Causal and Interventional Markov Boundaries (Supplementary Material)

Sofia Triantafillou[1] **Fattaneh Jabbari**[1] **Gregory F. Cooper**[1]

[1]Department of Biomedical Informatics, University of Pittsburgh, Pittsburgh, Pennsylvania, USA

Table S1: Closed-form solutions for Eq. 4 -8 in the main paper, for multinomial distributions with Dirichlet priors. Subscript jk refers variable Y taking its k-th configuration, and variable set $\mathbf{Z}$ taking its j-th configuration. α_{jk} is the prior for the Dirichlet distribution. We set $\alpha_{jk} = 1$ in all experiments. N^o_{jk}, N^e_{jk} corresponds to counts in the data where $Y = k$ and $\mathbf{Z} = j$ in D_o and D_e, respectively. N^o_j, N^e_j corresponds to counts in the data where $Z = j$. Tilde notation corresponds to the OMB $\mathbf{U}$.

Eq. number	Analytical Expression
Eq. 5	$P(D_e\|D_o, H^c_{\mathbf{Z}}) = \prod_{j=1}^{q} \frac{\Gamma(\alpha_j + N^o_j)}{\Gamma(\alpha_j + N^o_j + N^e_j)} \prod_{k=1}^{r} \frac{\Gamma(\alpha_{jk} + N^o_{jk} + N^e_{jk})}{\Gamma(\alpha_{jk} + N^o_{jk})}$
-	$P(D_e\|D_o, H^{\bar{c}}_{\mathbf{Z}}) = \prod_{j=1}^{q} \frac{\Gamma(\alpha_j)}{\Gamma(\alpha_j + N^e_j)} \prod_{k=1}^{r} \frac{\Gamma(\alpha_{jk} + N^e_{jk})}{\Gamma(\alpha_{jk})}$
Eq. 7	$P(D_o\|H^c_{\mathbf{Z}}) = \prod_{j=1}^{\tilde{q}} \frac{\Gamma(\tilde{\alpha}_j)}{\Gamma(\tilde{\alpha}_j + \tilde{N}^o_j)} \prod_{k=1}^{r} \frac{\Gamma(\tilde{\alpha}_{jk} + \tilde{N}^o_{jk})}{\Gamma(\tilde{\alpha}_{jk})}$
Eq. 7	$P(D_o\|H^{\bar{c}}_{\mathbf{Z}}) = \prod_{j=1}^{\tilde{q}} \frac{\Gamma(\tilde{\alpha}_j)}{\Gamma(\tilde{\alpha}_j + \tilde{N}^o_j)} \prod_{k=1}^{r} \frac{\Gamma(\tilde{\alpha}_{jk} + \tilde{N}^o_{jk})}{\Gamma(\tilde{\alpha}_{jk})}$
Terms in Eq. 8	$P(Y = k\|x, Z = j, D_e, D_o, H^c_{\mathbf{Z}}) = \frac{N^o_{jk} + N^e_{jk} + \alpha_{jk}}{N^o_j + N^e_j + \alpha_j}$
Terms in Eq. 8	$P(Y = k\|x, Z = j, D_e, D_o, H^{\bar{c}}_{\mathbf{Z}}) = \frac{N^e_{jk} + \alpha_{jk}}{N^e_j + \alpha_j}$

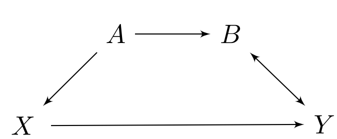

Figure S1: An example where a CMB does not necessarily correspond to an optimal adjustment set [Henckel et al., 2019]. $\mathbf{CMB}_X(Y) = \{\{X, A, B\}\}$, but the optimal adjustment set depends on the parameters.

Supplement for the Thirty-Seventh Conference on Uncertainty in Artificial Intelligence (UAI 2021).

1 PROOFS

In this section, we provide a proof that every causal Markov boundary is backdoor set, which is defined below (Definition 1.1). We make the following assumptions throughout the entire document:

- X causes Y
- all variables $\mathbf{V}$ are pre-treatment.

Definition 1.1 (Backdoor Set). $\mathbf{Z}$ is a backdoor set for X, Y if and only if $\mathbf{Z}$ m-separates X and Y in $\mathcal{G}_{\underline{X}}$.

We use the following definitions from [Shpitser and Pearl, 2006]:

Definition 1.2 (C-component). A C-component is as set of nodes S in $\mathcal{G}$ where every two nodes are connected by a bidirected path.

Definition 1.3 (C-forest). A graph $\mathcal{G}$ where the set of all of its nodes is a C-component, and each node has at most one child is a C-forest. The set of nodes $\mathbf{R}$ without children in the C-forest is called the root, and we say that $\mathcal{G}$ is an $\mathbf{R}$-rooted C-forest.

C-forests are useful for defining hedges:

Definition 1.4 (hedge). Let $\mathbf{X}$,$\mathbf{Y}$ be sets of variables in $\mathcal{G}$. Let F, F' be $\mathbf{R}$-rooted C-forests in $\mathcal{G}$ such that F' is a subgraph of F, $\mathbf{X}$ only occurs in F, and $\mathbf{R} \in An(\mathbf{Y})_{\mathcal{G}_{\overline{X}}}$. Then F, F' form a hedge for $P(\mathbf{Y}|do(\mathbf{X}))$.

The existence of a hedge for $P(\mathbf{Y}|do(\mathbf{X}))$ in $\mathcal{G}$ is equivalent to the non-identifiability of $P(\mathbf{Y}|do(\mathbf{X}))$ (see Theorem 4 in [Shpitser and Pearl, 2006]).

Lemma 1.5. *Let $\mathbf{Z}$ be a set that is not a subset of any backdoor set (i.e., there exists no set $\mathbf{Q} \subseteq (\mathbf{V} \setminus \mathbf{Z})$ such that $\mathbf{Q} \cup \mathbf{Z}$ m-separate X and Y in $\mathcal{G}_{\underline{X}}$). Then there exists in $\mathcal{G}$ a bi-directed path from X to Y where every collider has a descendant in $\mathbf{Z} \cup Y$.*

Proof. The proof is a special case of Theorem 4.2 $(iv) \Rightarrow (ii)$ in [Richardson et al., 2002] with $\mathbf{S} \leftarrow \mathbf{Z}, \mathbf{L} \leftarrow \emptyset, \mathcal{G} \leftarrow \mathcal{G}_{\underline{X}}$. The proof is for ancestral graphs, but it is straightforward to show that it holds for SMCMs, given that every SMCM $\mathcal{G}$ can be transformed to a maximal ancestral graph $\mathcal{M}$ over the same nodes (by adding some edges) such that (a) $\mathcal{G}$ and $\mathcal{M}$ entail the exact same m-separations and m-connections and (b) the exact same ancestral relationships hold in both graphs. The theorem proves that if $\forall \mathbf{Q} \subseteq (V \setminus \mathbf{Z}), \mathbf{Z} \cup \mathbf{Q}$ do not m-separate X and Y in $\mathcal{G}_{\underline{X}}$, then there exists a bidirected path between X and Y in $\mathcal{G}_{\underline{X}}$ where every variable is an ancestor of some variables in $\mathbf{Z} \cup \{X, Y\}$, which means that there exists a path in $\mathcal{G}$ a bi-directed path from X to Y where every collider has a descendant in $\mathbf{Z} \cup Y$ (since $X \rightarrow Y$ by assumption). □

Lemma 1.6. *Let $\mathbf{Z}$ be a set for which $P(Y|do(X), \mathbf{Z})$ is identifiable from $P(Y|X, \mathbf{Z})$, then $\mathbf{Z}$ is a subset of a backdoor set.*

Proof. First, notice that $P(Y|do(X), \mathbf{Z}) = \frac{P(Y, \mathbf{Z}|do(X))}{P(Z|do(X))} = \frac{P(Y, \mathbf{Z}|do(X))}{P(Z)}$. Therefore $P(Y|do(X), \mathbf{Z})$ is only identifiable if $P(Y, \mathbf{Z}|do(X))$ is identifiable. If $\mathbf{Z}$ is not a subset of a backdoor set, then there exists a bidirected path where every variable has a descendant in $\mathbf{Z} \cup Y$ in $\mathcal{G}$ by Lemma 1.5. Let $\mathcal{F}$ be the graph consisting of the bidirected path, and $\mathcal{F}$' be the same graph without X. Then $\mathcal{F}$, $\mathcal{F}$' are $\{Y, \mathbf{Z}\}$ rooted C-forests, and $\{Y, \mathbf{Z}\} \in An(\{Y, \mathbf{Z}\})$, so $\mathcal{F}$, $\mathcal{F}$' form a hedge for $\{Y, \mathbf{Z}\}$. Therefore, $P(Y, \mathbf{Z}|do(X))$ is not identifiable, and $P(Y|do(X), \mathbf{Z})$ is not identifiable. □

Theorem 3.3. *We assume that P_x and $\mathcal{G}_{\overline{X}}$ are faithful to each other. If $\mathbf{Z}$ is a causal Markov boundary for Y relative to X, then $\mathbf{W} = \mathbf{Z} \setminus X$ is a backdoor set.*

Proof. Assume $\mathbf{Z}$ is a causal Markov boundary, but $\mathbf{W}$ is not a backdoor set. Since $P(Y|do(X), \mathbf{W})$ is identifiable, by Lemma 1.6 $\mathbf{W}$ is a subset of a backdoor set $\mathbf{W} \cup \mathbf{Q}$, where $\mathbf{Q} \subseteq (\mathbf{V} \setminus \mathbf{W})$. Since by assumption $\mathbf{W}$ is not a backdoor set, $\mathbf{Q}$ is not the empty set (i.e., $\mathbf{W}$ is a proper subset of a backdoor set). We will show that $P(Y|do(X), \mathbf{W}, \mathbf{Q}) \neq P(Y|do(X), \mathbf{W})$. To show that, we only need to show that $\mathbf{Q}$ is not independent from $\mathbf{W}$ in $\mathcal{G}_{\overline{X}}$. Since $\mathbf{W}$ is not a backdoor set, there exists a backdoor path from X to Y that is m$-$connecting given $\mathbf{W}$, but blocked given $\mathbf{W} \cup \mathbf{Q}$. Thus, some $Q \in \mathbf{Q}$ is a non-collider on that path, therefore $\mathbf{Q}$ are not independent with Y given $\mathbf{W}$. Hence, $P(Y|do(X), \mathbf{W}, \mathbf{Q}) \neq P(Y|do(X), \mathbf{W})$ and therefore $\mathbf{Z}$ does not satisfy Condition (2), and $\mathbf{Z}$ is not a causal Markov boundary (Contradiction). □

Lemma 1.7. *Let $\mathbf{Z} \subseteq \mathbf{V}$ be a backdoor set for X, Y, and let $Q \in (\mathbf{Z} \setminus \text{MB}(Y))$ that has an m-connecting path $Q\pi_{QY}Y$ with Y given $\mathbf{Z} \setminus Q$. Then there exists a variable $W \in (\text{MB}(Y) \setminus \mathbf{Z})$ such that: $W \cup \mathbf{Z}$ is a backdoor set and $W \not\perp\!\!\!\perp Y|\mathbf{Z}$ in $\mathcal{G}_{\overline{X}}$.*

Proof. Let Q be a variable as described above. Then there exists a variable $W \in \text{MB}(Y)$ between Q and Y that is a non-collider on π, otherwise $Q \in Pa(Dis(Y))$, and therefore $Q \in \text{MB}(Y)$. In addition, $W \notin \mathbf{Z}$, otherwise $Q\pi_{QY}Y$ would be blocked given $\mathbf{Z} \setminus Q$. We will now show, by contradiction, that adding W to the conditioning set $\mathbf{Z}$ does not open any backdoor paths from X to Y; hence, $\mathbf{Z} \cup W$ is a backdoor set.

Assume that conditioning on W opens a path π_{XY} between X and Y that is blocked given just $\mathbf{Z}$. Then W must be a descendant of one or more colliders on that path. Let C be the collider closest to X on π_{XY} such that C is blocked on π_{XY} given $\mathbf{Z}$, but open given $\mathbf{Z} \cup W$. Then $X\pi_{XC}C$ is open given $\mathbf{Z}$, and W is a descendant of C. Let $C\pi_{CW}W$ be the (possibly empty) directed path from C to W, and let $W\pi_{WY}Y$ be the subpath of π_{CY} from W to Y. Since C is blocked on π_{XY} given $\mathbf{Z}$, no variable on π_{CW} can be in $\mathbf{Z}$. But then $X\pi_{XC}C\pi_{CW}W\pi_{WY}\,Y$ is an open path from X and Y given $\mathbf{Z}$ in $\mathcal{G}_{\overline{X}}$. Contradiction, since $\mathbf{Z}$ is a backdoor set. Thus, W does not open any backdoor paths, and $\mathbf{Z} \cup W$ is also a backdoor set.

Finally, W is not independent of Y given $\mathbf{Z}$ in $\mathcal{G}_{\overline{X}}$, since $W\pi_{WY}Y$ is open given $\mathbf{Z}$. □

Theorem 3.4. *We assume that P_x and $\mathcal{G}_{\overline{X}}$ are faithful to each other. Every causal Markov boundary $\mathbf{Z}$ of an outcome variable Y w.r.t a treatment variable X is a subset of the Markov boundary $\text{MB}(Y)$.*

Proof. We will show this by contradiction. Specifically, we will show that any set $\mathbf{Z}$ that includes variables $\mathbf{Q}$ not in the Markov boundary of Y cannot satisfy one of the Conditions (2) or (3) of the causal Markov boundary.

Assume that $\mathbf{Z}$ is a causal Markov boundary for Y with respect to X. and let $\mathbf{W} = \mathbf{Z} \setminus X$. Let $\mathbf{Q} = \mathbf{W} \setminus \text{MB}(Y)$ be the non-empty subset of $\mathbf{W}$ that is not a part of the Markov boundary of Y.

If there exists no $Q \in \mathbf{Q}$ that has an m-connecting path $Q\pi_{QY}Y$ to Y given $\mathbf{W} \setminus Q$, then $\mathbf{Q} \perp\!\!\!\perp Y|(\mathbf{W} \setminus \mathbf{Q})$ in $\mathcal{G}_{\overline{X}}$. Conditioning on X cannot open any paths from X to Y; therefore, $\mathbf{Q} \perp\!\!\!\perp Y|X, (\mathbf{W} \setminus \mathbf{Q})$ in $\mathcal{G}_{\overline{X}}$. Then by Rule 1 of the do-calculus [Pearl, 2000], $P(Y|do(X), \mathbf{W}) = P(Y|do(X), \mathbf{W} \setminus Q)$, and $\mathbf{Z}$ does not satisfy Condition (3) of the causal Markov boundary definition (Contradiction).

If there exists a $Q \in (\mathbf{W} \setminus \text{MB}(Y))$ that has an m-connecting path $Q\pi_{QY}Y$ with Y given $\mathbf{Z} \setminus Q$, then by Lemma 1.7, there exists a variable W in $\text{MB}(Y) \setminus \mathbf{Z}$ such that $\mathbf{Z} \cup W$ is also a backdoor set, and $W \not\perp\!\!\!\perp Y|X, \mathbf{Z}$ in $\mathcal{G}_{\overline{X}}$. Then $P(Y|do(X), \mathbf{Z}, W) \neq P(Y|do(X), \mathbf{Z})$. Thus, $\mathbf{Z}$ does not satisfy Condition (2) of the Causal Markov boundary definition (Contradiction).

Thus, $\mathbf{Z}$ cannot include any variables that are not in the Markov boundary of Y. □

Theorem 3.5. *Let $\mathcal{G}$ be a SMCM over X, Y, $\mathbf{V}$ with $\mathbf{V}$ occurring before X and Y. Let $\mathbf{Z} \subseteq \mathbf{V} \cup X$ be the IMB of Y relative to X. If $\mathbf{Z}$ is a causal Markov boundary, then $\text{MB}(Y) = \mathbf{Z}$.*

Proof. $\text{MB}_X(Y) \subseteq \text{MB}(Y)$, so we need to show that $\text{MB}(Y) \subseteq \text{MB}_X(Y)$ when $\text{MB}_X(Y) \in \mathbf{CMB}_X(Y)$. Assume that $\mathbf{Z}$ is both the $\text{MB}_X(Y)$ and a causal Markov boundary, but there exists a variable Q in $\mathbf{Z}$ that is not in $\text{MB}(Y)$. Then Q is reachable from Y through a bidirected path in $\mathcal{G}$ but not in $\mathcal{G}_{\overline{X}}$. Since $\mathcal{G}$ and $\mathcal{G}_{\overline{X}}$ only differ in edges that are into X, this path must be going through an edge that is incoming into X. Thus, $\mathcal{G}$ includes a bidirected path $Y \leftrightarrow \cdots \leftrightarrow X$, and every variable on this path is in $\text{MB}_X(Y)$ =$\mathbf{Z}$. But then $\mathbf{Z} \setminus X$ cannot be a backdoor set, and by Theorem 3.3 $\mathbf{Z}$ cannot be a causal Markov boundary. Contradiction. Thus, the Markov boundary of Y cannot include any more variables than $\mathbf{Z}$. □

2 CONVERGENCE PROOF FOR OBSERVATIONAL MARKOV BOUNDARY (OMB)

Definition 2.1 (Conditional Entropy). Let P be the full joint probability distribution over a set of variables $\mathbf{V}$, let $Y \in \mathbf{V}$ be a variable, and let $\mathbf{Z} \subseteq \mathbf{V} \setminus \{Y\}$ be a set of variables. Then, the conditional entropy of Y given $\mathbf{Z}$ is defined as follows [Cover, 1999]:

$$H(Y|\mathbf{Z}) = -\sum_{y}\sum_{z} P(y, z) \cdot \log P(y|z) \tag{S1}$$

where y and z denote the values of Y and $\mathbf{Z}$, respectively.

Lemma 2.2. *Let $X, Y \in \mathbf{V}$ be two variables and $\mathbf{Z} \subseteq \mathbf{V} \setminus \{X, Y\}$ be a set of variables. Then, $H(Y|\mathbf{Z}) \geq H(Y|X, \mathbf{Z})$, where the entropies are defined by Definition 2.1, and the equality holds if and only if $Y \perp\!\!\!\perp X|\mathbf{Z}$.*

Proof. Applying the chain rule of entropy, the conditional mutual information can be computed as follows [Cover, 1999]:

$$I(X;Y|\mathbf{Z}) = H(Y|\mathbf{Z}) - H(Y|X,\mathbf{Z})\,. \tag{S2}$$

Given that the mutual information is nonnegative (i.e., $I(X;Y|\mathbf{Z}) \geq 0$) and $I(X;Y|\mathbf{Z}) = 0$ if and only if $Y \perp\!\!\!\perp X|\mathbf{Z}$ (see [Cover, 1999], page 29), it follows that:

$$\begin{aligned} H(Y|\mathbf{Z}) - H(Y|X,\mathbf{Z}) \geq 0 \\ H(Y|\mathbf{Z}) \geq H(Y|X,\mathbf{Z})\,, \end{aligned} \tag{S3}$$

where the equality holds if and only if $Y \perp\!\!\!\perp X|\mathbf{Z}$. □

For brevity, let $\mathbf{V} = \{\mathbf{V} \cup X\}$, where X is a treatment variable, and let Y be an outcome variable in the remainder of this section.

Lemma 2.3. *All Markov blankets of Y have the same entropy.*

Proof. By definition, $\mathbf{Z}'$ is the Markov blanket of Y if and only if $P(Y|\mathbf{Z}',\mathbf{W}) = P(Y|\mathbf{Z}')$ for any $\mathbf{W} \subseteq \mathbf{V} \setminus \mathbf{Z}'$, which indicates that $Y \perp\!\!\!\perp \mathbf{W}|\mathbf{Z}'$. Also, according to Lemma 2.2, $H(Y|\mathbf{Z}') = H(Y|\mathbf{Z}',\mathbf{W})$ for any $\mathbf{W} \subseteq \mathbf{V} \setminus \mathbf{Z}'$. Let $\mathbf{Z}$ also be a Markov blanket of Y. By multiple applications of Lemma 2.2, we obtain:

$$H(Y|\mathbf{Z}') = H(Y|\mathbf{Z}',\mathbf{V} \setminus \mathbf{Z}') = H(Y|\mathbf{V}) = H(Y|\mathbf{Z},\mathbf{V} \setminus \mathbf{Z}') = H(Y|\mathbf{Z}) \tag{S4}$$

□

Lemma 2.4. *Let $\mathbf{Z}'$ be a Markov blanket of Y and let $\mathbf{Z}$ be a set of variables that is not a Markov blanket of Y. Then, $H(Y|\mathbf{Z}') < H(Y|\mathbf{Z})$, where the entropies are defined by Definition 2.1.*

Proof. Assume there is exists a set $\mathbf{W} \subseteq \mathbf{V} \setminus \mathbf{Z}$ such that $P(Y|\mathbf{Z},\mathbf{W}) \neq P(Y|\mathbf{Z})$. According to Lemma 2.2 we have:

$$H(Y|\mathbf{Z}, W) < H(Y|\mathbf{Z}). \tag{S5}$$

Also, given that $\mathbf{V}$ is a superset of $(\mathbf{Z} \cup \mathbf{W})$, we have:

$$H(Y|\mathbf{V}) \leq H(Y|\mathbf{Z},\mathbf{W}). \tag{S6}$$

Therefore,

$$H(Y|\mathbf{V}) < H(Y|\mathbf{Z}). \tag{S7}$$

Also, since $\mathbf{Z}'$ is a Markov blanket of Y, by Lemma 2.3 we have:

$$H(Y|\mathbf{Z}') = H(Y|\mathbf{V}). \tag{S8}$$

Combining Equations (S7) and (S8), we obtain:

$$H(Y|\mathbf{Z}') < H(Y|\mathbf{Z}). \tag{S9}$$

□

Lemma 2.5. *Given dataset D_o that contains samples from a strictly positive distribution P, which is a perfect map for a SMCM $\mathcal{G}$, the BD score [Heckerman et al., 1995] for $\log P(D_o|\mathbf{Z})$ is defined as follows in the large sample limit:*

$$\lim_{N\to\infty} \log P(D_o|\mathbf{Z}) = \lim_{N\to\infty} -N \cdot H(Y|\mathbf{Z}) - \frac{q \cdot (r-1)}{2} \log N + const., \tag{S10}$$

Proof. The BD score for $P(D_o|\mathbf{Z})$ is calculated as follows [Heckerman et al., 1995]:

$$P(D_o|\mathbf{Z}) = \prod_{j=1}^{q} \frac{\Gamma(\alpha_j)}{\Gamma(\alpha_j + N_j)} \cdot \prod_{k=1}^{r} \frac{\Gamma(\alpha_{jk} + N_{jk})}{\Gamma(\alpha_{jk})}, \tag{S11}$$

where q denotes instantiations of variables in $\mathbf{Z}$ and r denotes values of variable Y. The term N_{jk} is the number of cases in data in which variable $Y = k$ and its parent $\mathbf{Z} = j$; also, $N_j = \sum_{k=1}^{r} N_{jk}$. The term α_{jk} is a finite positive real number that is called Dirichlet prior parameter and may be interpreted as representing "pseudo-counts", where $\alpha_j = \sum_{k=1}^{r} \alpha_{jk}$. BD can be re-written in log form as follows:

$$\log P(D_o|\mathbf{Z}) = \sum_{j=1}^{q} \left[\log\Gamma(\alpha_j) - \log\Gamma(\alpha_j + N_j) + \sum_{k=1}^{r} [\log\Gamma(\alpha_{jk} + N_{jk}) - \log\Gamma(\alpha_{jk})] \right]. \tag{S12}$$

We can re-arrange the terms in Eq. (S12) to gather the constant terms as follows:

$$\begin{aligned} \log P(D_o|\mathbf{Z}) &= \sum_{j=1}^{q} \left[-\log\Gamma(\alpha_j + N_j) + \sum_{k=1}^{r} \log\Gamma(\alpha_{jk} + N_{jk}) \right] + \sum_{j=1}^{q} \left[\log\Gamma(\alpha_j) - \sum_{k=1}^{r} \log\Gamma(\alpha_{jk}) \right] \\ &= \sum_{j=1}^{q} \left[-\log\Gamma(\alpha_j + N_j) + \sum_{k=1}^{r} \log\Gamma(\alpha_{jk} + N_{jk}) \right] + const. \end{aligned} \tag{S13}$$

Using the Stirling's approximation of $\lim_{n\to\infty} \log\Gamma(n) = (n - \frac{1}{2})\log(n) - n + const.$, we can re-write Eq. (S13) as follows:

$$\begin{aligned} &\lim_{N\to\infty} \log P(D_o|\mathbf{Z}) = \lim_{N\to\infty} \sum_{j=1}^{q} \Bigg[-(\alpha_j + N_j - \frac{1}{2})\log(\alpha_j + N_j) + (\alpha_j + N_j) \\ &+ \sum_{k=1}^{r} \left((\alpha_{jk} + N_{jk} - \frac{1}{2})\log(\alpha_{jk} + N_{jk}) - (\alpha_{jk} + N_{jk}) \right) \Bigg] + const. \\ &= \lim_{N\to\infty} \sum_{j=1}^{q} \Bigg[-\alpha_j \log(\alpha_j + N_j) - N_j \log(\alpha_j + N_j) + \frac{1}{2}\log(\alpha_j + N_j) + \alpha_j + N_j \\ &+ \sum_{k=1}^{r} \left(\alpha_{jk} \log(\alpha_{jk} + N_{jk}) + N_{jk} \log(\alpha_{jk} + N_{jk}) - \frac{1}{2}\log(\alpha_{jk} + N_{jk}) - \alpha_{jk} - N_{jk} \right) \Bigg] + const. \\ &= \lim_{N\to\infty} \sum_{j=1}^{q} \left[-N_j \log(\alpha_j + N_j) + \sum_{k=1}^{r} N_{jk} \log(\alpha_{jk} + N_{jk}) \right] + \sum_{j=1}^{q} \left[-\alpha_j \log(\alpha_j + N_j) + \sum_{k=1}^{r} \alpha_{jk} \log(\alpha_{jk} + N_{jk}) \right] \\ &+ \frac{1}{2} \sum_{j=1}^{q} \left[\log(\alpha_j + N_j) - \sum_{k=1}^{r} \log(\alpha_{jk} + N_{jk}) + \alpha_j + N_j - \sum_{k=1}^{r} (\alpha_{jk} + N_{jk}) \right] + const. \\ &= \lim_{N\to\infty} \sum_{j=1}^{q} \left[-N_j \log(\alpha_j + N_j) + \sum_{k=1}^{r} N_{jk} \log(\alpha_{jk} + N_{jk}) \right] + \sum_{j=1}^{q} \left[-\alpha_j \log(\alpha_j + N_j) + \sum_{k=1}^{r} \alpha_{jk} \log(\alpha_{jk} + N_{jk}) \right] \\ &+ \frac{1}{2} \sum_{j=1}^{q} \left[\log(\alpha_j + N_j) - \sum_{k=1}^{r} \log(\alpha_{jk} + N_{jk}) \right] + const. \end{aligned} \tag{S14}$$

In the last step of Eq. (S14), we used the facts that $\sum_{k=1}^{r} N_{jk} = N_j$ and $\sum_{k=1}^{r} \alpha_{jk} = \alpha_j$, and we applied these identities again to that equation to obtain the following:

$$\begin{aligned} &\lim_{N\to\infty} \log P(D_o|\mathbf{Z}) = \\ &\lim_{N\to\infty} \sum_{j=1}^{q} \sum_{k=1}^{r} \left[N_{jk} \log(\frac{\alpha_{jk} + N_{jk}}{\alpha_j + N_j}) + \alpha_{jk} \log(\frac{\alpha_{jk} + N_{jk}}{\alpha_j + N_j}) \right] + \frac{1}{2} \sum_{j=1}^{q} \left[\log(\alpha_j + N_j) - \sum_{k=1}^{r} \log(\alpha_{jk} + N_{jk}) \right] + const. \end{aligned} \tag{S15}$$

Given that

$$\lim_{N\to\infty} \frac{\alpha_{jk} + N_{jk}}{\alpha_j + N_j} = \frac{N_{jk}}{N_j}$$

and

$$\lim_{N\to\infty} \sum_{j=1}^{q}\sum_{k=1}^{r} \alpha_{jk} \log(\frac{\alpha_{jk} + N_{jk}}{\alpha_j + N_j}) = const.,$$

in the limit, Eq. (S15) becomes:

$$\lim_{N\to\infty} \log P(D_o|\mathbf{Z}) = \lim_{N\to\infty} \sum_{j=1}^{q}\sum_{k=1}^{r} N_{jk} \log \frac{N_{jk}}{N_j} + \frac{1}{2}\sum_{j=1}^{q}\left[\log(\alpha_j + N_j) - \sum_{k=1}^{r} \log(\alpha_{jk} + N_{jk})\right] + const., \quad \text{(S16)}$$

or equivalently:

$$\begin{aligned}\lim_{N\to\infty} \log P(D_o|\mathbf{Z}) &= \lim_{N\to\infty} N \cdot \sum_{j=1}^{q}\sum_{k=1}^{r} \frac{N_{jk}}{N} \log \frac{N_{jk}}{N_j} + \frac{1}{2}\sum_{j=1}^{q}\left[\log(\alpha_j + N_j) - \sum_{k=1}^{r} \log(\alpha_{jk} + N_{jk})\right] + const. \\ &= \lim_{N\to\infty} -N \cdot H(Y|\mathbf{Z}) + \frac{1}{2}\sum_{j=1}^{q}\left[\log(\alpha_j + N_j) - \sum_{k=1}^{r} \log(\alpha_{jk} + N_{jk})\right] + const.\end{aligned} \quad \text{(S17)}$$

To simplify the second term in Eq. (S17), we divide the arguments in the log terms by N and equivalently add $\log N$ terms as follows:

$$\begin{aligned}&\lim_{N\to\infty} \frac{1}{2}\sum_{j=1}^{q}\left[\log(\alpha_j + N_j) - \sum_{k=1}^{r} \log(\alpha_{jk} + N_{jk})\right] \\ &= \lim_{N\to\infty} \frac{1}{2}\sum_{j=1}^{q}\left[\log(\frac{\alpha_j + N_j}{N}) + \log N - \sum_{k=1}^{r} \log(\frac{\alpha_{jk} + N_{jk}}{N}) + \log N\right] \\ &= \lim_{N\to\infty} \frac{1}{2}\sum_{j=1}^{q}\left(\log N - \sum_{k=1}^{r} \log N\right) + \frac{1}{2}\sum_{j=1}^{q}\left[\log(\frac{\alpha_j + N_j}{N}) - \sum_{k=1}^{r} \log(\frac{\alpha_{jk} + N_{jk}}{N})\right] \\ &= -\frac{q(r-1)}{2} \log N + const.\end{aligned} \quad \text{(S18)}$$

Combining Equations (S17) and (S18), we obtain:

$$\lim_{N\to\infty} \log P(D_o|\mathbf{Z}) = \lim_{N\to\infty} -N \cdot H(Y|\mathbf{Z}') - \frac{q \cdot (r-1)}{2} \log N + const. \quad \text{(S19)}$$

□

Theorem 4.2. *Given dataset D_o that contains samples from a strictly positive distribution P, which is a perfect map for a SMCM $\mathcal{G}$, the BD score [Heckerman et al., 1995] will assign the highest score to the OMB of Y in the large sample limit.*

Proof. Let $\mathbf{Z}'$ be the OMB of Y and $\mathbf{Z} \subseteq \mathbf{V}$ be an arbitrary set. We want to show that:

$$\lim_{N\to\infty} \frac{P(D_o|\mathbf{Z})}{P(D_o|\mathbf{Z}')} = \begin{cases} 1 & \text{iff } \mathbf{Z} \text{ is an OMB of } Y \\ 0 & \text{otherwise} \end{cases}, \quad \text{(S20)}$$

Applying Lemma 2.5 we have:

$$\lim_{N\to\infty} \log \frac{P(D_o|\mathbf{Z})}{P(D_o|\mathbf{Z}')} = \lim_{N\to\infty} N \cdot [H(Y|\mathbf{Z}') - H(Y|\mathbf{Z})] + \frac{(q'-q)\cdot(r-1)}{2} \log N. \quad \text{(S21)}$$

where q and q' are the number of possible parent instantiations of Y with $\mathbf{Z}$ and $\mathbf{Z}'$ as the set of parents. There are three possible cases:

Case 1: $\mathbf{Z}$ is a Markov blanket of Y and its OMB.

Since both $\mathbf{Z}'$ and $\mathbf{Z}$ are Markov blankets of Y, $H(Y|\mathbf{Z}) = H(Y|\mathbf{Z}')$ by Lemma 2.3. Thus, the first term in Eq. (S21) becomes 0. Also, given that $\mathbf{Z}'$ and $\mathbf{Z}$ are OMBs, they have the same number of parameters $q' = q$, by which the second term in Eq. (S21) becomes 0 in the limit as $N \to \infty$, or equivalently Eq. (S20) approaches to 1.

Case 2: $\mathbf{Z}$ is a Markov blanket of Y but not its OMB.

According to Lemma 2.3 $H(Y|\mathbf{Z}) = H(Y|\mathbf{Z}')$; therefore, the first term in Eq. (S21) becomes 0 and we obtain:

$$\lim_{N\to\infty} \frac{P(D_o|\mathbf{Z})}{P(D_o|\mathbf{Z}')} = \lim_{N\to\infty} \frac{(q'-q)\cdot(r-1)}{2} \log N. \tag{S22}$$

Given that $\mathbf{Z}'$ is the OMB with minimum number of variables, and therefore, minimum number of parameters $q' < q$. Thus, the term $(q' - q)$ becomes a negative constant. Also, the term $\frac{(r-1)}{2}$ is a positive constant. Consequently, Eq. (S22) goes to $-\infty$ in the limit as $N \to \infty$, which implies that Eq. (S20) approaches to 0.

Case 3: $\mathbf{Z}$ is not a Markov blanket of Y.

The first term in Eq. (S21) is of $O(N)$ and dominates the second term, which is $O(\log N)$. According to Lemma 2.4, $H(Y|\mathbf{Z}') < H(Y|\mathbf{Z})$; thus, the term $H(Y|\mathbf{Z}') - H(Y|\mathbf{Z})$ becomes a negative number. As a result, Eq. (S21) becomes $-\infty$, which equivalently implies that Eq. (S20) becomes 0. □

References

Thomas M Cover. *Elements of Information Theory*. John Wiley & Sons, 1999.

David Heckerman, Dan Geiger, and David M Chickering. Learning Bayesian networks: The combination of knowledge and statistical data. *Machine Learning*, 20(3):197–243, 1995.

Leonard Henckel, Emilija Perković, and Marloes H. Maathuis. Graphical criteria for efficient total effect estimation via adjustment in causal linear models. *arXiv preprint arXiv:1907.02435*, 2019.

J Pearl. *Causality: Models, Reasoning and Inference*. Cambridge University Press, 2000.

Thomas Richardson, Peter Spirtes, et al. Ancestral graph Markov models. *The Annals of Statistics*, 30(4):962–1030, 2002.

Ilya Shpitser and Judea Pearl. Identification of joint interventional distributions in recursive semi-Markovian causal models. In *Proceedings of the 21st National Conference on Artificial Intelligence*, pages 1219–1226, 2006.