a $m+1$ tuple that stands for ith term in the benefit function. If the ith term is $\alpha_i P(y_{i_1 x_1}, y_{i_2 x_2}, ..., y_{i_m x_m}|c)$, then $a[i] = (\alpha_i, i_1, i_2, ..., i_m)$.

lb, the lower bound of all possible terms obtained from Li-Pearl's theorems, where $lb[(i_1, i_2, ..., i_m)]$ is the lower bound of $P(y_{i_1 x_1}, y_{i_2 x_2}, ..., y_{i_m x_m}|c)$.

ub, the upper bound of all possible terms obtained from Li-Pearl's theorems, where $ub[(i_1, i_2, ..., i_m)]$ is the upper bound of $P(y_{i_1 x_1}, y_{i_2 x_2}, ..., y_{i_m x_m}|c)$.

e, the adjusted value of the benefit function.
The initial call of the algorithm is $BBF(a[1, ..., n^m], lb, ub, 0)$, where $a[1, ..., n^m]$ corresponding to the original benefit function.

All lists in this algorithm start with index 1.

Output: (lo, up), lower and upper bound of the benefit function.

```
Function $BBF(a, lb, ub, e)$:
$l = length(a)$;
// Base case, compute the bounds.
$up = e, lo = e$;
for $i = 1$ to $l$ do
    if $a[i][1] < 0$ then
        $lo = lo + a[i][1] * ub[(a[i][2], ..., a[i][m+1])]$;
        $up = up + a[i][1] * lb[(a[i][2], ..., a[i][m+1])]$;
    else
        $lo = lo + a[i][1] * lb[(a[i][2], ..., a[i][m+1])]$;
        $up = up + a[i][1] * ub[(a[i][2], ..., a[i][m+1])]$;
    end if
end for
// Build an equivalent benefit function by the fact that
  if $\exists 2 \le r \le (m+1) s.t., a[j_1][r] = ... = a[j_{n^{m-1}}][r]$,
  then the sum of these $n^{m-1}$ terms without coefficients
  is equal to $P(y_{a[j_1][r]_{x_r}})$, we then recursively solve the
  equivalent benefit function.
for every $n^{m-1}$ pair in $a$, $(a[j_1], ..., a[j_{n^{m-1}}])$, s.t., there
  $\exists 2 \le r \le (m+1) s.t., a[j_1][r] = ... = a[j_{n^{m-1}}][r]$ do
    for $k = 1$ to $n^{m-1}$ do
        $na = a$;
        $nc = e + a[j_k][1] * d[r-1][a[j_1][r]]$;
        for $t = 1$ to $n^{m-1}$ do
            if $t \ne k$ then
                $na[j_t][1] = na[j_t][1] - na[j_k][1]$;
            end if
        end for
        $Remove(na[j_k])$;
        $res = BBF(na, lb, ub, nc)$;
        $lo = \max\{lo, res[0]\}$
        $up = \min\{up, res[1]\}$
    end for
end for
Return $(lo, up)$;
```

	Vaccinated	Unvaccinated
Uninfected	14 People	121 People
Asymptomatic	933 People	65 People
Severe Condition	6 People	61 People
Overall	953 People	247 People

Table 2: Observational data of the clinical study. Here, 1200 people were free to the vaccine. 953 people chose to take the vaccine and 247 people chose to take no vaccine.

Now consider the expected difference between benefited and harmed individuals. Recall the benefited individuals include the individual who would be infected in a severe condition if unvaccinated and would be asymptomatic infected if vaccinated, the individual who would be infected in a severe condition if unvaccinated and would be uninfected if vaccinated, and the individual who would be asymptomatic infected if unvaccinated and would be uninfected if vaccinated. The harmed individuals include the individual who would be asymptomatic infected if unvaccinated and would be infected in a severe condition if vaccinated, the individual who would be uninfected if unvaccinated and would be infected in a severe condition if vaccinated, and the individual who would be uninfected if unvaccinated and would be asymptomatic infected if vaccinated. All others are unaffected individuals. In order to maximize the difference between benefited and harmed individuals; therefore, we assign 1 to benefited individuals, assign -1 to harmed individuals, and 0 to all others in the benefit vector. The objective function (i.e., benefit function) is then

$$\begin{aligned} f(c) &= 0P(y_{1_{x_1}}, y_{1_{x_2}}|c) + P(y_{1_{x_1}}, y_{2_{x_2}}|c) + \\ & \quad P(y_{1_{x_1}}, y_{3_{x_2}}|c) - P(y_{2_{x_1}}, y_{1_{x_2}}|c) + \\ & \quad 0P(y_{2_{x_1}}, y_{2_{x_2}}|c) + P(y_{2_{x_1}}, y_{3_{x_2}}|c) - \\ & \quad -P(y_{3_{x_1}}, y_{1_{x_2}}|c) - P(y_{3_{x_1}}, y_{2_{x_2}}|c) + \\ & \quad 0P(y_{3_{x_1}}, y_{3_{x_2}}|c). \end{aligned}$$

The experimental data in Table 1 provide the following estimates:

$$\begin{aligned} P(y_{1_{x_1}}|c) &= 52/600 = 0.087 \\ P(y_{2_{x_1}}|c) &= 512/600 = 0.853 \\ P(y_{3_{x_1}}|c) &= 36/600 = 0.060 \\ P(y_{1_{x_2}}|c) &= 329/600 = 0.548 \\ P(y_{2_{x_2}}|c) &= 58/600 = 0.097 \\ P(y_{3_{x_2}}|c) &= 213/600 = 0.355 \end{aligned}$$

The observational data Table 2 provide the following estimates:

$$\begin{aligned} P(x_1, y_1|c) &= 14/1200 = 0.012 \\ P(x_1, y_2|c) &= 933/1200 = 0.778 \\ P(x_1, y_3|c) &= 6/1200 = 0.005 \\ P(x_2, y_1|c) &= 121/1200 = 0.101 \\ P(x_2, y_2|c) &= 65/1200 = 0.054 \\ P(x_2, y_3|c) &= 61/1200 = 0.051 \end{aligned}$$

We plug the estimates and the benefit function into Theorem 1, the Algorithm 1 returns false (i.e., not identifiable by experimental data). We then plug the estimates and the benefit function into Theorem 2 to obtain the bounds

$$-0.228 \leq f(c) \leq -0.107$$

Thus, the expected difference between benefited and harmed individuals is at most -0.107 per individual. We can conclude that the vaccine is ineffective for the virus.

Task 2

The researcher of the clinic study claimed that the individual who would be infected in a severe condition if unvaccinated and would be uninfected if vaccinated and the individual who would be uninfected if unvaccinated and would be infected in a severe condition if vaccinated should be twice important than other individuals. Based on the clinical study, the number of severe condition patients dropped from 213 to only 36; therefore, the vaccine should be effective for the virus.

Now consider the expected difference between benefited and harmed individuals. The benefit vector should be the same except assigning 2 to the individual who would be infected in a severe condition if unvaccinated and would be uninfected if vaccinated and assigning -2 to the individual who would be uninfected if unvaccinated and would be infected in a severe condition if vaccinated.

The objective function (i.e., benefit function) is then

$$\begin{aligned} f(c) &= 0P(y_{1_{x_1}}, y_{1_{x_2}}|c) + P(y_{1_{x_1}}, y_{2_{x_2}}|c) + \\ & \quad 2P(y_{1_{x_1}}, y_{3_{x_2}}|c) - P(y_{2_{x_1}}, y_{1_{x_2}}|c) + \\ & \quad 0P(y_{2_{x_1}}, y_{2_{x_2}}|c) + P(y_{2_{x_1}}, y_{3_{x_2}}|c) - \\ & \quad -2P(y_{3_{x_1}}, y_{1_{x_2}}|c) - P(y_{3_{x_1}}, y_{2_{x_2}}|c) + \\ & \quad 0P(y_{3_{x_1}}, y_{3_{x_2}}|c). \end{aligned}$$

We plug the estimates and the benefit function into Theorem 1, the Algorithm 1 returns true (i.e., identifiable by experimental data) with value -0.167. The benefit function can be simplified as follow:

$$\begin{aligned} f(c) &= 2P(y_{1_{x_1}}|c) + P(y_{2_{x_1}}|c) - \\ & \quad -2P(y_{1_{x_2}}|c) - P(y_{2_{x_2}}|c) \\ &= -0.167. \end{aligned}$$

Thus, the expected difference between benefited and harmed individuals is exactly -0.167 per individual. We can conclude that the vaccine is still ineffective for the virus.

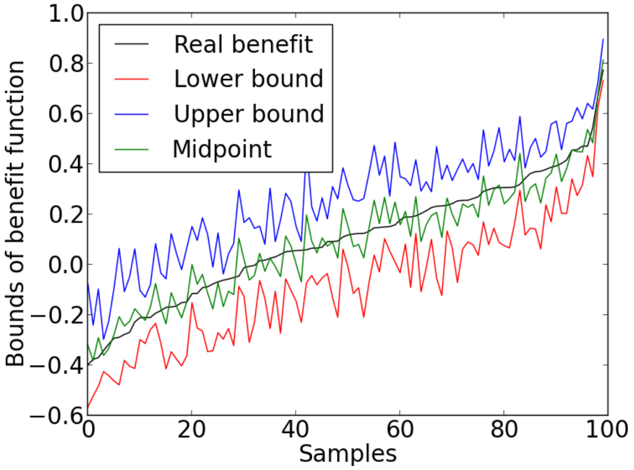

Figure 1: Bounds of the benefit function for 100 sample populations out of 1000 with the benefit vector $(0, 1, 1, -1, 0, 1, -1, -1, 0)$.

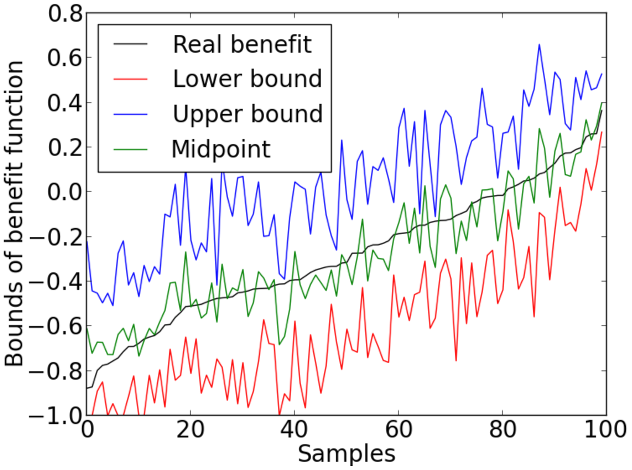

Figure 2: Bounds of the benefit function for 100 sample populations out of 1000 with the benefit vector $(-1, 1, 1, -1, -1, 1, -1, -1, -1)$.

Simulated Results

In this section, we show the quality of the bounds of the benefit function obtained by Theorem 2 using four common benefit vectors.

First, we set $m = 2$ (i.e., X has two values) and $n = 3$ (i.e., Y has three values). We set the benefit vector to one of the most common ones, $(0, 1, 1, -1, 0, 1, -1, -1, 0)$, which is to evaluate the expected difference between benefited and harmed individuals. We randomly generated 1000 populations where each population consists of different fractions of nine response types of individuals. For each population, we then generated sample distributions (observational data and experimental data) compatible with the fractions of response types (see the appendix for the generating algorithm). The advantage of this generating process is that we have the real benefit value (because we know the fractions of the response types) for comparison. Each sample population represents a different instantiate of the population-specific characteristics C in the model. The generating algorithm ensures that the experimental data and observational data satisfy the general relation (i.e., $P(x, y|c) \leq P(y_x|c) \leq P(x, y|c) + 1 - P(x|c)$). For a sample population i, let $[a_i, b_i]$ be the bounds of the benefit function from the proposed theorem. We summarized the following criteria for each population as illustrated in Figure 1:

- lower bound : a_i;
- upper bound : b_i;
- midpoint : $(a_i + b_i)/2$;
- real benefit : dot product of the benefit vector and the fractions of response types;

From Figure 1, it is clear that the proposed bounds obtained from Theorem 2 are a good estimation of the real benefit. The lower and upper bounds are closely around the real benefit and the midpoints are almost identified with the real benefit. Besides, the average gap of the bounds, $\frac{\sum(b_i - a_i)}{1000}$, is 0.330, which is also small compared to the largest possible gap of 6.

Second, we set the benefit vector to another common one, $(-1, 1, 1, -1, -1, 1, -1, -1, -1)$, which is to evaluate the expected difference between benefited and unbenefited (i.e., unaffected and harmed) individuals. We again randomly generated 1000 populations where each population consists of different fractions of nine response types. The data generating process and all other factors remain the same. We summarized the same criteria for each population as illustrated in Figure 2.

From Figure 2, it is clear that the proposed bounds obtained from Theorem 2 are a good estimation of the real benefit. The lower and upper bounds are closely around the real benefit and the midpoints are almost identified with the real benefit. Besides, the average gap of the bounds, $\frac{\sum(b_i - a_i)}{1000}$, is 0.6520, which is also small compared to the largest possible gap of 9.

Third, we set the benefit vector to another common one, $(0, 1, 1, 0, 0, 1, 0, 0, 0)$, which is to evaluate the expected benefited individuals. We again randomly generated 1000 populations where each population consists of different fractions of nine response types. The data generating process and all other factors still remain the same. We summarized the same criteria for each population as illustrated in Figure 3.

From Figure 3, it is clear that the proposed bounds obtained from Theorem 2 are a good estimation of the real benefit. The lower and upper bounds are closely around the real benefit and the midpoints are almost identified with the real benefit. Besides, the average gap of the bounds, $\frac{\sum(b_i - a_i)}{1000}$, is 0.3284, which is also small compared to the largest possible gap of 3.

Lastly, we set the benefit vector to the last common one, $(0, 0, 0, -1, 0, 0, -1, -1, 0)$, which is to evaluate the expected harmed individuals (we set the benefit vector to -1 because we want to minimize the harmed individuals). We again randomly generated 1000 populations where each

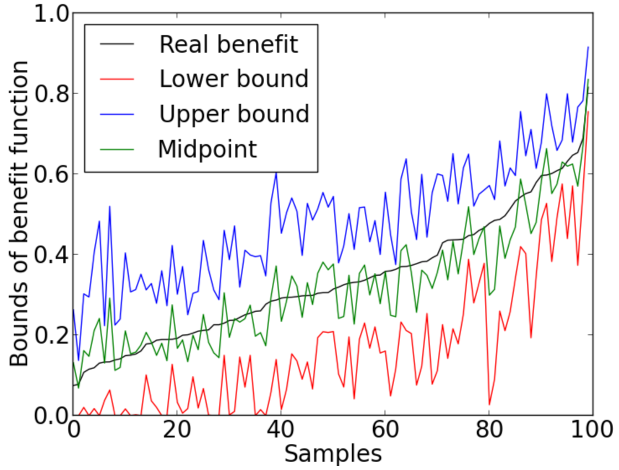

Figure 3: Bounds of the benefit function for 100 sample populations out of 1000 with the benefit vector $(0, 1, 1, 0, 0, 1, 0, 0, 0)$.

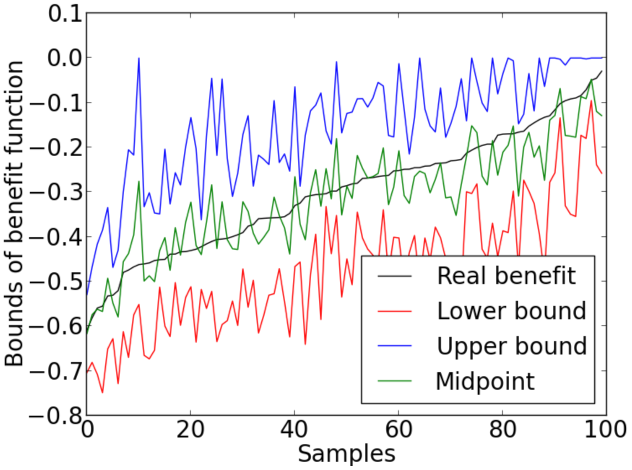

Figure 4: Bounds of the benefit function for 100 sample populations out of 1000 with the benefit vector $(0, 0, 0, -1, 0, 0, -1, -1, 0)$.

population consists of different fractions of nine response types. The data generating process and all other factors still remain the same. We summarized the same criteria for each population as illustrated in Figure 4.

From Figure 4, it is clear that the proposed bounds obtained from Theorem 2 are a good estimation of the real benefit. The lower and upper bounds are closely around the real benefit and the midpoints are almost identified with the real benefit. Besides, the average gap of the bounds, $\frac{\sum(b_i-a_i)}{1000}$, is 0.3266, which is also small compared to the largest possible gap of 3.

Discussion

We have shown that the proposed theorems are a good estimation of the non-binary benefit function using examples and simulated studies. One may concern about the computation complexity of Algorithms 1 and 2. They are for sure in exponential time. However, the m and n (i.e., values of X and Y) are usually small constant, therefore, we do not need to worry about too much.

Conclusion and Future Work

We demonstrated the formalization of the general benefit function with nonbinary treatment and effect. We provided the algorithm to compute the bounds of the general benefit function and the algorithm to check whether the benefit function is identifiable with purely experimental data. Examples and simulation results are provided to support the proposed theorems.

Future studies could assess the statistical properties of the proposed bounds. How tight would the bounds be? Does it sufficient to make decisions? Which data, experimental or observational, would affect the bounds more? How would the number of values in treatment and effect affect the quality of the bounds?

Another future direction could be to improve the bounds using covariate information as Li and Pearl (Li and Pearl 2019) did for the binary benefit function.

Acknowledgements

This research was supported in parts by grants from the National Science Foundation [#IIS-2106908], Office of Naval Research [#N00014-17-S-12091 and #N00014-21-1-2351], and Toyota Research Institute of North America [#PO-000897].

References

Berson, A.; Smith, S.; and Thearling, K. 1999. *Building data mining applications for CRM*. McGraw-Hill Professional.

Bottou, L.; Peters, J.; Quiñonero-Candela, J.; Charles, D. X.; Chickering, D. M.; Portugaly, E.; Ray, D.; Simard, P.; and Snelson, E. 2013. Counterfactual reasoning and learning systems: The example of computational advertising. *The Journal of Machine Learning Research*, 14(1): 3207–3260.

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

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.

Hung, S.-Y.; Yen, D. C.; and Wang, H.-Y. 2006. Applying data mining to telecom churn management. *Expert Systems with Applications*, 31(3): 515–524.

Lejeune, M. A. 2001. Measuring the impact of data mining on churn management. *Internet Research*, 11(5): 375–387.

Li, A.; and Pearl, J. 2019. Unit Selection Based on Counterfactual Logic. In *Proceedings of the 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. Probabilities of Causation with Non-binary Treatment and Effect. Technical Report R-516, Department of Computer Science, University of California, Los Angeles, CA.

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

Li, L.; Chen, S.; Kleban, J.; and Gupta, A. 2014. Counterfactual estimation and optimization of click metrics for search engines. *arXiv preprint arXiv:1403.1891*.

Mueller, S.; Li, A.; and Pearl, J. 2021. Causes of effects: Learning individual responses from population data. Technical Report R-505, <http://ftp.cs.ucla.edu/pub/stat_ser/r505.pdf>, Department of Computer Science, University of California, Los Angeles, CA.

Sun, W.; Wang, P.; Yin, D.; Yang, J.; and Chang, Y. 2015. Causal inference via sparse additive models with application to online advertising. In *AAAI*, 297–303.

Tsai, C.-F.; and Lu, Y.-H. 2009. Customer churn prediction by hybrid neural networks. *Expert Systems with Applications*, 36(10): 12547–12553.

Yan, J.; Liu, N.; Wang, G.; Zhang, W.; Jiang, Y.; and Chen, Z. 2009. How much can behavioral targeting help online advertising? In *Proceedings of the 18th international conference on World Wide Web*, 261–270. ACM.

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.

Appendix

Proof of Theorems

Theorem 1. *Suppose variables X has m values $x_1, ..., x_m$ and Y has n values $y_1, ..., y_n$. Then the benefit function $f(c)$ is identifiable if Algorithm 1 returns (True, res), and res is the value of the benefit function.*

$$\begin{aligned} f(c) = \quad & \alpha_1 P(y_{1_{x_1}}, y_{1_{x_2}}, ..., y_{1_{x_m}}|c) + \\ & \alpha_2 P(y_{1_{x_1}}, y_{1_{x_2}}, ..., y_{2_{x_m}}|c) + ... \\ & \alpha_n P(y_{1_{x_1}}, y_{1_{x_2}}, ..., y_{n_{x_m}}|c) + ... \\ & \alpha_{n^{m-1}+1} P(y_{2_{x_1}}, y_{1_{x_2}}, ..., y_{1_{x_m}}|c) + ... \\ & \alpha_{n^m} P(y_{n_{x_1}}, y_{n_{x_2}}, ..., y_{n_{x_m}}|c). \end{aligned}$$

Proof. The proof is simple.
Lines 1 to 12 in Algorithm 1 simply check whether the given benefit function f (encoded as a in the algorithm) is identifiable.
On line 24 of the algorithm, we recursively call on another benefit function f' (encoded as na in the algorithm).
Now lets consider how we obtain f'.
if there exist n^{m-1} terms in a and r, s.t., $a[j_1][r] = ... = a[j_{n^{m-1}}][r]$, these n^{m-1} terms are $P(..., y_{a[j_1][r]_{x_{r-1}}}, ...|c)$, and the sum of these n^{m-1} terms is equal to $P(y_{a[j_1][r]_{x_{r-1}}}|c)$.
We obtain f' by eliminating kth of the n^{m-1} terms in f, replacing kth term by other $n^{m-1} - 1$ terms and their sum $P(y_{a[j_1][r]_{x_{r-1}}}|c)$.
Therefore, $f = f' + nc$ where $nc = a[j_k][1] * P(y_{a[j_1][r]_{x_{r-1}}}|c)$. Thus, f is identifiable if and only if f' is identifiable and $f = f' + nc$. □

Theorem 2. *Suppose variables X has m values $x_1, ..., x_m$ and Y has n values $y_1, ..., y_n$. Then the bounds of the benefit function $f(c)$ is obtained by Algorithm 2.*

$$\begin{aligned} f(c) = \quad & \alpha_1 P(y_{1_{x_1}}, y_{1_{x_2}}, ..., y_{1_{x_m}}|c) + \\ & \alpha_2 P(y_{1_{x_1}}, y_{1_{x_2}}, ..., y_{2_{x_m}}|c) + ... \\ & \alpha_n P(y_{1_{x_1}}, y_{1_{x_2}}, ..., y_{n_{x_m}}|c) + ... \\ & \alpha_{n^{m-1}+1} P(y_{2_{x_1}}, y_{1_{x_2}}, ..., y_{1_{x_m}}|c) + ... \\ & \alpha_{n^m} P(y_{n_{x_1}}, y_{n_{x_2}}, ..., y_{n_{x_m}}|c). \end{aligned}$$

Proof. Similarly to Theorem 1,
lines 1 to 12 in Algorithm 2 simply compute the bounds of the given benefit function f (encoded as a in the algorithm).
On line 24 of the algorithm, we recursively call on another benefit function f' (encoded as na in the algorithm).
Now lets consider how we obtain f'.
if there exist n^{m-1} terms in a and r, s.t., $a[j_1][r] = ... = a[j_{n^{m-1}}][r]$, these n^{m-1} terms are $P(..., y_{a[j_1][r]_{x_{r-1}}}, ...|c)$, and the sum of these n^{m-1} terms is equal to $P(y_{a[j_1][r]_{x_{r-1}}}|c)$.
We obtain f' by eliminating kth of the n^{m-1} terms in f, replacing kth term by other $n^{m-1} - 1$ terms and their sum $P(y_{a[j_1][r]_{x_{r-1}}}|c)$.
Therefore, $f = f' + nc$ where $nc = a[j_k][1] * P(y_{a[j_1][r]_{x_{r-1}}}|c)$. Thus, the bounds of $f' + nc$ is the bounds of f. □

Li-Pearl's Bounds of Probabilities of Causation

The input, lb, ub, in Algorithm 2 depends on the bounds of probabilities of causation. The bounds of probabilities of causation recently proposed by Li and Pearl (Li and Pearl 2022b) is not conditional C. However, nothing is changed if conditioning on a variable C that is not affected by X. We listed the conditional version of the eight theorems proposed by Li and Pearl. The proof of eight theorems is exactly the same, except every probability should be conditioned on C.

Theorem 3. *Suppose variable X has m values $x_1, ..., x_m$ and Y has n values $y_1, ..., y_n$, and variable C is not affected by X, then the probability of causation $P(y_{i_{x_j}}, y_i|c)$, where $1 \leq i \leq n, 1 \leq j \leq m$, is bounded as following:*

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

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

Theorem 4. *Suppose variable X has m values $x_1, ..., x_m$ and Y has n values $y_1, ..., y_n$, and variable C is not affected by X, then the probability of causation $P(y_{i_{x_j}}, y_k|c)$, where $1 \leq i, k \leq n, 1 \leq j \leq m, i \neq k$, is bounded as following:*

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

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

Theorem 5. *Suppose variable X has m values $x_1, ..., x_m$ and Y has n values $y_1, ..., y_n$, and variable C is not affected by X, then the probability of causation $P(y_{i_{x_j}}, x_k|c)$, where $1 \leq i \leq n, 1 \leq j, k \leq m, j \neq k$, is bounded as following:*

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

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

Theorem 6. *Suppose variable X has m values $x_1, ..., x_m$ and Y has n values $y_1, ..., y_n$, and variable C is not affected by X, then the probability of causation $P(y_{i_{x_j}}, y_k, x_p|c)$,* *where $1 \leq i,k \leq n, 1 \leq j,p \leq m, j \neq p$, is bounded as following:*

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

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

Theorem 7. *Suppose variable X has m values $x_1,...,x_m$ and Y has n values $y_1,...,y_n$, and variable C is not affected by X, then the probability of causation $P(y_{i_1 x_{j_1}},...,y_{i_k x_{j_k}}|c)$, where $1 \leq i_1,...,i_k \leq n, 1 \leq j_1,...,j_k \leq m, j_1 \neq ... \neq j_k$, is bounded as following:*

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

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

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 $P(y_{i_1 x_{j_1}},...,y_{i_{r-1} x_{j_{r-1}}},y_{i_{r+1} x_{j_{r+1}}},...,y_{i_k x_{j_k}},x_{j_r},y_{j_r}|c)$ are given by Theorem 6 or 10, the bounds of $P(y_{i_1 x_{j_1}},...,y_{i_k x_{j_k}},x_p|c)$ are given by Theorem 5 or 8, and the bounds of $P(y_{i_1 x_{j_1}},...,y_{i_{t-1} x_{j_{t-1}}},y_{i_{t+1} x_{j_{t+1}}},...,y_{i_k x_{j_k}}|c)$ are given by Theorem 7 or experimental data if $k=2$.

Theorem 8. *Suppose variable X has m values $x_1,...,x_m$ and Y has n values $y_1,...,y_n$, and variable C is not affected by X, then the probability of causation $P(y_{i_1 x_{j_1}},...,y_{i_k x_{j_k}},x_p|c)$, where $1 \leq i_1,...,i_k \leq n, 1 \leq j_1,...,j_k,p \leq m, j_1 \neq ... \neq j_k \neq p$, is bounded as following:*

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

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

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 $P(y_{i_1 x_{j_1}},...,y_{i_{t-1} x_{j_{t-1}}},y_{i_{t+1} x_{j_{t+1}}},...,y_{i_k x_{j_k}}|c)$ are given by Theorem 7 or experimental data if $k=2$ and the bounds of $P(y_{i_t x_{i_t}},x_p|c)$ are given by Theorem 5.

Theorem 9. *Suppose variable X has m values $x_1,...,x_m$ and Y has n values $y_1,...,y_n$, and variable C is not affected by X, then the probability of causation $P(y_{i_1 x_{j_1}},...,y_{i_k x_{j_k}},y_q|c)$, where $1 \leq i_1,...,i_k,q \leq n, 1 \leq j_1,...,j_k \leq m, j_1 \neq ... \neq j_k$, is bounded as following:*

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

$$P(y_{i_1 x_{j_1}}, ..., y_{i_k x_{j_k}}, y_q|c) \leq$$

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

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 $P(y_{i_1 x_{j_1}}, ..., y_{i_{r-1} x_{j_{r-1}}}, y_{i_{r+1} x_{j_{r+1}}}, ..., y_{i_k x_{j_k}}, x_{j_r}, y_{j_r}|c)$, $P(y_{i_1 x_{j_1}}, ..., y_{i_k x_{j_k}}, x_p, y_q|c)$ are given by Theorem 6 or 10, the bounds of $P(y_{i_1 x_{j_1}}, ..., y_{i_{t-1} x_{j_{t-1}}}, y_{i_{t+1} x_{j_{t+1}}}, ..., y_{i_k x_{j_k}}|c)$ are given by Theorem 7 or experimental data if $k = 2$, and the bounds of $P(y_{i_t x_{i_t}}, y_q|c)$ are given by Theorem 3 or 4.

Theorem 10. *Suppose variable X has m values $x_1, ..., x_m$ and Y has n values $y_1, ..., y_n$, and variable C is not affected by X, then the probability of causation $P(y_{i_1 x_{j_1}}, ..., y_{i_k x_{j_k}}, x_p, y_q|c)$, where $1 \leq i_1, ..., i_k, q \leq n, 1 \leq j_1, ..., j_k, p \leq m, j_1 \neq ... \neq j_k \neq p$, is bounded as following:*

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

$$\leq P(y_{i_1 x_{j_1}}, ..., y_{i_k x_{j_k}}, x_p, y_q|c)$$

$$P(y_{i_1 x_{j_1}}, ..., y_{i_k x_{j_k}}, x_p, y_q|c) \leq$$

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

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 $P(y_{i_1 x_{j_1}}, ..., y_{i_{t-1} x_{j_{t-1}}}, y_{i_{t+1} x_{j_{t+1}}}, ..., y_{i_k x_{j_k}}|c)$ are given by Theorem 7 or experimental data if $k = 2$ and the bounds of $P(y_{i_t x_{i_t}}, x_p, y_q|c)$ are given by Theorem 6.

Calculation in the Example

Task 1 First by Theorem 7, we have,

$$\begin{aligned} 0 \leq P(y_{1_{x_1}}, y_{1_{x_2}}|c) &\leq 0.087, \\ 0 \leq P(y_{1_{x_1}}, y_{2_{x_2}}|c) &\leq 0.066, \\ 0 \leq P(y_{1_{x_1}}, y_{3_{x_2}}|c) &\leq 0.063, \\ 0.431 \leq P(y_{2_{x_1}}, y_{1_{x_2}}|c) &\leq 0.523, \\ 0.026 \leq P(y_{2_{x_1}}, y_{2_{x_2}}|c) &\leq 0.097, \\ 0.287 \leq P(y_{2_{x_1}}, y_{3_{x_2}}|c) &\leq 0.355, \\ 0 \leq P(y_{3_{x_1}}, y_{1_{x_2}}|c) &\leq 0.060, \\ 0 \leq P(y_{3_{x_1}}, y_{2_{x_2}}|c) &\leq 0.059, \\ 0 \leq P(y_{3_{x_1}}, y_{3_{x_2}}|c) &\leq 0.056. \end{aligned}$$

By Algorithm 2, the lower bound came from the following steps,

$$\begin{aligned} f(c) &= 0P(y_{1_{x_1}}, y_{1_{x_2}}|c) + P(y_{1_{x_1}}, y_{2_{x_2}}|c) + \\ &\quad P(y_{1_{x_1}}, y_{3_{x_2}}|c) - P(y_{2_{x_1}}, y_{1_{x_2}}|c) + \\ &\quad 0P(y_{2_{x_1}}, y_{2_{x_2}}|c) + P(y_{2_{x_1}}, y_{3_{x_2}}|c) - \\ &\quad -P(y_{3_{x_1}}, y_{1_{x_2}}|c) - P(y_{3_{x_1}}, y_{2_{x_2}}|c) + \\ &\quad 0P(y_{3_{x_1}}, y_{3_{x_2}}|c) \\ &= P(y_{1_{x_1}}, y_{2_{x_2}}|c) + \\ &\quad P(y_{1_{x_1}}, y_{3_{x_2}}|c) - P(y_{2_{x_1}}, y_{1_{x_2}}|c) + \\ &\quad 0P(y_{2_{x_1}}, y_{2_{x_2}}|c) + P(y_{2_{x_1}}, y_{3_{x_2}}|c) - \\ &\quad -P(y_{3_{x_1}}, y_{1_{x_2}}|c) - P(y_{3_{x_1}}, y_{2_{x_2}}|c) + \\ &\quad 0P(y_{3_{x_1}}, y_{3_{x_2}}|c) \\ &= P(y_{1_{x_1}}, y_{2_{x_2}}|c) + \\ &\quad P(y_{1_{x_1}}, y_{3_{x_2}}|c) - P(y_{2_{x_1}}, y_{1_{x_2}}|c) + \\ &\quad P(y_{2_{x_1}}, y_{3_{x_2}}|c) - \\ &\quad -P(y_{3_{x_1}}, y_{1_{x_2}}|c) - P(y_{3_{x_1}}, y_{2_{x_2}}|c) + \\ &\quad 0P(y_{3_{x_1}}, y_{3_{x_2}}|c) \\ &= P(y_{1_{x_1}}, y_{2_{x_2}}|c) - P(y_{2_{x_1}}, y_{1_{x_2}}|c) + \\ &\quad 0P(y_{2_{x_1}}, y_{3_{x_2}}|c) - \\ &\quad -P(y_{3_{x_1}}, y_{1_{x_2}}|c) - P(y_{3_{x_1}}, y_{2_{x_2}}|c) - \\ &\quad -P(y_{3_{x_1}}, y_{3_{x_2}}|c) + P(y_{3_{x_2}}|c) \\ &= P(y_{1_{x_1}}, y_{2_{x_2}}|c) - P(y_{2_{x_1}}, y_{1_{x_2}}|c) + \\ &\quad 0P(y_{2_{x_1}}, y_{3_{x_2}}|c) + 0P(y_{3_{x_1}}, y_{2_{x_2}}|c) + \\ &\quad 0P(y_{3_{x_1}}, y_{3_{x_2}}|c) + P(y_{3_{x_2}}|c) - P(y_{3_{x_1}}|c) \\ &\geq LB(P(y_{1_{x_1}}, y_{2_{x_2}}|c)) - UB(P(y_{2_{x_1}}, y_{1_{x_2}}|c)) + \\ &\quad +0 + 0 + 0 + 213/600 - 36/600 \\ &= 0 - 0.523 + 0.295 \\ &= -0.228. \end{aligned}$$

and the upper bounds came from the following steps,

$$\begin{aligned}
f(c) &= 0P(y_{1_{x_1}}, y_{1_{x_2}}|c) + P(y_{1_{x_1}}, y_{2_{x_2}}|c) + \\
&\quad P(y_{1_{x_1}}, y_{3_{x_2}}|c) - P(y_{2_{x_1}}, y_{1_{x_2}}|c) + \\
&\quad 0P(y_{2_{x_1}}, y_{2_{x_2}}|c) + P(y_{2_{x_1}}, y_{3_{x_2}}|c) - \\
&\quad -P(y_{3_{x_1}}, y_{1_{x_2}}|c) - P(y_{3_{x_1}}, y_{2_{x_2}}|c) + \\
&\quad 0P(y_{3_{x_1}}, y_{3_{x_2}}|c) \\
&= P(y_{1_{x_1}}, y_{2_{x_2}}|c) + \\
&\quad P(y_{1_{x_1}}, y_{3_{x_2}}|c) - P(y_{2_{x_1}}, y_{1_{x_2}}|c) + \\
&\quad 0P(y_{2_{x_1}}, y_{2_{x_2}}|c) + P(y_{2_{x_1}}, y_{3_{x_2}}|c) - \\
&\quad -P(y_{3_{x_1}}, y_{1_{x_2}}|c) - P(y_{3_{x_1}}, y_{2_{x_2}}|c) + \\
&\quad 0P(y_{3_{x_1}}, y_{3_{x_2}}|c) \\
&= P(y_{1_{x_1}}, y_{3_{x_2}}|c) - P(y_{2_{x_1}}, y_{1_{x_2}}|c) - \\
&\quad -P(y_{2_{x_1}}, y_{2_{x_2}}|c) + P(y_{2_{x_1}}, y_{3_{x_2}}|c) - \\
&\quad -P(y_{3_{x_1}}, y_{1_{x_2}}|c) - 2P(y_{3_{x_1}}, y_{2_{x_2}}|c) + \\
&\quad 0P(y_{3_{x_1}}, y_{3_{x_2}}|c) + P(y_{2_{x_2}}|c) \\
&= -P(y_{2_{x_1}}, y_{1_{x_2}}|c) - \\
&\quad -P(y_{2_{x_1}}, y_{2_{x_2}}|c) + 0P(y_{2_{x_1}}, y_{3_{x_2}}|c) - \\
&\quad -P(y_{3_{x_1}}, y_{1_{x_2}}|c) - 2P(y_{3_{x_1}}, y_{2_{x_2}}|c) - \\
&\quad -P(y_{3_{x_1}}, y_{3_{x_2}}|c) + P(y_{2_{x_2}}|c) + P(y_{3_{x_2}}|c) \\
&= 0P(y_{2_{x_1}}, y_{2_{x_2}}|c) + 1P(y_{2_{x_1}}, y_{3_{x_2}}|c) - \\
&\quad -P(y_{3_{x_1}}, y_{2_{x_2}}|c) + 0P(y_{3_{x_1}}, y_{3_{x_2}}|c) + \\
&\quad +P(y_{2_{x_2}}|c) + P(y_{3_{x_2}}|c) - P(y_{2_{x_1}}|c) - P(y_{3_{x_1}}|c) \\
&\leq 0 + UB(P(y_{2_{x_1}}, y_{3_{x_2}}|c)) - \\
&\quad -LB(P(y_{3_{x_1}}, y_{2_{x_2}}|c)) + 0 + \\
&\quad +58/600 + 213/600 - 512/600 - 36/600 \\
&= 0 + 0.355 - 0 + 0 - 0.462 \\
&= -0.107.
\end{aligned}$$

Thus,

$$-0.228 \leq f(c) \leq -0.107.$$

Task 2 By Algorithm 1, the result came from the following steps,

$$\begin{aligned}
f(c) &= 0P(y_{1_{x_1}}, y_{1_{x_2}}|c) + P(y_{1_{x_1}}, y_{2_{x_2}}|c) + \\
&\quad 2P(y_{1_{x_1}}, y_{3_{x_2}}|c) - P(y_{2_{x_1}}, y_{1_{x_2}}|c) + \\
&\quad 0P(y_{2_{x_1}}, y_{2_{x_2}}|c) + P(y_{2_{x_1}}, y_{3_{x_2}}|c) - \\
&\quad -2P(y_{3_{x_1}}, y_{1_{x_2}}|c) - P(y_{3_{x_1}}, y_{2_{x_2}}|c) + \\
&\quad 0P(y_{3_{x_1}}, y_{3_{x_2}}|c) \\
&= P(y_{1_{x_1}}, y_{2_{x_2}}|c) + \\
&\quad 2P(y_{1_{x_1}}, y_{3_{x_2}}|c) - P(y_{2_{x_1}}, y_{1_{x_2}}|c) + \\
&\quad 0P(y_{2_{x_1}}, y_{2_{x_2}}|c) + P(y_{2_{x_1}}, y_{3_{x_2}}|c) - \\
&\quad -2P(y_{3_{x_1}}, y_{1_{x_2}}|c) - P(y_{3_{x_1}}, y_{2_{x_2}}|c) + \\
&\quad 0P(y_{3_{x_1}}, y_{3_{x_2}}|c) \\
&= 2P(y_{1_{x_1}}, y_{3_{x_2}}|c) - P(y_{2_{x_1}}, y_{1_{x_2}}|c) - \\
&\quad -P(y_{2_{x_1}}, y_{2_{x_2}}|c) + P(y_{2_{x_1}}, y_{3_{x_2}}|c) - \\
&\quad -2P(y_{3_{x_1}}, y_{1_{x_2}}|c) - 2P(y_{3_{x_1}}, y_{2_{x_2}}|c) + \\
&\quad 0P(y_{3_{x_1}}, y_{3_{x_2}}|c) + P(y_{2_{x_2}}|c) \\
&= -P(y_{2_{x_1}}, y_{1_{x_2}}|c) - \\
&\quad -P(y_{2_{x_1}}, y_{2_{x_2}}|c) - P(y_{2_{x_1}}, y_{3_{x_2}}|c) - \\
&\quad -2P(y_{3_{x_1}}, y_{1_{x_2}}|c) - 2P(y_{3_{x_1}}, y_{2_{x_2}}|c) - \\
&\quad -2P(y_{3_{x_1}}, y_{3_{x_2}}|c) + P(y_{2_{x_2}}|c) + 2P(y_{3_{x_2}}|c) \\
&= 0P(y_{2_{x_1}}, y_{2_{x_2}}|c) + 0P(y_{2_{x_1}}, y_{3_{x_2}}|c) - \\
&\quad -2P(y_{3_{x_1}}, y_{1_{x_2}}|c) - 2P(y_{3_{x_1}}, y_{2_{x_2}}|c) - \\
&\quad -2P(y_{3_{x_1}}, y_{3_{x_2}}|c) + \\
&\quad +P(y_{2_{x_2}}|c) + 2P(y_{3_{x_2}}|c) - P(y_{2_{x_1}}|c) \\
&= 0P(y_{2_{x_1}}, y_{2_{x_2}}|c) + 0P(y_{2_{x_1}}, y_{3_{x_2}}|c) + \\
&\quad +0P(y_{3_{x_1}}, y_{2_{x_2}}|c) + 0P(y_{3_{x_1}}, y_{3_{x_2}}|c) + \\
&\quad P(y_{2_{x_2}}|c) + 2P(y_{3_{x_2}}|c) - \\
&\quad -P(y_{2_{x_1}}|c) - 2P(y_{3_{x_1}}|c) \\
&= 58/600 + 426/600 - 512/600 - 72/600 \\
&= -0.167.
\end{aligned}$$

Distribution Generating Algorithm

Here, the sample distribution generating algorithm in the simulated studies is presented. It generated both experimental and observational data compatible with the fractions of response types of individuals. The data satisfy the general relation between experimental and observational data. Note that all four simulated studies shared the same distribution generating algorithm but with different benefit vectors.

Algorithm 3: Generate sample distributions for simulated studies

Input: num, number of samples needed.

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

```
$count = 0$;
while $count < num$ do
    //$rand(0, 1)$ is the function that random uniformly generate a number from 0 to 1.
    $a = []$;
    for $i = 1$ to 8 do
        $a.append(rand(0, 1))$;
    end for
    $a.append(1.0)$;
    $a.sort()$;
    // Each $c_k$ corresponding to a sample distribution.
    $k = count$;
    //$f$ is the fractions of response types of individuals, $f[0] = P(y_{1_{x_1}}, y_{1_{x_2}}|c_k), ..., f[8] = P(y_{3_{x_1}}, y_{3_{x_2}}|c_k)$.
    $f = []$;
    $f[0] = a[0]$;
    for $i = 1$ to 8 do
        $f[i] = a[i] - a[i - 1]$;
    end for
    // Generate experimental data.
    $P(y_{1_{x_1}}|c_k) = f[0] + f[1] + f[2]$;
    $P(y_{2_{x_1}}|c_k) = f[3] + f[4] + f[5]$;
    $P(y_{3_{x_1}}|c_k) = f[6] + f[7] + f[8]$;
    $P(y_{1_{x_2}}|c_k) = f[0] + f[3] + f[6]$;
    $P(y_{2_{x_2}}|c_k) = f[1] + f[4] + f[7]$;
    $P(y_{3_{x_2}}|c_k) = f[2] + f[5] + f[8]$;
    // Generate observational data.
    $P(x_1, y_1|c_k) = rand(0, P(y_{1_{x_1}}|c_k))$;
    $P(x_1, y_2|c_k) = rand(0, P(y_{2_{x_1}}|c_k))$;
    $P(x_1|c_k) = rand(P(x_1, y_1|c_k) + P(x_1, y_2|c_k), \min\{P(x_1, y_1|c_k) + 1 - P(y_{1_{x_1}}|c_k), P(x_1, y_2|c_k) + 1 - P(y_{1_{x_2}}|c_k)\})$;
    $P(x_1, y_3|c_k) = P(x_1|c_k) - P(x_1, y_1|c_k) - P(x_1, y_2|c_k)$;
    $P(x_2|c_k) = 1 - P(x_1|c_k)$
    $P(x_2, y_1|c_k) = rand(0, \min\{P(y_{1_{x_2}}|c_k), P(x_2|c_k)\})$;
    $P(x_2, y_2|c_k) = rand(0, \min\{P(y_{2_{x_2}}|c_k), P(x_2|c_k) - P(x_2, y_1|c_k)\})$;
    $P(x_2, y_3|c_k) = P(x_2|c_k) - P(x_2, y_1|c_k) - P(x_2, y_2|c_k)$;
    //Validate the data, the experimental data and observational data should satisfies the following: $P(x, y|c_k) \leq P(y_x|c_k) \leq P(x, y|c_k) + 1 - P(x|c_k)$.
    $mark = True$
    for $i = 1$ to 3 do
        for $j = 1$ to 2 do
            if $P(y_{i_{x_j}}|c_k) < P(x_j, y_i|c_k)$ or $P(y_{i_{x_j}}|c_k) > P(x_j, y_i|c_k) + 1 - P(x_j|c_k)$ then
                $mark = False$;
            end if
        end for
    end for
    if $mark == False$ then
        $continue$;
    end if
    $count = count + 1$;
end while
```