Appendix

Proof of Theorem 4

Theorem 4. Given a causal diagram $G$ and a distribution compatible with $G$, let $W \cup U$ be a set of variables satisfying the back-door criterion in $G$ relative to an ordered pair $(X, Y)$, where $W \cup U$ is partially observable, i.e., only probabilities $P(X, Y, W)$ and $P(U)$ are given, the causal effects of $X$ on $Y$ are then bounded as follows:

$$LB \leq P(y|do(x)) \leq UB$$

where $LB$ is the solution to the non-linear optimization problem in Equation 9 and $UB$ is the solution to the non-linear optimization problem in Equation 10.

$$LB = \min \sum_{w,u} \frac{a_{w,u}b_{w,u}}{c_{w,u}},$$

$$UB = \max \sum_{w,u} \frac{a_{w,u}b_{w,u}}{c_{w,u}},$$

where,

$$\sum_{u} a_{w,u} = P(x, y, w), \sum_{u} b_{w,u} = P(w),$$

$$\sum_{u} c_{w,u} = P(x, w) \text{ for all } w \in W;$$

and for all $w \in W$ and $u \in U$,

$$b_{w,u} \geq c_{w,u} \geq a_{w,u};$$

$$\max \{0, p(x, y, w) + p(u) - 1\} \leq a_{w,u},$$

$$\min \{P(x, y, w), p(u)\} \geq a_{w,u},$$

$$\max \{0, p(w) + p(u) - 1\} \leq b_{w,u},$$

$$\min \{P(w), p(u)\} \geq b_{w,u},$$

$$\max \{0, p(x, w) + p(u) - 1\} \leq c_{w,u},$$

$$\min \{P(x, w), p(u)\} \geq c_{w,u}.$$  

Proof. To show that the LB and UB bound the actual causal effects, we only need to show that there exists a point in feasible space of the non-linear optimization that $\frac{a_{w,u}b_{w,u}}{c_{w,u}}$ is equal to the actual causal effects. Since $W \cup U$ satisfies the back-door criterion, by adjustment formula in Equation 1, we have,

$$P(y|do(x)) = \sum_{w,u} P(y|x, w, u) P(w, u)$$

$$= \sum_{w,u} \frac{P(x, y, w, u)P(w, u)}{P(x, w, u)}$$

Let

$$a_{w,u} = P(x, y, w, u)$$

$$b_{w,u} = P(w, u)$$

$$c_{w,u} = P(x, w, u)$$

We now show that the above set of $a_{w,u}, b_{w,u}, c_{w,u}$ are in feasible space.

We have,

for $w \in W$,

$$\sum_{u} a_{w,u} = \sum_{u} P(x, y, w, u) = P(x, y, w),$$

$$\sum_{u} b_{w,u} = \sum_{u} P(w, u) = P(w),$$

$$\sum_{u} c_{w,u} = \sum_{u} P(x, w, u) = P(x);$$

and,

for all $w \in W$ and $u \in U$,

$$b_{w,u} = P(w, u) \geq P(x, w, u) = c_{w,u},$$

$$c_{w,u} = P(x, w, u) \leq a_{w,u},$$

$$a_{w,u} = P(x, y, w, u) \leq \min \{P(x, y, w), p(u)\},$$

$$b_{w,u} = P(w, u) \leq \min \{P(w, p(u))\},$$

$$c_{w,u} = P(x, w, u) \leq \min \{P(x, w), p(u)\},$$

$$a_{w,u} = P(x, y, w, u) \geq \max \{0, p(x, y, w) + p(u) - 1\},$$

$$b_{w,u} = P(w, u) \geq \max \{0, p(w) + p(u) - 1\},$$

$$c_{w,u} = P(x, w, u) \geq \max \{0, p(x, w) + p(u) - 1\}.$$  

Therefore, the above set of $a_{w,u}, b_{w,u}, c_{w,u}$ are in feasible space, and thus, the UB and LB bound the actual causal effects. \qed

Proof of Theorem 5

Theorem 5. Given a causal diagram $G$ and distribution compatible with $G$, let $W \cup U$ be a set of variables satisfying the front-door criterion in $G$ relative to an ordered pair $(X, Y)$, where $W \cup U$ is partially observable, i.e., only probabilities $P(X, Y, W)$ and $P(U)$ are given and $P(X, W, U) > 0$, the causal effects of $X$ on $Y$ are then bounded as follows:

$$LB \leq P(y|do(x)) \leq UB$$

where $LB$ is the solution to the non-linear optimization problem in Equation 11 and $UB$ is the solution to the non-linear optimization problem in Equation 12.

$$LB = \min \sum_{w,u} \frac{b_{x,u}a_{x,u}}{c_{x,u}b_{x,u}},$$

$$UB = \max \sum_{w,u} \frac{b_{x,u}a_{x,u}}{c_{x,u}b_{x,u}},$$

where,

$$\sum_{u} a_{x,u} = P(x, y, w), \sum_{u} b_{x,u} = P(x, w)$$

for all $x \in X$ and $w \in W;$$

and for all $x \in X, w \in W$, and $u \in U$,

$$b_{x,u} \geq a_{x,u};$$

$$\max \{0, p(x, y, w) + p(u) - 1\} \leq a_{x,u},$$

$$\min \{P(x, y, w), p(u)\} \geq a_{x,u},$$

$$\max \{0, p(x, w) + p(u) - 1\} \leq b_{x,u},$$

$$\min \{P(x, w), p(u)\} \geq b_{x,u}.$$
Theorem 7. Let $P$ be a causal diagram containing nodes $\{V_1, \ldots, V_{n-3}, X, Y, Z\}$. Let $O$ be any observational data compatible with $G$. Suppose there exists a set of variables that satisfies the back-door or front-door criterion relative to $(X, Y)$ in $G$, then, $(G, O)$ is equivalent to $(G', O')$ $(G' \supseteq G)$ containing nodes $\{V_1, \ldots, V_{n-3}, X, Y, W, U\}$; $O'$ is observational data compatible with $G'$, where the number of states in $W$ times the number of states in $U$ is equal to the number of states in $Z$, and the structure of $G'$ and the observational data $O'$ are obtained as follows:

Structure of $G'$:
Let $\text{Parents}_{G'}(H)$ be the parents of $H$ in causal diagram $G$. 
\[ \text{Parents}_{G'}(U) = \text{Parents}_{G'}(Z), \quad \text{Parents}_{G'}(W) = \text{Parents}_{G'}(Z) \cup \{U\}, \]
\[ \text{Parents}_{G'}(H) = \text{Parents}_{G'}(H) \text{ if } Z \notin \text{Parents}_{G'}(H) \text{ for } H \in \{V_1, \ldots, V_{n-3}, X, Y\}, \]
\[ \text{Parents}_{G'}(H) = \text{Parents}_{G'}(H) \setminus \{Z\} \cup \{W, U\} \text{ if } Z \in \text{Parents}_{G'}(H) \text{ for } H \in \{V_1, \ldots, V_{n-3}, X, Y\}. \]

Observe that, let $Q$ be the set of variables in $G$ that satisfies the back-door or front-door criterion relative to $(X, Y)$, then $Q'$ satisfies the back-door or front-door criterion relative to $(X, Y)$ in $G'$, where 
$Q' = Q$ if $Z \notin Q$, 
$Q' = Q \setminus \{Z\} \cup \{W, U\}$ if $Z \in Q$.

Observational data:
Let the number of states in $W$ be $p$, and let the number of states in $U$ be $q$.
The states of $Z$ is the Cartesian product of the states of $W$ and the states of $U$.

In detail, $(w_j, u_k)$ is equivalent to $z_{(j-1)s+q+k}$; $w_j$ is equivalent to $\cap_{k=1}^{q} (w_j, u_k)$; $u_k$ is equivalent to $\cup_{j=1}^{p} (w_j, u_k)$; $z_{(j-1)s+q+k}$, i.e., $P(w_j, u_k | V) = P(z_{(j-1)s+q+k} | V)$ for any $V \subseteq \{V_1, \ldots, V_{n-3}, X, Y\}$.

Proof. First, we show that $Q'$ satisfies the back-door or front-door criterion relative to $(X, Y)$ in $G'$.

If $Q$ satisfies the back-door criterion relative to $(X, Y)$ in $G$, we need to show that,

- no node in $Q'$ is a descendant of $X$.
- $Q'$ blocks every path between $X$ and $Y$ that contains an arrow into $X$.

It is easy to show that if there is a node in $Q'$ that is a descendant of $X$ in $G'$, then there is a node in $Q$ that is a descendant of $X$ in $G$. And if there is a path between $X$ and $Y$ that contains an arrow into $X$ does not blocked by $Q'$ in $G'$, then there is a path between $X$ and $Y$ that contains an arrow into $X$ does not blocked by $Q$ in $G$. Thus, $Q'$ satisfies the back-door criterion relative to $(X, Y)$ in $G'$. Similarly, we can show that if $Q$ satisfies the front-door criterion relative to $(X, Y)$ in $G$, then $Q'$ satisfies the front-door criterion relative to $(X, Y)$ in $G'$.

Now, we show that $(G, O)$ is equivalent to $(G', O')$, i.e., show that $P(y|do(x))$ is the same between $(G, O)$ and $(G', O')$. Suppose $Q$ satisfies the back-door criterion relative to $(X, Y)$ in $G$. By adjustment formula in Equation 1, we have,

$P(y|do(x)) = \sum_{q \in Q} P(y|x, q) \times P(q) = \sum_{q \in Q} \frac{P(y|x, q) \times P(q)}{P(x,q)}$.

And in $G'$,

$P(y|do(x)) = \sum_{q \in Q'} P(y|x, q) \times P(q) = \sum_{q \in Q'} \frac{P(y|x, q) \times P(q)}{P(x,q)}$.

It is obviously that these two causal effects are the same, because $P(w_j, u_k | V) = P(z_{(j-1)s+k} | V)$ for any $V \subseteq \{V_1, \ldots, V_{n-3}, X, Y\}$.

Similarly, we can show that if $Q$ satisfies the front-door criterion relative to $(X, Y)$ in $G$, $(G, O)$ is equivalent to $(G', O')$.

Simulation Algorithm for Generating Sample Distributions

The two sample distributions generated in the paper (in two Simulation Results sections) were generated by Algorithm 2 with $D$ equal to the uniform distribution.
Algorithm 2: Generate-cpt()

**Input:** $n$ causal diagram nodes $(X_1, ..., X_n)$; Distribution $D$.

**Output:** $n$ conditional probability tables for $P(X_i | \text{Parents}(X_i))$.

```
1: for $i = 1$ to $n$ do
2:   $s = \text{num\_instantiates}(X_i)$;
3:   $p = \text{num\_instantiates}(\text{Parents}(X_i))$
4:   for $k = 1$ to $p$ do
5:     $\text{sum} = 0$
6:     for $j = 1$ to $s$ do
7:       $a_j = \text{sample}(D)$
8:       $\text{sum} = \text{sum} + a_j$
9:     end for
10:    for $j = 1$ to $s$ do
11:      $P(x_i | \text{Parents}(X_i)_k) = a_j / \text{sum}$
12:   end for
13: end for
```

**Construction of the Data in Table 4**

- $P(u, w) = P(z_1)$,
- $P(u, w') = P(z_2)$,
- $P(u', w) = P(z_3)$,
- $P(u', w') = P(z_4)$,
- $P(u) = P(u, w) + P(u, w') = P(z_1) + P(z_2) = 0.5$,
- $P(w|u) = P(u, w)/P(u) = P(z_1)/P(u) = 0.3/0.5 = 0.6$,
- $P(w|u') = P(u', w)/P(u') = P(z_4)/(1 - P(u)) = 0.2/0.5 = 0.4$,
- $P(x|u, w) = P(x|z_1) = 0.1$,
- $P(x|u, w') = P(x|z_2) = 0.4$,
- $P(x|u', w) = P(x|z_3) = 0.5$,
- $P(x|u', w') = P(x|z_4) = 0.7$,
- $P(y|x, u, w) = P(y|x, z_1) = 0.2$,
- $P(y|x', u, w) = P(y|x', z_1) = 0.3$,
- $P(y|x, u, w') = P(y|x, z_2) = 0.7$,
- $P(y|x', u, w') = P(y|x', z_2) = 0.1$,
- $P(y|x, u', w) = P(y|x, z_3) = 0.6$,
- $P(y|x', u', w) = P(y|x', z_3) = 0.5$,
- $P(y|x, u', w') = P(y|x, z_4) = 0.5$,
- $P(y|x', u', w') = P(y|x', z_4) = 0.4$.

**Construction of the Distribution in the Example of Dimensionality Reduction**

Here is how the data used in the example of Dimensionality Reduction were generated (both $P(X, Y, Z)$ and $P(X, Y, W), P(U)$). Instead of providing the resulting 1024 rows of the observational data, we provide the details for regenerating the observational data as following steps.

- Let $P(X, Y, w_j, u_k) = P(X, Y, z_{(j-1)\times16+k})$.
- Let $P(X, Y, w_j) = \sum_{k=1}^{8} P(X, Y, w_j, u_k)$.
- Let $P(X, Y, u_k) = \sum_{j=1}^{p} P(X, Y, w_j, u_k)$.
- Let $P(u_k) = \sum_{x,y} P(X, Y, u_k)$.

For example,

$P(u_1)$

$= \sum_{X,Y} P(X, Y, u_1)$

$= P(x, y, u_1) + P(x, y', u_1) + P(x', y, u_1) + P(x', y', u_1)$

$= \sum_{j=1}^{16} P(x, y, w_j, u_1) + \sum_{j=1}^{16} P(x, y', w_j, u_1) + \sum_{j=1}^{16} P(x', y, w_j, u_1) + \sum_{j=1}^{16} P(x', y', w_j, u_1)$

$= \sum_{j=1}^{16} P(x, y, z_{(j-1)\times16+1}) + \sum_{j=1}^{16} P(x, y', z_{(j-1)\times16+1}) + \sum_{j=1}^{16} P(x', y, z_{(j-1)\times16+1}) + \sum_{j=1}^{16} P(x', y', z_{(j-1)\times16+1})$

$= \sum_{k=1}^{16} P(x, y, w_1, u_k)$

$= \sum_{k=1}^{16} P(x, y, z_k)$. 

- Generate $P(X, Y, Z)$ using Algorithm 2.