CRAB Proof Extensions

Ved Sriraman

December 2024

1 Introduction

This document contains extensions of proofs from the CRAB and LEWIS papers (Galhotra et. al).

2 Extending Section 3 of CRAB

2.1 Extension of Proposition 3.3

We first focus on the case of no external information. We will explore how to deal with uncertainty in multiple edges in a causal graph, and how that affects our bounds on the target distribution and thus the fairness bounds.

2.1.1 Preliminaries

Recall that we define fairness as such:

$$\mathsf{F}(\Omega) = \frac{1}{|\mathbf{A}|} \sum_{\mathbf{a} \in Dom(\mathbf{A})} \Pr^{+}_{\Omega}(h(x) \mid s_1, \mathbf{a}) - \Pr^{+}_{\Omega}(h(x) \mid s_0, \mathbf{a})$$

We will now examine two conditions in a causal graph for "perfect fairness" (PF).

PF1. If $(h(x) \perp\!\!\!\perp_d S \mid \mathbf{A})$, then $\Pr_{\Omega}(h(x) \mid S, \mathbf{A}) = \Pr_{\Omega}(h(x) \mid \mathbf{A}) \implies \mathsf{F}(\Omega) = 0$.

That is, if $h(x)$ is d-separated from S by $\mathbf{A}$, then we achieve perfect fairness as measured by $\mathsf{F}(\Omega)$.

PF2. If S is somehow isolated from $h(x)$, meaning that there is no path between S and $h(x)$, then $\Pr_{\Omega}(h(x) \mid S, \mathbf{A}) = \Pr_{\Omega}(h(x) \mid \mathbf{A}) \implies \mathsf{F}(\Omega) = 0$.

That is, knowing anything about S tells us absolutely nothing about Y, so we achieve perfect fairness as measured by $\mathsf{F}(\Omega)$.

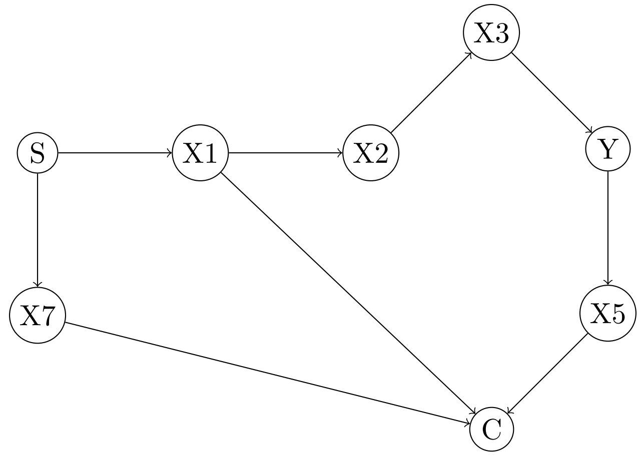
We will now apply these conditions to prove a statement about the conditional independence requirement in Proposition 3.3.

2.1.2 Uncertainty in only edge

Consider the causal graph $\mathcal{G} = (E, V)$ below that represents the true underlying model.

Here, we assume that all $x \in X$ are used to compute the target variable $h(x)$. Suppose that in a real-world application, we are unsure about the uncertainty of some causal link $X_i \to C$, and we can estimate from observational data that this exists in our causal graph with probability p. If $\mathcal{G}' = (E, V)$, with edges and nodes identical to the original causal graph $\mathcal{G}$, then we denote this uncertain edge as e_p. This represents our belief that the edge e_p exists with probability p. Note that, if $p = 1$, the edge is no longer uncertain, and $\mathcal{G}' = \mathcal{G}$. Let us also define that for the modified graph $\mathcal{G}'$,

$$E_{\mathbf{U}} = \left\{ e = (x_1, x_2) \in E \mid x_1 \in \mathbf{U} \text{ and } x_2 = C \right\}$$



That is, $E_{\mathbf{U}}$ is the set of all edges having a source in the set $\mathbf{U}$ from the original graph $\mathcal{G}$ and the sink as C itself. We want to show that if $e_p \notin E_{\mathbf{U}}$, $(\mathbf{X} \perp\!\!\!\perp C \,|\, S, \mathbf{A}, \mathbf{U})$ still holds true. We will assume that neither C nor Y are parents of any other nodes in the causal graph.

Lemma 2.1. *Let $\mathcal{G}$ be the original causal graph with no uncertain edges and let $\mathcal{G}'$ be the modified causal graph with uncertain edge e_p. The independence condition $(\boldsymbol{X} \perp\!\!\!\perp C \,|\, S, \boldsymbol{A}, \boldsymbol{U})$ still holds in $\mathcal{G}'$ if we use the same $\mathbf{U}$ that satisfied the independence condition in $\mathcal{G}$, and the fairness query on the modified graph $\mathcal{G}'$ will be bounded by the fairness query on the original graph $\mathcal{G}$.*

Proof. All edges in $E_{\mathbf{U}}$ exist with certainty. Since $\mathbf{U}$ still forms the Markov boundary of C, conditioning on $\mathbf{U}$ will d-separate any other node in X from C. Thus, $(X \perp\!\!\!\perp_d C \mid S, \mathbf{A}, \mathbf{U}) \implies (\mathbf{X} \perp\!\!\!\perp C \,|\, S, \mathbf{A}, \mathbf{U})$. □

Now, we can leverage this Lemma to derive improved bounds on the fairness query. WLOG, assume $p \in [0, \frac{1}{2}]$.

- If w.p. p we have a graph $\mathcal{G}'$ that satisfies *PF1* or *PF2* $\implies \mathsf{F}(\Omega) = 0$, we get the following two cases:

Case 1. if, w.p. $1-p$, we have a graph $\mathcal{G}'$ that satisfies *PF1* or *PF2*
$\implies \mathsf{F}(\Omega) = 0$,

Case 2. else, w.p. $1-p$, we have a graph in which $U, S, \mathbf{A}$ are unchanged, yet neither *PF1* nor *PF2* are satisfied
$\implies \mathsf{F}(\Omega) \leq (1-p) \cdot CUB$

- Else, w.p. p, we have a graph in which $U, S, \mathbf{A}$ are unchanged, yet neither *PF1* nor *PF2* are satisfied $\implies \mathsf{F}(\Omega) \leq p \cdot CUB$

Case 3. if, w.p. $1-p$, we have a graph $\mathcal{G}'$ that satisfies *PF1* or *PF2*
$\implies \mathsf{F}(\Omega) = 0$,

Case 4. else w.p. $1-p$ we have a graph in which $\mathbf{U}, S, \mathbf{A}$ are unchanged, yet neither *PF1* nor *PF2* are satisfied
$\implies \mathsf{F}(\Omega) \leq (1-p) \cdot CUB$

Case 1 exhibits perfect fairness because we obtain $\mathsf{F}(\Omega) = 0$ w.p. 1. Case 4 experiences no change in the fairness query as compared to the graph $\mathcal{G}$. Cases 2 and 3 achieve improved fairness bounds of $\mathsf{F}(\Omega) \leq (1-p) \cdot CUB$ and $\mathsf{F}(\Omega) \leq p \cdot CUB$, respectively.

Thus, by Lemma 2.1, we do not have to consider any uncertainty in edges that do not change the Markov boundary of C, i.e., a valid set U. We can shift our focus to the case where there exist uncertain edges in the set $E_{\mathbf{U}}$. For the singular case, take the example of one uncertain edge $e_p \in E_{\mathbf{U}}$ with source u_1. Let $\mathbf{U}' = \mathbf{U} \setminus \{u_1\}$. With a naive analysis, we can form the following consistent upper bound (CUB) as defined in the CRAB paper:

$$CUB = \frac{1}{|A|} \sum_{\mathbf{a} \in \mathrm{Dom}(\mathbf{A})} \left(p \cdot \Big(\max_{\mathbf{u} \in \mathrm{Dom}(\mathbf{U})} \Pr^{+}_{\Delta}(h(x) \mid s_1, \mathbf{a}, \mathbf{u}) - \min_{\mathbf{u} \in \mathrm{Dom}(\mathbf{U})} \Pr^{+}_{\Delta}(h(x) \mid s_0, \mathbf{a}, \mathbf{u}) \Big) \right)$$

$$+\frac{1}{|A|}\sum_{\mathbf{a}\in\mathrm{Dom}(\mathbf{A})}\left((1-p)\cdot\Big(\max_{\mathbf{u}'\in\mathrm{Dom}(\mathbf{U}')}\Pr^{+}_{\Delta}(h(x)\mid s_1,\mathbf{a},\mathbf{u}')-\min_{\mathbf{u}'\in\mathrm{Dom}(\mathbf{U}')}\Pr^{+}_{\Delta}(h(x)\mid s_0,\mathbf{a},\mathbf{u}')\Big)\right)$$

We will prove an upper bound for each $\mathbf{a}\in\mathbf{A}$ in the definition of $\mathsf{F}_{h,\mathbf{A}}(\Omega)$, which extends to showing a lower bound and providing the desired result above.

Proof.

$$\begin{aligned}
\Pr_\Omega(h(x)\mid s_1,\mathbf{a}) &= p\cdot\sum_{\mathbf{u}\in\mathrm{Dom}(\mathbf{U})}\Pr_\Omega(h(x)\mid s_1,\mathbf{a},\mathbf{u})\cdot\Pr_\Omega(\mathbf{u}\mid s_1,\mathbf{a})\\
&+(1-p)\cdot\sum_{\mathbf{u}'\in\mathrm{Dom}(\mathbf{U}')}\Pr_\Omega(h(x)\mid s_1,\mathbf{a},\mathbf{u}')\cdot\Pr_\Omega(\mathbf{u}'\mid s_1,\mathbf{a})\\
&= p\cdot\sum_{\mathbf{u}\in\mathrm{Dom}(\mathbf{U})}\Pr_\Omega(h(x)\mid s_1,\mathbf{a},\mathbf{u},C=1)\cdot\Pr_\Omega(\mathbf{u}\mid s_1,\mathbf{a})\\
&+(1-p)\cdot\sum_{\mathbf{u}'\in\mathrm{Dom}(\mathbf{U}')}\Pr_\Omega(h(x)\mid s_1,\mathbf{a},\mathbf{u}',C=1)\cdot\Pr_\Omega(\mathbf{u}'\mid s_1,\mathbf{a})\\
&\le p\cdot\sum_{\mathbf{u}\in\mathrm{Dom}(\mathbf{U})}\max_{u\in\mathrm{Dom}(\mathbf{U})}\Pr_\Omega(h(x)\mid s_1,\mathbf{a},\mathbf{u},C=1)\cdot\Pr_\Omega(\mathbf{u}\mid s_1,\mathbf{a})\\
&+(1-p)\cdot\sum_{\mathbf{u}'\in\mathrm{Dom}(\mathbf{U}')}\max_{u'\in\mathrm{Dom}(\mathbf{U}')}\Pr_\Omega(h(x)\mid s_1,\mathbf{a},\mathbf{u}',C=1)\cdot\Pr_\Omega(\mathbf{u}'\mid s_1,\mathbf{a})\\
&= p\cdot\max_{u\in\mathrm{Dom}(\mathbf{U})}\Pr_\Omega(h(x)\mid s_1,\mathbf{a},\mathbf{u},C=1)\sum_{\mathbf{u}\in\mathrm{Dom}(\mathbf{U})}\Pr_\Omega(\mathbf{u}\mid s_1,\mathbf{a})\\
&+(1-p)\cdot\max_{u'\in\mathrm{Dom}(\mathbf{U}')}\Pr_\Omega(h(x)\mid s_1,\mathbf{a},\mathbf{u}',C=1)\cdot\sum_{\mathbf{u}'\in\mathrm{Dom}(\mathbf{U}')}\Pr_\Omega(\mathbf{u}'\mid s_1,\mathbf{a})\\
&= p\cdot\Pr_\Delta(h(x)\mid s_1,\mathbf{a},\mathbf{u}^*)+(1-p)\cdot\Pr_\Delta(h(x)\mid s_1,\mathbf{a},\mathbf{u}'^*)
\end{aligned}$$

□

This naive approach requires us to iterate through possibilities of $\mathbf{u}$ and $\mathbf{u}'$. Instead, we can show a simplified approach, which allows us to form the following CUB:

$$CUB=\frac{1}{|A|}\sum_{\mathbf{a}\in\mathrm{Dom}(\mathbf{A})}\left((1+p\cdot\beta-p)\cdot\max_{\mathbf{u}'\in\mathrm{Dom}(\mathbf{U}')}\Pr^{+}_{\Delta}(h(x)\mid s_1,\mathbf{a},\mathbf{u}')-(1+p\cdot\beta-p)\cdot\min_{\mathbf{u}'\in\mathrm{Dom}(\mathbf{U}')}\Pr^{+}_{\Delta}(h(x)\mid s_0,\mathbf{a},\mathbf{u}')\right)$$

Before we begin the proof, we will show some auxiliary work. We will consider $\mathbf{u}\in\mathrm{Dom}(\mathbf{U})$ and $\mathbf{u}'\in\mathrm{Dom}(\mathbf{U}')$, and abbreviate $\Pr_\Omega(X)$ as $\Pr(X)$ in order to show the following simplification:

$$\begin{aligned}
\sum_{\mathbf{u}}\Pr(h(x)\mid s_1,\mathbf{a},\mathbf{u},C=1)\cdot\Pr(\mathbf{u}\mid s_1,\mathbf{a}) &= \sum_{\mathbf{u}}\Pr(h(x)\mid s_1,\mathbf{a},\mathbf{u}'\wedge u_1,C=1)\cdot\Pr(\mathbf{u}'\wedge u_1\mid s_1,\mathbf{a}) &(1)\\
&=\sum_{\mathbf{u}}\frac{\Pr(h(x),s_1,\mathbf{a},\mathbf{u}',u_1,C=1)}{\Pr(s_1,\mathbf{a},\mathbf{u}',u_1,C=1)}\cdot\frac{\Pr(\mathbf{u}',u_1,s_1,\mathbf{a})}{\Pr(s_1,\mathbf{a})} &(2)\\
&=\sum_{\mathbf{u}}\frac{\Pr(h(x),s_1,\mathbf{a},\mathbf{u}',u_1,C=1)}{\Pr(s_1,\mathbf{a},\mathbf{u}',u_1,C=1)}\cdot\frac{\Pr(\mathbf{u}',u_1,s_1,\mathbf{a})}{\Pr(s_1,\mathbf{a})}\cdot\frac{\Pr(\mathbf{u}',s_1,\mathbf{a},C=1)}{\Pr(\mathbf{u}',s_1,\mathbf{a},C=1)} &(3)\\
&=\sum_{\mathbf{u}}\Pr(h(x),u_1\mid s_1,\mathbf{a},\mathbf{u}',C=1)\cdot\Pr(\mathbf{u}',C=1\mid s_1,\mathbf{a})
\end{aligned}$$

$$
\begin{aligned}
&\quad \cdot \frac{\Pr(\mathbf{u}', u_1, s_1, \mathbf{a})}{\Pr(\mathbf{u}', u_1, s_1, \mathbf{a}, C=1)} && (4)\\
&\le \sum_{\mathbf{u}'} \Pr(h(x) \mid s_1, \mathbf{a}, \mathbf{u}', C=1) \cdot \Pr(\mathbf{u}', C=1 \mid s_1, \mathbf{a}) \\
&\quad \cdot \left\{ \frac{\Pr(\mathbf{u}', u_1=0, s_1, \mathbf{a})}{\Pr(\mathbf{u}', u_1=0, s_1, \mathbf{a}, C=1)} + \frac{\Pr(\mathbf{u}', u_1=1, s_1, \mathbf{a})}{\Pr(\mathbf{u}', u_1=1, s_1, \mathbf{a}, C=1)} \right\} && (5)\\
&\le \sum_{\mathbf{u}'} \Bigg[\max_{\mathbf{u}'} \Pr(h(x) \mid s_1, \mathbf{a}, \mathbf{u}', C=1) \\
&\quad \cdot \left\{ \frac{\Pr(\mathbf{u}', u_1=0, s_1, \mathbf{a})}{\Pr(\mathbf{u}', u_1=0, s_1, \mathbf{a}, C=1)} + \frac{\Pr(\mathbf{u}', u_1=1, s_1, \mathbf{a})}{\Pr(\mathbf{u}', u_1=1, s_1, \mathbf{a}, C=1)} \right\} \Bigg] \\
&\quad \cdot \Pr(\mathbf{u}', C=1 \mid s_1, \mathbf{a}) && (6)\\
&= \Pr(h(x) \mid s_1, \mathbf{a}, \mathbf{u}'^*, C=1) \\
&\quad \cdot \left\{ \frac{\Pr(\mathbf{u}'^*, u_1=0, s_1, \mathbf{a})}{\Pr(\mathbf{u}'^*, u_1=0, s_1, \mathbf{a}, C=1)} + \frac{\Pr(\mathbf{u}'^*, u_1=1, s_1, \mathbf{a})}{\Pr(\mathbf{u}'^*, u_1=1, s_1, \mathbf{a}, C=1)} \right\} \\
&\quad \cdot \sum_{\mathbf{u}'} \Pr(\mathbf{u}', C=1 \mid s_1, \mathbf{a}) && (7)\\
&= \Pr(h(x) \mid s_1, \mathbf{a}, \mathbf{u}'^*, C=1) \cdot \beta \cdot \sum_{\mathbf{u}'} \Pr(\mathbf{u}', C=1 \mid s_1, \mathbf{a}) && (8)\\
&= \beta \cdot \Pr(h(x) \mid s_1, \mathbf{a}, \mathbf{u}'^*, C=1) \cdot \Pr(C=1 \mid s_1, \mathbf{a}) && (9)\\
&= \beta \cdot \Pr(h(x) \mid s_1, \mathbf{a}, \mathbf{u}'^*, C=1) \cdot (1 - \Pr(C=0 \mid s_1, \mathbf{a})) && (10)\\
&\le \beta \cdot \Pr(h(x) \mid s_1, \mathbf{a}, \mathbf{u}'^*, C=1) && (11)\\
&= \beta \cdot \Pr_\Delta(h(x) \mid s_1, \mathbf{a}, \mathbf{u}'^*) && (12)
\end{aligned}
$$

Now, we can provide an upper bound on the expected value of the fairness query given one uncertain edge as follows:

Proof.

$$
\begin{aligned}
\Pr_\Omega(h(x) \mid s_1, \mathbf{a}) &= p \cdot \sum_{\mathbf{u} \in \mathrm{Dom}(\mathbf{U})} \Pr_\Omega(h(x) \mid s_1, \mathbf{a}, \mathbf{u}) \cdot \Pr_\Omega(\mathbf{u} \mid s_1, \mathbf{a}) \\
&\quad + (1-p) \cdot \sum_{\mathbf{u}' \in \mathrm{Dom}(\mathbf{U}')} \Pr_\Omega(h(x) \mid s_1, \mathbf{a}, \mathbf{u}') \cdot \Pr_\Omega(\mathbf{u}' \mid s_1, \mathbf{a}) \\
&= p \cdot \sum_{\mathbf{u} \in \mathrm{Dom}(\mathbf{U})} \Pr_\Omega(h(x) \mid s_1, \mathbf{a}, \mathbf{u}, C=1) \cdot \Pr_\Omega(\mathbf{u} \mid s_1, \mathbf{a}) \\
&\quad + (1-p) \cdot \sum_{\mathbf{u}' \in \mathrm{Dom}(\mathbf{U}')} \Pr_\Omega(h(x) \mid s_1, \mathbf{a}, \mathbf{u}', C=1) \cdot \Pr_\Omega(\mathbf{u}' \mid s_1, \mathbf{a}) \\
&\le p \cdot \beta \cdot \Pr_\Delta(h(x) \mid s_1, \mathbf{a}, \mathbf{u}'^*) + (1-p) \cdot \Pr_\Delta(h(x) \mid s_1, \mathbf{a}, \mathbf{u}'^*) \\
&= (1 + p \cdot \beta - p) \cdot \Pr_\Delta(h(x) \mid s_1, \mathbf{a}, \mathbf{u}'^*)
\end{aligned}
$$

We can show the result above by applying similar logic to find the minimum. □

However, upon taking a closer look at the β factor in this second analysis, we note that this bound is impractical because it is not possible to calculate $\Pr(\mathbf{u}'^*, u_1 = 0, s_1, \mathbf{a})$ or $\Pr(\mathbf{u}'^*, u_1 = 1, s_1, \mathbf{a})$ from observational data under the influence of selection bias. Instead, we can make a slight adjustment to the first part of our previous analysis, and we will use the following result to upper bound $\Pr_\Omega(h(x) \mid s_1, \mathbf{a})$ itself.

$$
\sum_{\mathbf{u}} \Pr(h(x) \mid s_1, \mathbf{a}, \mathbf{u}, C=1) \cdot \Pr(\mathbf{u} \mid s_1, \mathbf{a}) = \sum_{\mathbf{u}} \Pr(h(x) \mid s_1, \mathbf{a}, \mathbf{u}' \wedge u_1, C=1) \cdot \Pr(\mathbf{u}' \wedge u_1 \mid s_1, \mathbf{a})
$$

$$
\begin{aligned}
&= \sum_{\mathbf{u}} \frac{\Pr(h(x), s_1, \mathbf{a}, \mathbf{u}', u_1, C=1)}{\Pr(s_1, \mathbf{a}, \mathbf{u}', u_1, C=1)} \cdot \frac{\Pr(\mathbf{u}', u_1, s_1, \mathbf{a})}{\Pr(s_1, \mathbf{a})} \\
&= \sum_{\mathbf{u}} \frac{\Pr(h(x), s_1, \mathbf{a}, \mathbf{u}', u_1, C=1)}{\Pr(s_1, \mathbf{a}, \mathbf{u}', u_1, C=1)} \cdot \frac{\Pr(\mathbf{u}', u_1, s_1, \mathbf{a})}{\Pr(s_1, \mathbf{a})} \cdot \frac{\Pr(\mathbf{u}', s_1, \mathbf{a}, C=1)}{\Pr(\mathbf{u}', s_1, \mathbf{a}, C=1)} \\
&= \sum_{\mathbf{u}} \frac{\Pr(h(x), u_1 \mid s_1, \mathbf{a}, \mathbf{u}', C=1)}{\Pr(C=1 \mid s_1, u_1, \mathbf{u}', \mathbf{a})} \cdot \Pr(C=1, \mathbf{u}' \mid s_1, \mathbf{a}) \\
&= \sum_{\mathbf{u}'} \sum_{\mathbf{u_1}} \frac{\Pr(h(x), u_1 \mid s_1, \mathbf{a}, \mathbf{u}', C=1)}{\Pr(C=1 \mid u_1, \mathbf{u}')} \cdot \Pr(C=1, \mathbf{u}' \mid s_1, \mathbf{a}) \\
&\leq \sum_{\mathbf{u}'} \sum_{\mathbf{u_1}} \frac{\Pr(h(x), u_1 \mid s_1, \mathbf{a}, \mathbf{u}', C=1)}{\min_{\mathbf{u}} \Pr(C=1 \mid \mathbf{u})} \cdot \Pr(C=1, \mathbf{u}' \mid s_1, \mathbf{a}) \\
&\leq \frac{1}{\min_{\mathbf{u}} \Pr(C=1 \mid \mathbf{u})} \sum_{\mathbf{u}'} \sum_{\mathbf{u_1}} \Pr(h(x), u_1 \mid s_1, \mathbf{a}, \mathbf{u}', C=1) \cdot \Pr(C=1, \mathbf{u}' \mid s_1, \mathbf{a} \\
&= \frac{1}{\min_{\mathbf{u}} \Pr(C=1 \mid \mathbf{u})} \sum_{\mathbf{u}'} \Pr(C=1, \mathbf{u}' \mid s_1, \mathbf{a}) \sum_{\mathbf{u_1}} \Pr(h(x), u_1 \mid s_1, \mathbf{a}, \mathbf{u}', C=1) \\
&= \frac{1}{\min_{\mathbf{u}} \Pr(C=1 \mid \mathbf{u})} \sum_{\mathbf{u}'} \Pr(C=1, \mathbf{u}' \mid s_1, \mathbf{a}) \cdot \Pr(h(x) \mid s_1, \mathbf{a}, \mathbf{u}', C=1) \\
&\leq \frac{\max_{\mathbf{u}'} \Pr(h(x) \mid s_1, \mathbf{a}, \mathbf{u}', C=1)}{\min_{\mathbf{u}} \Pr(C=1, \mathbf{u}' \mid \mathbf{u})} \cdot \sum_{\mathbf{u}'} \Pr(C=1, \mathbf{u}' \mid s_1, \mathbf{a}) \\
&= \frac{\max_{\mathbf{u}'} \Pr(h(x) \mid s_1, \mathbf{a}, \mathbf{u}', C=1)}{\min_{\mathbf{u}} \Pr(C=1 \mid \mathbf{u})} \cdot \Pr(C=1 \mid s_1, \mathbf{a}) \\
&= \frac{\Pr(C=1 \mid s_1, \mathbf{a})}{\min_{\mathbf{u}} \Pr(C=1 \mid \mathbf{u})} \cdot \max_{\mathbf{u}'} \Pr_{\Delta}(h(x) \mid s_1, \mathbf{a}, \mathbf{u}') \\
&= \alpha \cdot \Pr_{\Delta}(h(x) \mid s_1, \mathbf{a}, \mathbf{u}')
\end{aligned}
$$

Next, we will form the new CUB by considering the uncertainty in one edge. In the case of one uncertain edge, we will take the maximum of the case in which u_1 exists in the causal graph and the case in which u_1 does not exist.

$$
\begin{aligned}
\Pr_{\Omega}(h(x) \mid s_1, \mathbf{a}) &\leq \max \left\{ \sum_{\mathbf{u}} \Pr(h(x) \mid s_1, \mathbf{a}, \mathbf{u}, C=1) \cdot \Pr(\mathbf{u} \mid s_1, \mathbf{a}), \sum_{\mathbf{u}'} \Pr(h(x) \mid s_1, \mathbf{a}, \mathbf{u}', C=1) \cdot \Pr(\mathbf{u}' \mid s_1, \mathbf{a}) \right\} \\
&= \max \left\{ \alpha \cdot \max_{\mathbf{u}'} \Pr_{\Delta}(h(x) \mid s_1, \mathbf{a}, \mathbf{u}'), \max_{\mathbf{u}'} \Pr_{\Delta}(h(x) \mid s_1, \mathbf{a}, \mathbf{u}') \right\} \\
&= \max \{\alpha, 1\} \cdot \max_{\mathbf{u}'} \Pr_{\Delta}(h(x) \mid s_1, \mathbf{a}, \mathbf{u}')
\end{aligned}
$$

Thus, we have the following CUB:

$$
CUB = \frac{1}{|A|} \sum_{\mathbf{a} \in \mathrm{Dom}(\mathbf{A})} \left(\max\{\alpha, 1\} \cdot \max_{\mathbf{u}' \in \mathrm{Dom}(\mathbf{U}')} \Pr_{\Delta}^{+}(h(x) \mid s_1, \mathbf{a}, \mathbf{u}') - \min\{\alpha, 1\} \cdot \min_{\mathbf{u}' \in \mathrm{Dom}(\mathbf{U}')} \Pr_{\Delta}^{+}(h(x) \mid s_0, \mathbf{a}, \mathbf{u}') \right)
$$

This approach is feasible because we can calculate $\min_{\mathbf{u}} \Pr(C = 1 \mid \mathbf{u})$ easily through observational data, and our analysis is tighter. This is less intensive than calculating the conditional distribution we approximated over $\mathbf{u} \in \mathbf{U}$. In this analysis, the only approximation we have to make is on the conditional distribution of the target $h(x)$ by dropping the intersection with u_1 to reduce the complexity.

2.1.3 Uncertainty in multiple edges

With multiple uncertainties in causal links provided by nodes in $\mathbf{U}$, we must consider the maximum over all possible sets $\mathbf{U}$. If there are ℓ uncertain edges, then 2^{ℓ} observable sets of $\mathbf{U}$ are possible, all of which include the nodes with no uncertainty. Consider an example with two uncertain edges from u_1 and u_2, and $\mathbf{u}' \in \mathbf{U}' = \mathbf{U} \setminus \{u_1, u_2\}$. We will take the max over four separate cases.

$$\Pr_\Omega(h(x) \mid s_1, \mathbf{a}) \leq \max \Bigg\{ \sum_{\mathbf{u}'} \Pr(h(x) \mid s_1, \mathbf{a}, \mathbf{u}', C = 1) \cdot \Pr(\mathbf{u}' \mid s_1, \mathbf{a}),$$
$$\sum_{\mathbf{u}' \wedge u_1} \Pr(h(x) \mid s_1, \mathbf{a}, \mathbf{u}' \wedge u_1, C = 1) \cdot \Pr(\mathbf{u}' \wedge u_1 \mid s_1, \mathbf{a}),$$
$$\sum_{\mathbf{u}' \wedge u_2} \Pr(h(x) \mid s_1, \mathbf{a}, \mathbf{u}' \wedge u_2, C = 1) \cdot \Pr(\mathbf{u}' \wedge u_2 \mid s_1, \mathbf{a}),$$
$$\sum_{\mathbf{u}} \Pr(h(x) \mid s_1, \mathbf{a}, \mathbf{u}, C = 1) \cdot \Pr(\mathbf{u} \mid s_1, \mathbf{a}) \Bigg\}$$

We will simplify each of the four cases below:

Case 1. $u_1, u_2 \notin \mathbf{U}$ We cite the proof of Proposition 3.1 from the CRAB paper, since $(\mathbf{X} \perp\!\!\!\perp C \,|\, S, \mathbf{A}, \mathbf{U})$ is satisfied.

$$\sum_{\mathbf{u}'} \Pr(h(x) \mid s_1, \mathbf{a}, \mathbf{u}', C = 1) \cdot \Pr(\mathbf{u}' \mid s_1, \mathbf{a}) \leq \max_{\mathbf{u}'} \Pr_\Delta(h(x) \mid s_1, \mathbf{a}, \mathbf{u}')$$

Case 2. $u_1 \in \mathbf{U}, u_2 \notin \mathbf{U}$ The result extends exactly from the work in Case 1. However, step (7) is less lossy because we only discard u_2 from the conditional joint distribution.

$$\sum_{\mathbf{u}' \wedge u_1} \Pr(h(x) \mid s_1, \mathbf{a}, \mathbf{u}' \wedge u_1, C = 1) \cdot \Pr(\mathbf{u}' \wedge u_1 \mid s_1, \mathbf{a}) = \sum_{\mathbf{u}' \wedge u_1} \Pr(h(x) \mid s_1, \mathbf{a}, \mathbf{u}', u_1, C = 1) \cdot \Pr(\mathbf{u}'\mathbf{R}, u_1 \mid s_1, \mathbf{a}) \tag{1}$$
$$= \sum_{\mathbf{u}' \wedge u_1} \frac{\Pr(h(x), s_1, \mathbf{a}, \mathbf{u}', u_1, C = 1)}{\Pr(s_1, \mathbf{a}, \mathbf{u}', u_1, C = 1)} \cdot \frac{\Pr(\mathbf{u}', u_1, s_1, \mathbf{a})}{\Pr(s_1, \mathbf{a})} \tag{2}$$
$$= \sum_{\mathbf{u}' \wedge u_1} \frac{\Pr(h(x), s_1, \mathbf{a}, \mathbf{u}', u_1, C = 1)}{\Pr(s_1, \mathbf{a}, \mathbf{u}', u_1, C = 1)} \cdot \frac{\Pr(\mathbf{u}', u_1, s_1, \mathbf{a})}{\Pr(s_1, \mathbf{a})} \tag{3}$$
$$\cdot \frac{\Pr(\mathbf{u}', s_1, \mathbf{a}, C = 1)}{\Pr(\mathbf{u}', s_1, \mathbf{a}, C = 1)} \tag{4}$$
$$= \sum_{\mathbf{u}' \wedge u_1} \frac{\Pr(h(x), u_1 \mid s_1, \mathbf{a}, \mathbf{u}', C = 1)}{\Pr(C = 1 \mid s_1, u_1, \mathbf{u}', \mathbf{a})} \cdot \Pr(C = 1, \mathbf{u}' \mid s_1, \mathbf{a}) \tag{5}$$
$$= \sum_{\mathbf{u}' \wedge u_1} \frac{\Pr(h(x), u_1 \mid s_1, \mathbf{a}, \mathbf{u}', C = 1)}{\Pr(C = 1 \mid u_1, \mathbf{u}')} \cdot \Pr(C = 1, \mathbf{u}' \mid s_1, \mathbf{a}) \tag{6}$$
$$\leq \frac{1}{\min_{\mathbf{u}' \wedge u_1} \Pr(C = 1 \mid u_1, \mathbf{u}')} \sum_{\mathbf{u}' \wedge u_1} \Pr(h(x), u_1 \mid s_1, \mathbf{a}, \mathbf{u}', C = 1) \tag{7}$$
$$\cdot \Pr(C = 1, \mathbf{u}' \mid s_1, \mathbf{a}) \tag{8}$$
$$= \frac{1}{\min_{\mathbf{u}' \wedge u_1} \Pr(C = 1 \mid u_1, \mathbf{u}')} \sum_{\mathbf{u}'} \Pr(C = 1, \mathbf{u}' \mid s_1, \mathbf{a}) \tag{9}$$

$$\sum_{u_1} \Pr(h(x), u_1 \mid s_1, \mathbf{a}, \mathbf{u'}, C=1) \tag{10}$$

$$= \frac{1}{\min_{\mathbf{u'} \wedge u_1} \Pr(C=1 \mid u_1, \mathbf{u'})} \sum_{\mathbf{u'}} \Pr(C=1, \mathbf{u'} \mid s_1, \mathbf{a}) \tag{11}$$

$$\cdot \Pr(h(x) \mid s_1, \mathbf{a}, \mathbf{u'}, C=1) \tag{12}$$

$$\leq \frac{\max_{\mathbf{u'}} \Pr(h(x) \mid s_1, \mathbf{a}, \mathbf{u'}, C=1)}{\min_{\mathbf{u'} \wedge u_1} \Pr(C=1 \mid u_1, \mathbf{u'})} \sum_{\mathbf{u'}} \Pr(C=1, \mathbf{u'} \mid s_1, \mathbf{a}) \tag{13}$$

$$= \frac{\max_{\mathbf{u'}} \Pr(h(x) \mid s_1, \mathbf{a}, \mathbf{u'}, C=1)}{\min_{\mathbf{u'} \wedge u_1} \Pr(C=1 \mid u_1, \mathbf{u'})} \cdot \Pr(C=1 \mid s_1, \mathbf{a}) \tag{14}$$

$$= \frac{\max_{\mathbf{u'}} \Pr_{\Delta}(h(x) \mid s_1, \mathbf{a}, \mathbf{u'})}{\min_{\mathbf{u'} \wedge u_1} \Pr(C=1 \mid u_1, \mathbf{u'})} \cdot \Pr(C=1 \mid s_1, \mathbf{a}) \tag{15}$$

$$\tag{16}$$

Case 3. $u_2 \in \mathbf{U}, u_1 \notin \mathbf{U}$

Similar reasoning to Case 2.

Case 4. $u_1, u_2 \in \mathbf{U}$

Note that $(\mathbf{X} \not\perp\!\!\!\perp C \,|\, S, \mathbf{A}, \mathbf{U'})$.

$$\sum_{\mathbf{u}} \Pr(h(x) \mid s_1, \mathbf{a}, \mathbf{u}, C=1) \cdot \Pr(\mathbf{u} \mid s_1, \mathbf{a}) = \sum_{\mathbf{u}} \Pr(h(x) \mid s_1, \mathbf{a}, \mathbf{u'} \wedge \{u_1, u_2\}, C=1) \cdot \Pr(\mathbf{u'} \wedge \{u_1, u_2\} \mid s_1, \mathbf{a}) \tag{1}$$

$$= \sum_{\mathbf{u}} \frac{\Pr(h(x), s_1, \mathbf{a}, \mathbf{u'}, u_1, u_2, C=1)}{\Pr(s_1, \mathbf{a}, \mathbf{u'}, u_1, u_2, C=1)} \cdot \frac{\Pr(\mathbf{u'}, u_1, u_2, s_1, \mathbf{a})}{\Pr(s_1, \mathbf{a})} \tag{2}$$

$$= \sum_{\mathbf{u}} \frac{\Pr(h(x), s_1, \mathbf{a}, \mathbf{u'}, u_1, u_2, C=1)}{\Pr(s_1, \mathbf{a}, \mathbf{u'}, u_1, u_2, C=1)} \cdot \frac{\Pr(\mathbf{u'}, u_1, u_2, s_1, \mathbf{a})}{\Pr(s_1, \mathbf{a})} \cdot \frac{\Pr(\mathbf{u'}, s_1, \mathbf{a}, C=1)}{\Pr(\mathbf{u'}, s_1, \mathbf{a}, C=1)} \tag{3}$$

$$= \sum_{\mathbf{u}} \frac{\Pr(h(x), u_1, u_2 \mid s_1, \mathbf{a}, \mathbf{u'}, C=1)}{\Pr(C=1 \mid s_1, u_1, u_2, \mathbf{u'}, \mathbf{a})} \cdot \Pr(C=1, \mathbf{u'} \mid s_1, \mathbf{a}) \tag{4}$$

$$= \sum_{\mathbf{u}} \frac{\Pr(h(x), u_1, u_2 \mid s_1, \mathbf{a}, \mathbf{u'}, C=1)}{\Pr(C=1 \mid u_1, u_2, \mathbf{u'})} \cdot \Pr(C=1, \mathbf{u'} \mid s_1, \mathbf{a}) \tag{5}$$

$$\leq \frac{1}{\min_{\mathbf{u}} \Pr(C=1 \mid u_1, u_2, \mathbf{u'})} \sum_{\mathbf{u}} \Pr(h(x), u_1, u_2 \mid s_1, \mathbf{a}, \mathbf{u'}, C=1) \tag{6}$$

$$\cdot \Pr(C=1, \mathbf{u'} \mid s_1, \mathbf{a}) \tag{7}$$

$$= \frac{1}{\min_{\mathbf{u}} \Pr(C=1 \mid u_1, u_2, \mathbf{u'})} \sum_{\mathbf{u'}} \Pr(C=1, \mathbf{u'} \mid s_1, \mathbf{a}) \tag{8}$$

$$\sum_{u_1, u_2} \Pr(h(x), u_1, u_2 \mid s_1, \mathbf{a}, \mathbf{u'}, C=1) \tag{9}$$

$$= \frac{1}{\min_{\mathbf{u}} \Pr(C=1 \mid \mathbf{u})} \sum_{\mathbf{u'}} \Pr(C=1, \mathbf{u'} \mid s_1, \mathbf{a}) \tag{10}$$

$$\cdot \Pr(h(x) \mid s_1, \mathbf{a}, \mathbf{u'}, C=1) \tag{11}$$

$$\leq \frac{\max_{\mathbf{u'}} \Pr(h(x) \mid s_1, \mathbf{a}, \mathbf{u'}, C=1)}{\min_{\mathbf{u}} \Pr(C=1 \mid \mathbf{u})} \sum_{\mathbf{u'}} \Pr(C=1, \mathbf{u'} \mid s_1, \mathbf{a}) \tag{12}$$

$$= \frac{\max_{\mathbf{u}'} \Pr(h(x) \mid s_1, \mathbf{a}, \mathbf{u}', C = 1)}{\min_{\mathbf{u}} \Pr(C = 1 \mid \mathbf{u})} \cdot \Pr(C = 1 \mid s_1, \mathbf{a}) \tag{13}$$

$$= \frac{\max_{\mathbf{u}'} \Pr_{\Delta}(h(x) \mid s_1, \mathbf{a}, \mathbf{u}')}{\min_{\mathbf{u}} \Pr(C = 1 \mid \mathbf{u})} \cdot \Pr(C = 1 \mid s_1, \mathbf{a}) \tag{14}$$

$$\tag{15}$$

$$\begin{aligned}
\Pr_{\Omega}(h(x) \mid s_1, \mathbf{a}) \le \max \Bigg\{ & \sum_{\mathbf{u}'} \Pr(h(x) \mid s_1, \mathbf{a}, \mathbf{u}', C = 1) \cdot \Pr(\mathbf{u}' \mid s_1, \mathbf{a}), \\
& \sum_{\mathbf{u}' \wedge u_1} \Pr(h(x) \mid s_1, \mathbf{a}, \mathbf{u}' \wedge u_1, C = 1) \cdot \Pr(\mathbf{u}' \wedge u_1 \mid s_1, \mathbf{a}), \\
& \sum_{\mathbf{u}' \wedge u_2} \Pr(h(x) \mid s_1, \mathbf{a}, \mathbf{u}' \wedge u_2, C = 1) \cdot \Pr(\mathbf{u}' \wedge u_2 \mid s_1, \mathbf{a}), \\
& \sum_{\mathbf{u}} \Pr(h(x) \mid s_1, \mathbf{a}, \mathbf{u}, C = 1) \cdot \Pr(\mathbf{u} \mid s_1, \mathbf{a}) \Bigg\} \\
= \max \Bigg\{ & \max_{\mathbf{u}'} \Pr_{\Delta}(h(x) \mid s_1, \mathbf{a}, \mathbf{u}'), \\
& \frac{\max_{\mathbf{u}'} \Pr_{\Delta}(h(x) \mid s_1, \mathbf{a}, \mathbf{u}')}{\min_{\mathbf{u}' \wedge u_1} \Pr(C = 1 \mid u_1, \mathbf{u}')} \cdot \Pr(C = 1 \mid s_1, \mathbf{a}), \\
& \frac{\max_{\mathbf{u}'} \Pr_{\Delta}(h(x) \mid s_1, \mathbf{a}, \mathbf{u}')}{\min_{\mathbf{u}' \wedge u_2} \Pr(C = 1 \mid u_2, \mathbf{u}')} \cdot \Pr(C = 1 \mid s_1, \mathbf{a}), \\
& \frac{\max_{\mathbf{u}'} \Pr_{\Delta}(h(x) \mid s_1, \mathbf{a}, \mathbf{u}')}{\min_{\mathbf{u}} \Pr(C = 1 \mid \mathbf{u})} \cdot \Pr(C = 1 \mid s_1, \mathbf{a}) \Bigg\} \\
= \max \Bigg\{ & 1, \underbrace{\frac{\Pr(C = 1 \mid s_1, \mathbf{a})}{\min_{\mathbf{u} \wedge u_1} \Pr(C = 1 \mid u_1, \mathbf{u}')}}_{\text{empirical}}, \frac{\Pr(C = 1 \mid s_1, \mathbf{a})}{\min_{\mathbf{u}' \wedge u_2} \Pr(C = 1 \mid u_2, \mathbf{u}')}, \frac{\Pr(C = 1 \mid s_1, \mathbf{a})}{\min_{\mathbf{u}} \Pr(C = 1 \mid \mathbf{u})} \Bigg\} \\
& \cdot \max_{\mathbf{u}'} \Pr_{\Delta}(h(x) \mid s_1, \mathbf{a}, \mathbf{u}')
\end{aligned}$$

2.1.4 Computability of Our Bounds

The two types of noise we are dealing with here are (1) inherent in determining what the selection variable C is, and (2), determining what the parents of C are.

To quantify the overall complexity of this bound, we can separate the calculation into two parts:

1. We must choose a set of ℓ uncertain edges, meaning that, which takes $\mathcal{O}(2^n)$ time if every edge is considered. As ℓ increases, the number of terms in the max grows exponentially as 2^ℓ.

2. For each set of edges that constitute $\mathbf{U}$ in the causal graph, we must evaluate α in terms of $\mathrm{Dom}(Pa(C))$.

Thus, the total complexity would be on the order of $\mathcal{O}\left(2^n \times \mathrm{Dom}(Pa(C))\right.$. However, in real-world applications, since we are usually certain with high probability that only a certain number of k features are non-latent variables in the causal graph, this complexity effectively reduces to the following:

$$\mathcal{O}\left(\binom{n}{k} \times \mathrm{Dom}(Pa(C))\right)$$

Note that the complexity term of $\binom{n}{k}$ can be greedily approximated using a decision tree searching algorithm, which reduces the complexity of this term to $n \times k$.

In the case of binary values for nodes in the causal graph, iterating over all possible assignments to each of the k variables would generate $\mathcal{O}(2^k)$ complexity. To further reduce the complexity of this bound, we could greedily choose only important nodes from $\mathbf{U} \setminus \mathbf{U}'$ that we believe contribute most to our bound, and leave out others. This would reduce our complexity by a factor of 2 for each one left out of consideration. However, this leads to a more lossy bound.

3 Extending LEWIS Explanation Scores

In this section, we calculate necessity and sufficiency scores exactly in the presence of binary attribute values and selection bias. We first study certain assumptions about the causal setting that allow us to calculate necessity and sufficiency scores **exactly**, even in the presence of selection bias. Then, we move to a more general setting, where we find upper bounds on these scores.

3.1 Calculating Exact Scores

We can calculate necessity and sufficiency exactly in certain scenarios. We will analyze the setting in which the following conditional independence assumption holds (motivated by Galhotra and Halpern):

$$(\mathbf{O}, \mathbf{X} \perp\!\!\!\perp C \mid \mathbf{K})$$

We claim that the result holds in either of the following cases:

1. There is no parent of $\mathbf{O}$ or $\mathbf{X}$ other than $\mathbf{K}$, and C is not a child of $\mathbf{O}$ or $\mathbf{X}$.
2. There is no parent of C other than $\mathbf{K}$.

Note that we will abbreviate $\Pr_\Omega(\cdot)$ as $\Pr(\cdot)$ throughout this section. Also, we take on binary attribute values in this setting. Given these assumptions, the following interventional independence condition holds:

$$\Pr(o'_{\mathbf{X}\leftarrow\mathbf{x}}, o_{\mathbf{X}\leftarrow\mathbf{x}'} \mid \mathbf{x}, \mathbf{k}) = \Pr(o'_{\mathbf{X}\leftarrow\mathbf{x}} \mid \mathbf{x}, \mathbf{k}) \cdot \Pr(o_{\mathbf{X}\leftarrow\mathbf{x}'} \mid \mathbf{x}, \mathbf{k}) \tag{1}$$

Using this condition, the exact scores can be calculated as:

$$\textsc{Nec}_{\mathbf{x}}(\mathbf{k}) = \frac{\big(1 - \Pr_\Delta(o', \mathbf{x} \mid \mathbf{k})\big) \cdot \Pr_\Delta(o' \mid \mathtt{do}(\mathbf{x}'), \mathbf{k}) - \Pr_\Delta(o' \mid \mathbf{k}) + \Pr_\Delta(o' \mid \mathbf{x}, \mathbf{k}) \cdot \big(1 - \Pr_\Delta(o, \mathbf{x}' \mid \mathbf{k})\big)}{\Pr_\Delta(o, \mathbf{x} \mid \mathbf{k})} \tag{2}$$

$$\textsc{Suf}_{\mathbf{x}}(\mathbf{k}) = \frac{\big(\Pr_\Delta(o \mid \mathbf{x}', \mathbf{k}) - 1\big) \cdot \Pr_\Delta(o' \mid \mathtt{do}(\mathbf{x}), \mathbf{k}) + \Pr_\Delta(o' \mid \mathbf{k}) - \Pr_\Delta(o \mid \mathbf{x}', \mathbf{k}) \cdot \Pr_\Delta(o', \mathbf{x} \mid \mathbf{k})}{\Pr_\Delta(o', \mathbf{x}' \mid \mathbf{k})} \tag{3}$$

$$\textsc{NeSuf}_{\mathbf{x}}(\mathbf{k}) = \Pr_\Delta(o, \mathbf{x} \mid \mathbf{k}) \cdot \textsc{Nec}_{\mathbf{x}}(\mathbf{k}) + \Pr_\Delta(o', \mathbf{x}' \mid \mathbf{k}) \cdot \textsc{Suf}_{\mathbf{x}}(\mathbf{k}) \tag{4}$$

Proof. We prove the bounds for (2); (3) and (4) are proved similarly. The following equations are obtained from the law of total probability:

$$\Pr(o'_{\mathbf{X}\leftarrow\mathbf{x}}, \mathbf{x} \mid \mathbf{k}) = \Pr(o'_{\mathbf{X}\leftarrow\mathbf{x}}, o'_{\mathbf{X}\leftarrow\mathbf{x}'}, \mathbf{x} \mid \mathbf{k}) + \Pr(o'_{\mathbf{X}\leftarrow\mathbf{x}}, o_{\mathbf{X}\leftarrow\mathbf{x}'}, \mathbf{x} \mid \mathbf{k}) \tag{5}$$

$$\Pr(o'_{\mathbf{X}\leftarrow\mathbf{x}'}, \mathbf{x} \mid \mathbf{k}) = \Pr(o'_{\mathbf{X}\leftarrow\mathbf{x}'}, o'_{\mathbf{X}\leftarrow\mathbf{x}}, \mathbf{x} \mid \mathbf{k}) + \Pr(o'_{\mathbf{X}\leftarrow\mathbf{x}'}, o_{\mathbf{X}\leftarrow\mathbf{x}}, \mathbf{x} \mid \mathbf{k}) \tag{6}$$

$$\Pr(o'_{\mathbf{X}\leftarrow\mathbf{x}'} \mid \mathbf{k}) = \Pr(o'_{\mathbf{X}\leftarrow\mathbf{x}'}, \mathbf{x} \mid \mathbf{k}) + \Pr(o'_{\mathbf{X}\leftarrow\mathbf{x}'}, \mathbf{x}' \mid \mathbf{k}) \tag{7}$$

By rearranging (5) and (6), we obtain the following equality:

$$\Pr(o'_{\mathbf{X}\leftarrow\mathbf{x}'}, o_{\mathbf{X}\leftarrow\mathbf{x}}, \mathbf{x} \mid \mathbf{k}) = \Pr(o'_{\mathbf{X}\leftarrow\mathbf{x}'}, \mathbf{x} \mid \mathbf{k}) - \Pr(o'_{\mathbf{X}\leftarrow\mathbf{x}}, \mathbf{x} \mid \mathbf{k}) + \Pr(o'_{\mathbf{X}\leftarrow\mathbf{x}}, o_{\mathbf{X}\leftarrow\mathbf{x}'}, \mathbf{x} \mid \mathbf{k}) \tag{8}$$

The following bounds for the LHS of (8) are obtained using the interventional independence assumption in (1):

$$\begin{aligned}
\text{LHS} &= \Pr(o'_{\mathbf{X}\leftarrow\mathbf{x}'}, \mathbf{x} \mid \mathbf{k}) - \Pr(o'_{\mathbf{X}\leftarrow\mathbf{x}}, \mathbf{x} \mid \mathbf{k}) + \Pr(o'_{\mathbf{X}\leftarrow\mathbf{x}}, o_{\mathbf{X}\leftarrow\mathbf{x}'}, \mathbf{x} \mid \mathbf{k}) \cdot \frac{\Pr(\mathbf{x} \mid \mathbf{k})}{\Pr(\mathbf{x} \mid \mathbf{k})} \\
&= \Pr(o'_{\mathbf{X}\leftarrow\mathbf{x}'}, \mathbf{x} \mid \mathbf{k}) - \Pr(o'_{\mathbf{X}\leftarrow\mathbf{x}}, \mathbf{x} \mid \mathbf{k}) + \Pr(o'_{\mathbf{X}\leftarrow\mathbf{x}}, o_{\mathbf{X}\leftarrow\mathbf{x}'} \mid \mathbf{x}, \mathbf{k}) \cdot \Pr(\mathbf{x} \mid \mathbf{k})
\end{aligned}$$

$$
\begin{aligned}
&= \Pr(o'_{\mathbf{X}\leftarrow\mathbf{x}'}, \mathbf{x} \mid \mathbf{k}) - \Pr(o'_{\mathbf{X}\leftarrow\mathbf{x}}, \mathbf{x} \mid \mathbf{k}) + \Pr(o'_{\mathbf{X}\leftarrow\mathbf{x}} \mid \mathbf{x}, \mathbf{k}) \cdot \Pr(o_{\mathbf{X}\leftarrow\mathbf{x}'} \mid \mathbf{x}, \mathbf{k}) \cdot \Pr(\mathbf{x} \mid \mathbf{k}) \\
&= \Big(\Pr(o'_{\mathbf{X}\leftarrow\mathbf{x}'} \mid \mathbf{k}) - \Pr(o'_{\mathbf{X}\leftarrow\mathbf{x}'}, \mathbf{x}' \mid \mathbf{k})\Big) - \Pr(o'_{\mathbf{X}\leftarrow\mathbf{x}}, \mathbf{x} \mid \mathbf{k}) + \Pr(o'_{\mathbf{X}\leftarrow\mathbf{x}} \mid \mathbf{x}, \mathbf{k}) \cdot \Pr(o_{\mathbf{X}\leftarrow\mathbf{x}'}, \mathbf{x} \mid \mathbf{k}) \\
&\qquad \texttt{(obtained from Eq.}(7)\texttt{)} \\
&= \Pr(o'_{\mathbf{X}\leftarrow\mathbf{x}'} \mid \mathbf{k}) - \Pr(o'_{\mathbf{X}\leftarrow\mathbf{x}'}, \mathbf{x}' \mid \mathbf{k}) - \Pr(o'_{\mathbf{X}\leftarrow\mathbf{x}}, \mathbf{x} \mid \mathbf{k}) \\
&\quad + \Pr(o'_{\mathbf{X}\leftarrow\mathbf{x}} \mid \mathbf{x}, \mathbf{k}) \cdot \Big(\Pr(o_{\mathbf{X}\leftarrow\mathbf{x}'} \mid \mathbf{k}) - \Pr(o_{\mathbf{X}\leftarrow\mathbf{x}'}, \mathbf{x}' \mid \mathbf{k})\Big) \\
&= \Pr(o'_{\mathbf{X}\leftarrow\mathbf{x}'} \mid \mathbf{k}) - \Pr(o', \mathbf{x}' \mid \mathbf{k}) - \Pr(o', \mathbf{x} \mid \mathbf{k}) + \Pr(o' \mid \mathbf{x}, \mathbf{k}) \cdot \Big(\Pr(o_{\mathbf{X}\leftarrow\mathbf{x}'} \mid \mathbf{k}) - \Pr(o, \mathbf{x}' \mid \mathbf{k})\Big) \\
&\qquad \texttt{(obtained from the consistency rule)} \\
&= \Pr(o'_{\mathbf{X}\leftarrow\mathbf{x}'} \mid \mathbf{k}) - \Pr(o' \mid \mathbf{k}) + \Pr(o' \mid \mathbf{x}, \mathbf{k}) \cdot \big(\Pr(o_{\mathbf{X}\leftarrow\mathbf{x}'} \mid \mathbf{k}) - \Pr(o, \mathbf{x}' \mid \mathbf{k})\big) \\
&= \Pr(o'_{\mathbf{X}\leftarrow\mathbf{x}'} \mid \mathbf{k}) - \Pr(o' \mid \mathbf{k}) + \Pr(o' \mid \mathbf{x}, \mathbf{k}) \cdot \big(1 - \Pr(o'_{\mathbf{X}\leftarrow\mathbf{x}'} \mid \mathbf{k}) - \Pr(o, \mathbf{x}' \mid \mathbf{k})\big) \\
&= \big(1 - \Pr(o', \mathbf{x} \mid \mathbf{k})\big) \cdot \Pr(o'_{\mathbf{X}\leftarrow\mathbf{x}'} \mid \mathbf{k}) - \Pr(o' \mid \mathbf{k}) + \Pr(o' \mid \mathbf{x}, \mathbf{k}) \cdot \big(1 - \Pr(o, \mathbf{x}' \mid \mathbf{k})\big) \\
&= \big(1 - \Pr(o', \mathbf{x} \mid \mathbf{k})\big) \cdot \Pr(o' \mid \mathtt{do}(\mathbf{x}'), \mathbf{k}) - \Pr(o' \mid \mathbf{k}) + \Pr(o' \mid \mathbf{x}, \mathbf{k}) \cdot \big(1 - \Pr(o, \mathbf{x}' \mid \mathbf{k})\big) \\
&\qquad \texttt{(obtained from the backdoor criterion)} \\
&= \big(1 - \Pr(o', \mathbf{x} \mid \mathbf{k}, C=1)\big) \cdot \Pr(o' \mid \mathtt{do}(\mathbf{x}'), \mathbf{k}, C=1) - \Pr(o' \mid \mathbf{k}, C=1) \\
&\quad + \Pr(o' \mid \mathbf{x}, \mathbf{k}, C=1) \cdot \big(1 - \Pr(o, \mathbf{x}' \mid \mathbf{k}, C=1)\big) \\
&\qquad \texttt{(obtained from the conditional independence assumption)} \\
&= \big(1 - \Pr_\Delta(o', \mathbf{x} \mid \mathbf{k})\big) \cdot \Pr_\Delta(o' \mid \mathtt{do}(\mathbf{x}'), \mathbf{k}) - \Pr_\Delta(o' \mid \mathbf{k}) \\
&\quad + \Pr_\Delta(o' \mid \mathbf{x}, \mathbf{k}) \cdot \big(1 - \Pr_\Delta(o, \mathbf{x}' \mid \mathbf{k})\big)
\end{aligned}
$$

Dividing through by $\Pr(o, \mathbf{x} \mid \mathbf{k})$, we obtain the final result. $\square$

3.2 General case

In the general case, if we assume that $\mathrm{Pa}(\mathbf{O})$, $\mathbf{O}$, and $\mathbf{X}$ are all disjoint from $\mathrm{Pa}(C)$, then we can calculate bounds for the scores by borrowing from the CRAB analysis.

3.2.1 No External Information

We first focus on the setting under no external information. Note that we are not restricted to the binary attribute setting anymore. There are two cases of interest explored below. We take the max over the two cases to write the final bound.

For the first case, we have the following bound:

$$
\begin{aligned}
\text{Nec. Case 1} \leq \Bigg(& \Pr_\Delta(o' \mid \mathtt{do}(\mathbf{x}'), \mathrm{Pa}(\mathbf{O}), \mathbf{k}) \\
& - \max_{\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_4} \Big\{ \Pr_\Delta(o', \mathbf{x}' \mid \mathrm{Pa}(\mathbf{O}), \mathbf{k}, \mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_4) \Big\} \Bigg) \\
& \cdot \Pr_\Delta(o \mid \mathbf{x}, \mathrm{Pa}(\mathbf{O}), \mathbf{k}) \cdot \max_{\mathbf{u}} \Big\{ \Pr_\Delta(\mathrm{Pa}(\mathbf{O}) \mid \mathbf{k}, \mathbf{u}) \Big\}
\end{aligned}
$$

and similarly for the second case:

$$
\begin{aligned}
\text{Nec. Case 2} \leq \Bigg(& \max_{\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_4} \{ \Pr_\Delta(o' \mid \mathtt{do}(\mathbf{x}'), \mathrm{Pa}(\mathbf{O}), \mathbf{k}, \mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_4) \} \\
& - \max_{\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_4} \Big\{ \Pr_\Delta(o', \mathbf{x}' \mid \mathrm{Pa}(\mathbf{O}), \mathbf{k}, \mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_4) \Big\} \Bigg)
\end{aligned}
$$

$$\cdot \max_{\mathbf{u}_1,\mathbf{u}_2,\mathbf{u}_4}\{\Pr_\Delta(o \mid \mathbf{x}, \mathrm{Pa}(\mathbf{O}), \mathbf{k}, \mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_4)\} \cdot \max_{\mathbf{u}}\left\{\Pr_\Delta(\mathrm{Pa}(\mathbf{O}) \mid \mathbf{k}, \mathbf{u})\right\}$$

We can now represent the bounds for the scores succinctly as follows:

$$\textsc{Nec}_{\mathbf{x}}(\mathbf{k}) \leq \frac{1}{\Pr_\Delta(o, \mathbf{x} \mid \mathbf{k})} \cdot \max\{\text{Nec. Case 1}, \text{Nec. Case 2}\} \tag{9}$$

$$\textsc{Suf}_{\mathbf{x}}(\mathbf{k}) \leq tbd \tag{10}$$

$$\textsc{NeSuf}_{\mathbf{x}}(\mathbf{k}) \leq tbd \tag{11}$$

Proof. We prove the bounds for (9); (10) and (11) are proved similarly.

$$\begin{aligned}
\text{LHS} &= \sum_{\mathrm{Pa}(\mathbf{O})\setminus\{\mathbf{x}\}} \Pr(o'_{\mathbf{X}\leftarrow\mathbf{x}'}, o_{\mathbf{X}\leftarrow\mathbf{x}} \mid \mathbf{x}, \mathrm{Pa}(\mathbf{O}), \mathbf{k}) \cdot \Pr(\mathrm{Pa}(\mathbf{O}), \mathbf{x} \mid \mathbf{k}) \\
&= \sum_{\mathrm{Pa}(\mathbf{O})\setminus\{\mathbf{x}\}} \Pr(o'_{\mathbf{X}\leftarrow\mathbf{x}'} \mid \mathbf{x}, \mathrm{Pa}(\mathbf{O}), \mathbf{k}) \cdot \Pr(o_{\mathbf{X}\leftarrow\mathbf{x}} \mid \mathbf{x}, \mathrm{Pa}(\mathbf{O}), \mathbf{k}) \cdot \Pr(\mathrm{Pa}(\mathbf{O}), \mathbf{x} \mid \mathbf{k}) \\
&= \sum_{\mathrm{Pa}(\mathbf{O})\setminus\{\mathbf{x}\}} \Pr(o'_{\mathbf{X}\leftarrow\mathbf{x}'} \mid \mathbf{x}, \mathrm{Pa}(\mathbf{O}), \mathbf{k}) \cdot \Pr(o_{\mathbf{X}\leftarrow\mathbf{x}} \mid \mathbf{x}, \mathrm{Pa}(\mathbf{O}), \mathbf{k}) \cdot \Pr(\mathrm{Pa}(\mathbf{O}), \mathbf{x} \mid \mathbf{k}) \\
&\quad \texttt{(obtained from the consistency rule)} \\
&= \sum_{\mathrm{Pa}(\mathbf{O})\setminus\{\mathbf{x}\}} \frac{\Pr(o'_{\mathbf{X}\leftarrow\mathbf{x}'}, \mathbf{x}, \mathrm{Pa}(\mathbf{O}), \mathbf{k})}{\Pr(\mathbf{x}, \mathrm{Pa}(\mathbf{O}), \mathbf{k})} \cdot \Pr(o_{\mathbf{X}\leftarrow\mathbf{x}} \mid \mathbf{x}, \mathrm{Pa}(\mathbf{O}), \mathbf{k}) \cdot \Pr(\mathrm{Pa}(\mathbf{O}), \mathbf{x} \mid \mathbf{k}) \\
&= \sum_{\mathrm{Pa}(\mathbf{O})\setminus\{\mathbf{x}\}} \frac{\Pr(o'_{\mathbf{X}\leftarrow\mathbf{x}'}, \mathbf{x} \mid \mathrm{Pa}(\mathbf{O}), \mathbf{k})}{\Pr(\mathbf{x} \mid \mathrm{Pa}(\mathbf{O}), \mathbf{k})} \cdot \Pr(o_{\mathbf{X}\leftarrow\mathbf{x}} \mid \mathbf{x}, \mathrm{Pa}(\mathbf{O}), \mathbf{k}) \cdot \Pr(\mathrm{Pa}(\mathbf{O}), \mathbf{x} \mid \mathbf{k}) \\
&= \sum_{\mathrm{Pa}(\mathbf{O})\setminus\{\mathbf{x}\}} \frac{\Pr(o'_{\mathbf{X}\leftarrow\mathbf{x}'} \mid \mathrm{Pa}(\mathbf{O}), \mathbf{k}) - \Pr(o'_{\mathbf{X}\leftarrow\mathbf{x}'}, \mathbf{x}' \mid \mathrm{Pa}(\mathbf{O}), \mathbf{k})}{\Pr(\mathbf{x} \mid \mathrm{Pa}(\mathbf{O}), \mathbf{k})} \cdot \Pr(o_{\mathbf{X}\leftarrow\mathbf{x}} \mid \mathbf{x}, \mathrm{Pa}(\mathbf{O}), \mathbf{k}) \cdot \Pr(\mathrm{Pa}(\mathbf{O}), \mathbf{x} \mid \mathbf{k}) \\
&= \sum_{\mathrm{Pa}(\mathbf{O})\setminus\{\mathbf{x}\}} \frac{\Pr(o'_{\mathbf{X}\leftarrow\mathbf{x}'} \mid \mathrm{Pa}(\mathbf{O}), \mathbf{k}) - \Pr(o', \mathbf{x}' \mid \mathrm{Pa}(\mathbf{O}), \mathbf{k})}{\Pr(\mathbf{x} \mid \mathrm{Pa}(\mathbf{O}), \mathbf{k})} \cdot \Pr(o \mid \mathbf{x}, \mathrm{Pa}(\mathbf{O}), \mathbf{k}) \cdot \Pr(\mathrm{Pa}(\mathbf{O}), \mathbf{x} \mid \mathbf{k}) \\
&\quad \texttt{(obtained from the consistency rule)} \\
&= \sum_{\mathrm{Pa}(\mathbf{O})\setminus\{\mathbf{x}\}} \frac{\Pr(o' \mid \mathtt{do}(x'), \mathrm{Pa}(\mathbf{O}), \mathbf{k}) - \Pr(o', \mathbf{x}' \mid \mathrm{Pa}(\mathbf{O}), \mathbf{k})}{\Pr(\mathbf{x} \mid \mathrm{Pa}(\mathbf{O}), \mathbf{k})} \cdot \Pr(o \mid \mathbf{x}, \mathrm{Pa}(\mathbf{O}), \mathbf{k}) \cdot \Pr(\mathrm{Pa}(\mathbf{O}), \mathbf{x} \mid \mathbf{k}) \\
&= \sum_{\mathrm{Pa}(\mathbf{O})\setminus\{\mathbf{x}\}} \big(\Pr(o' \mid \mathtt{do}(x'), \mathrm{Pa}(\mathbf{O}), \mathbf{k}) - \Pr(o', \mathbf{x}' \mid \mathrm{Pa}(\mathbf{O}), \mathbf{k})\big) \cdot \Pr(o \mid \mathbf{x}, \mathrm{Pa}(\mathbf{O}), \mathbf{k}) \cdot \Pr(\mathrm{Pa}(\mathbf{O}) \mid \mathbf{k})
\end{aligned}$$

Now we apply the CRAB analysis in the case of no external information. Let $\mathbf{U} \subseteq \mathbf{X} \times \mathbf{O}$ where $\mathbf{U} \cap (\mathrm{Pa}(\mathbf{O}) \cup K) = \emptyset$, such that the following conditional independence is satisfied:

$$(\mathrm{Pa}(\mathbf{O}) \perp\!\!\!\perp C \mid \mathbf{U}, \mathbf{K})$$

Here, we are also implicitly assuming that $\mathrm{Pa}(C) \cap \mathrm{Pa}(\mathbf{O}) = \emptyset$, because if variables in the parents of $\mathbf{O}$ directly influenced the selection variable, then by definition, the conditional independence assumption would not hold.

$$\begin{aligned}
\Pr\left(\mathrm{Pa}(\mathbf{O}) \mid \mathbf{k}\right) &= \sum_{\mathbf{u}\in Dom(\mathbf{U})} \Pr(\mathrm{Pa}(\mathbf{O}) \mid \mathbf{k}, \mathbf{u}) \cdot \Pr(\mathbf{u} \mid \mathbf{k}) \\
&= \sum_{\mathbf{u}\in Dom(\mathbf{U})} \Pr(\mathrm{Pa}(\mathbf{O}) \mid \mathbf{k}, \mathbf{u}, C = 1) \cdot \Pr(\mathbf{u} \mid \mathbf{k}) \\
&\leq \sum_{\mathbf{u}\in Dom(\mathbf{U})} \max_{\mathbf{u}}\{\Pr(\mathrm{Pa}(\mathbf{O}) \mid \mathbf{k}, \mathbf{u}, C = 1)\} \cdot \Pr(\mathbf{u} \mid \mathbf{k})
\end{aligned}$$

$$= \max_{\mathbf{u} \in Dom(\mathbf{U})} \{\Pr(\mathrm{Pa}(\mathbf{O}) \mid \mathbf{k}, \mathbf{u}, C=1)\} \cdot \sum_{\mathbf{u} \in Dom(\mathbf{U})} \Pr(\mathbf{u} \mid \mathbf{k})$$
$$= \max_{\mathbf{u} \in Dom(\mathbf{U})} \{\Pr(\mathrm{Pa}(\mathbf{O}) \mid \mathbf{k}, \mathbf{u}, C=1)\}$$
$$= \max_{\mathbf{u} \in Dom(\mathbf{U})} \{\Pr_\Delta(\mathrm{Pa}(\mathbf{O}) \mid \mathbf{k}, \mathbf{u})\}$$

Thus, we have:

$$\text{LHS} \leq \sum_{\mathrm{Pa}(\mathbf{O}) \setminus \{\mathbf{x}\}} \Bigg(\underbrace{\Pr(o' \mid \mathtt{do}(\mathbf{x}'), \mathrm{Pa}(\mathbf{O}), \mathbf{k})}_{\text{Term I}} - \underbrace{\Pr(o', \mathbf{x}' \mid \mathrm{Pa}(\mathbf{O}), \mathbf{k})}_{\text{Term II}} \Bigg) \cdot \underbrace{\Pr(o \mid \mathbf{x}, \mathrm{Pa}(\mathbf{O}), \mathbf{k})}_{\text{Term III}} \cdot \max_{\mathbf{u} \in Dom(\mathbf{U})} \{\Pr_\Delta(\mathrm{Pa}(\mathbf{O}) \mid \mathbf{k}, \mathbf{u})\}$$

We still need to convert Terms I, II, and III, which are probabilities over the true population distribution, to the biased data distribution.

There are two cases of interest.

1. If $\mathbf{O}$ has no children, the conditional independence $\Pr(\mathbf{O} \perp\!\!\!\perp C \mid \mathrm{Pa}(\mathbf{O}))$ holds trivially. We can directly change Terms I and III to the biased data distribution, and we will have to use the CRAB bound on Term II.
2. If $\mathbf{O}$ has children, we will have to use the CRAB bound on all of the terms.

Since we are already conditioning on $\mathrm{Pa}(\mathbf{O})$ and $\mathbf{K}$, we do not have to condition on as many other variables in the support to achieve the conditional independence between $(\mathbf{O}, \mathbf{X})$ and the selection variable C.

Case 1.

For Terms I and III, we can directly change the probabilities to be over the biased data distribution. We will have that:

$$\text{Term I} \leq \Pr_\Delta(o' \mid \mathtt{do}(\mathbf{x}'), \mathrm{Pa}(\mathbf{O}), \mathbf{k})$$

and

$$\text{Term III} \leq \Pr_\Delta(o \mid \mathbf{x}, \mathrm{Pa}(\mathbf{O}), \mathbf{k})$$

For Term II, we apply the CRAB analysis as follows. Define the disjoint subsets $\mathbf{U}_1, \mathbf{U}_2, \mathbf{U}_3, \mathbf{U}_4 \subseteq \mathbf{U}$, such that all are disjoint from $\mathrm{Pa}(\mathbf{O})$ and $\mathbf{K}$. Consider the graphical construction below:

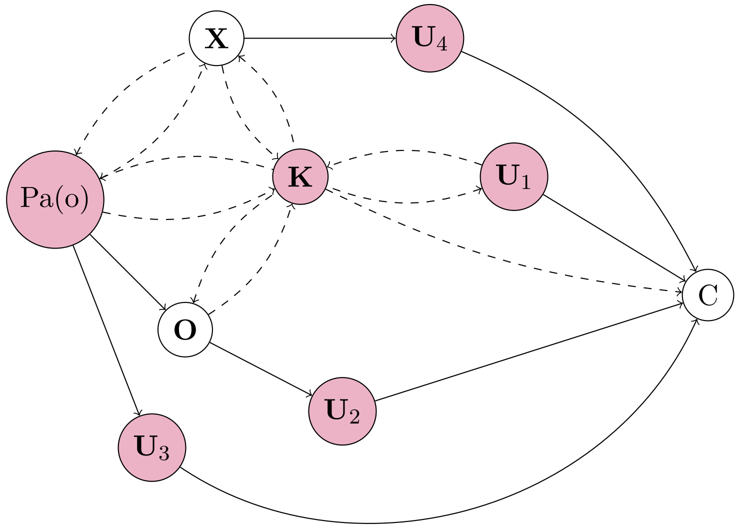


Figure 1: Bounding probabilities with only $\mathbf{O}$

In the above figure, shading of the nodes represents conditioning. Any combination of the dashed arrows, along with all of the solid arrows, constitute a valid graph.

From the above model, observe that we will have to condition on $\mathbf{U}_1$, $\mathbf{U}_2$, and $\mathbf{U}_4$, as well as $\mathrm{Pa}(\mathbf{O})$ and $\mathbf{K}$, in observe that in order to apply the CRAB bound as follows:

$$\text{Term II} \leq \max_{\mathbf{u}_1,\mathbf{u}_2,\mathbf{u}_4} \{\Pr_\Delta(o', \mathbf{x}' \mid \mathrm{Pa}(\mathbf{O}), \mathbf{k}, \mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_4)\}$$

Case 2.

As in the previous case, we can apply the CRAB analysis to get the following bounds:

$$\begin{aligned}
\text{Term I} &\leq \max_{\mathbf{u}_1,\mathbf{u}_2,\mathbf{u}_4} \{\Pr_\Delta(o' \mid \mathtt{do}(\mathbf{x}'), \mathrm{Pa}(\mathbf{O}), \mathbf{k}, \mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_4)\} \\
\text{Term II} &\leq \max_{\mathbf{u}_1,\mathbf{u}_2,\mathbf{u}_4} \{\Pr_\Delta(o', \mathbf{x}' \mid \mathrm{Pa}(\mathbf{O}), \mathbf{k}, \mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_4)\} \\
\text{Term III} &\leq \max_{\mathbf{u}_1,\mathbf{u}_2,\mathbf{u}_4} \{\Pr_\Delta(o \mid \mathbf{x}, \mathrm{Pa}(\mathbf{O}), \mathbf{k}, \mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_4)\}
\end{aligned}$$

However, in certain scenarios, it may be possible to further simplify Terms Term I and III. Consider the setting of Figure 1, as depicted below:

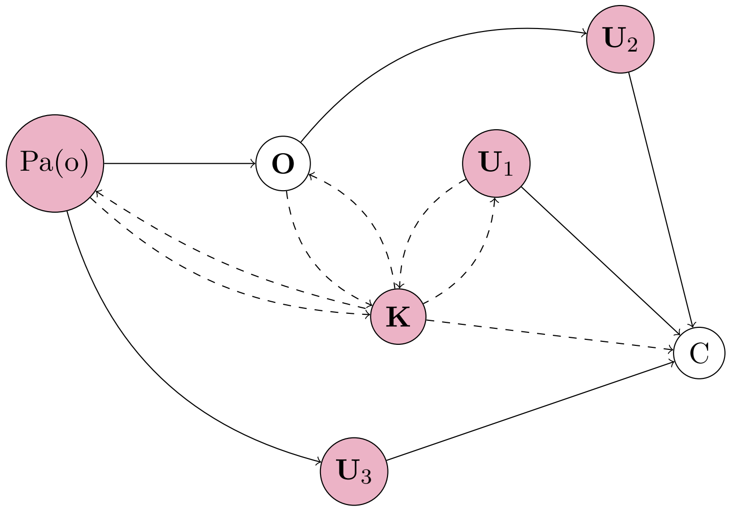


Figure 2: Bounding probabilities with only o

Note that Figure 1 is a generalization of Figure 2. In this scenario, we can form the conditional independence $(\mathbf{O} \perp\!\!\!\perp C \mid \mathbf{U}_1, \mathbf{U}_2, \mathrm{Pa(o)})$. By applying the CRAB analysis, we have:

$$\text{Term I} \leq \max_{\mathbf{u}_1,\mathbf{u}_2} \{ \Pr_\Delta(o' \mid \mathtt{do}(\mathbf{x}'), \mathrm{Pa}(\mathbf{O}), \mathbf{k}, \mathbf{u}_1, \mathbf{u}_2) \}$$

and

$$\text{Term III} \leq \max_{\mathbf{u}_1,\mathbf{u}_2} \{ \Pr_\Delta(o \mid \mathbf{x}, \mathrm{Pa}(\mathbf{O}), \mathbf{k}, \mathbf{u}_1, \mathbf{u}_2) \}$$

Note that since we already condition on Pa(o) and $\mathbf{K}$, we do not have to condition on $\mathbf{U}_3$.

$\square$