SUPPLEMENTARY MATERIAL OF "CAUSAL INFERENCE IN DEGENERATE SYSTEMS: AN IMPOSSIBILITY RESULT"

S.1. **Proof of Lemma 1.** We first present a proposition based on the weak union property of probability distributions (Pearl, 1988).

Proposition S1. *Any superset of a Markov blanket is still a Markov blanket.*

Now consider two Markov boundaries $\mathcal{M}_1$, $\mathcal{M}_2$ within $\{X\} \cup \mathcal{K}$. Let $\mathcal{M}_1 = \{X\} \cup \mathcal{Z}^1$, $X \notin \mathcal{M}_2$, $\mathcal{M}_2 \backslash \mathcal{M}_1 = \mathcal{Z}^2$, $\mathcal{K} \cup \{X\} \backslash (\mathcal{M}_1 \cup \mathcal{M}_2) = \mathcal{Z}^3$, where $\mathcal{Z}^1 = \{Z_1, \ldots, Z_n\}$, $\mathcal{Z}^2 = \{Z'_1, \ldots, Z'_m\}$, $\mathcal{Z}^3 = \{Z''_1, \ldots, Z''_l\}$. Therefore $\mathcal{K} = \mathcal{Z}^1 \cup \mathcal{Z}^2 \cup \mathcal{Z}^3$.

Fix $z_0^1 \in \mathbb{Z}^1$ such that $f(z_0^1) > 0$. Assume that for $x_i \in \mathbb{X}$, $f(x_i, z_0^1) > 0$ is true for $i \in \{1, \ldots, p\}$. Assume that for $z_j^2 \in \mathbb{Z}^2$, $f(z_0^1, z_j^2) > 0$ is true for $j \in \{1, \ldots, q\}$. Consider any $y \in \mathbb{Y}$.

To obtain contradiction, we assume that $f(x_i, z_0^1, z_j^2) > 0$ for all $i \in \{1, \ldots, p\}$ and all $j \in \{1, \ldots, q\}$.

Since $X \perp\!\!\!\perp Y \mid (\mathcal{Z}^1, \mathcal{Z}^2)$ (Proposition S1) for all $i, r \in \{1, \ldots, p\}$ and all $j \in \{1, \ldots, q\}$,

$$f(y \mid x_i, z_0^1, z_j^2) = f(y \mid x_r, z_0^1, z_j^2).$$

Since $\mathcal{Z}^2 \perp\!\!\!\perp Y \mid (X, \mathcal{Z}^1)$ for all $r \in \{1, \ldots, p\}$ and all $j, s \in \{1, \ldots, q\}$,

$$f(y \mid x_r, z_0^1, z_j^2) = f(y \mid x_r, z_0^1, z_s^2).$$

All the conditions have positive probabilities, so the conditional probabilities are well-defined.

Then we have

$$f(y \mid x_i, z_0^1, z_j^2) = f(y \mid x_r, z_0^1, z_s^2),$$

for all $i, r \in \{1, \ldots, p\}$ and all $j, s \in \{1, \ldots, q\}$.

Since this is true for any possible values of X and $\mathcal{Z}^2$ when $\mathcal{Z}^1 = z_0^1$, we know that

$$f(y \mid x_i, z_0^1, z_j^2) = f(y \mid z_0^1).$$

Therefore, for all $z_1^1 \in \mathcal{Z}^1$ with $f(z_1^1) > 0$, all $y \in \mathbb{Y}$ and all i, j,

$$f(x_i, z_j^2, y \mid z_1^1) = f(x_i, z_j^2 \mid z_1^1) f(y \mid z_1^1)$$

is valid.

This implies that $(X, \mathcal{Z}^2) \perp\!\!\!\perp Y \mid \mathcal{Z}^1$, therefore $X \perp\!\!\!\perp Y \mid \mathcal{Z}^1$, $\textsc{mi}(Y, \mathcal{Z}^1) = \textsc{mi}(Y, (X, \mathcal{Z}^1))$. Since $\mathcal{M}_1 = \{X\} \cup \mathcal{Z}^1$, $\textsc{mi}(Y, (X, \mathcal{Z}^1)) = \textsc{mi}(Y, \mathcal{K})$. Thus $\textsc{mi}(Y, \mathcal{Z}^1) = \textsc{mi}(Y, \{X\} \cup \mathcal{K})$, implying that $\mathcal{Z}^1$ is a Markov blanket, which is a contradiction. So there exists $x \in \mathbb{X}$, $z_0^1 \in \mathbb{Z}^1$, $z_1^2 \in \mathbb{Z}^2$ such that $f(x, z_0^1) > 0$

(implies $f(x) > 0$), $f(z_0^1, z_1^2) > 0$, but $f(x, z_0^1, z_1^2) = 0$. Choose $z_1^3 \in \mathbb{Z}^3$ such that $f(z_0^1, z_1^2, z_1^3) > 0$, and let $k = (z_0^1, z_1^2, z_1^3)$, then $f(x) > 0$, $f(k) > 0$, but $f(x, k) = 0$.

S.2. **Proof of Proposition 2.** In this setting, when n is much larger than fixed m, due to the property of Dirichlet distribution, with probability at least $1 - \delta/2$, we can modify $\mathfrak{p}$ to $\bar{\mathfrak{p}}$ such that three pre-chosen variables X, Y, Z are independent under $\bar{\mathfrak{p}}$, and $\mathrm{d}(\mathfrak{p}, \bar{\mathfrak{p}}) < \epsilon/2$. Then construct $\bar{X}, \bar{Y}, \bar{Z}$: $\bar{X}, \bar{Y}, \bar{Z}$ equal X, Y, Z if none of X, Y, Z is 1; $\bar{X}, \bar{Y}, \bar{Z}$ equal 1 if at least one of X, Y, Z is 1. Now either all $\bar{X}, \bar{Y}, \bar{Z}$ equal 1, or none of them equals 1 (they are independent in this case). Substitute X, Y, Z by $\bar{X}, \bar{Y}, \bar{Z}$ to obtain a new distribution $\mathfrak{p}'$. When m is large enough, $\mathrm{d}(\mathfrak{p}', \bar{\mathfrak{p}}) < \epsilon/2$. Now under $\mathfrak{p}'$, $\bar{X}$ and $\bar{Z}$ contain exactly the same unique information of $\bar{Y}$, thus there exist multiple Markov boundaries. Besides, $\mathrm{d}(\mathfrak{p}, \mathfrak{p}') < \epsilon/2$.

S.3. **Proof of Lemma 2.** In the following we will assume there is only one pair of (x, l) such that $f(x) > 0$, $f(l) > 0$, $f(x, l) = 0$. If there are multiple pairs, we can treat them one by one.

We construct a family of probability distributions $\mathfrak{p}_i^\eta$ with mass functions f_i^η based on $\mathfrak{p}$. For $(x', l') \neq (x, l)$, $f_i^\eta(x', y, l') = (1-\eta) f(x', y, l')$. $f_i^\eta(x, l) = \eta > 0$, $f_i^\eta(y_j \mid x, l) = \alpha_i^j$, where $\alpha_i^j \geq 0$, $\sum_j \alpha_i^j = 1$. Then for each i, $\textsc{cs}[\mathfrak{p}_i^\eta](X \to Y)$ can be defined, and when $\eta \to 0$, f_i^η converges to f. The total variation distance between f and f_i^η is η.

When $\eta \to 0$,

$$\begin{aligned}
\textsc{cs}[\mathfrak{p}_i^\eta](X \to Y) = &\sum_{x' \in \mathbb{X}} \sum_{y' \in \mathbb{Y}} \sum_{l' \neq l} f_i^\eta(x', y', l') \log \frac{f_i^\eta(y' \mid x', l')}{\sum_{x'' \in \mathbb{X}} f_i^\eta(y' \mid x'', l') f_i^\eta(x'')} \\
&+ \sum_{x' \in \mathbb{X}} \sum_{y' \in \mathbb{Y}} f_i^\eta(x', y', l) \log \frac{f_i^\eta(y' \mid x', l)}{\sum_{x'' \in \mathbb{X}} f_i^\eta(y' \mid x'', l) f_i^\eta(x'')} \\
\to &\sum_{x' \in \mathbb{X}} \sum_{y' \in \mathbb{Y}} \sum_{l' \neq l} f(x', y', l') \log \frac{f(y' \mid x', l')}{\sum_{x'' \in \mathbb{X}} f(y' \mid x'', l') f(x'')} \\
&+ \sum_{x' \neq x} \sum_{y' \in \mathbb{Y}} f(x', y', l) \log f(y' \mid x', l) \\
&- \sum_j f(y_j, l) \log\{f(x)\alpha_i^j + \sum_{x' \neq x} f(x') f(y_j \mid x', l)\}.
\end{aligned}$$

For different i, when we let $\eta \to 0$, the only different terms are

$$-\sum_j f(y_j, l) \log\{f(x)\alpha_i^j + \sum_{x' \neq x} f(x') f(y_j \mid x', l)\}.$$

We will show that the above term is not a constant with $\{\alpha_i^j\}$. Therefore we can find two groups of $\{\alpha_i^j\}$ for $i = 1, 2$ such that $g_1 = \lim_{\eta \to 0} \textsc{cs}[\mathfrak{p}_1^\eta](X \to Y) < \lim_{\eta \to 0} \textsc{cs}[\mathfrak{p}_2^\eta](X \to Y) = g_2$.

If there is only one y_1 such that $f(y_1, l) > 0$, then

$$-\sum_j f(y_j, l) \log\{f(x)\alpha_i^j + \sum_{x' \neq x} f(x') f(y_j \mid x', l)\}$$
$$= -f(y_1, l) \log\{f(x)\alpha_i^1 + \sum_{x' \neq x} f(x') f(y_1 \mid x', l)\}.$$

It is not a constant when we change α_i^1.

If there are at least two values y_1, y_2 of Y, such that $f(y_1, l) > 0$, $f(y_2, l) > 0$, then we can change α_i^1 while keeping $\alpha_i^1 + \alpha_i^2 = d$, and leave other α_i^j fixed.

Set $f(y_1, l) = a_1$, $f(y_2, l) = a_2$, $f(x) = c$, $\sum_{x' \neq x} f(x') f(y_1 \mid x', l) = b_1$, $\sum_{x' \neq x} f(x') f(y_2 \mid x', l) = b_2$. All these terms are positive. Then in $-\sum_j f(y_j, l) \log\{f(x)\alpha_i^j + \sum_{x' \neq x} f(x') f(y_j \mid x', l)\}$, terms containing α_i^1 and α_i^2 are

$$-a_1 \log(c\alpha_i^1 + b_1) - a_2 \log\{c(d - \alpha_i^1) + b_2\}.$$

Its derivative with respect to α_i^1 is

$$-\frac{a_1 c}{c\alpha_i^1 + b_1} + \frac{a_2 c}{c(d - \alpha_i^1) + b_2}.$$

If the derivative always equal 0 in an interval, then we should have

$$\frac{a_1}{a_2} \equiv \frac{c\alpha_i^1 + b_1}{c(d - \alpha_i^1) + b_2},$$

which is incorrect.

Now we have two groups of $\{\alpha_i^j\}$ for $i = 1, 2$ such that

$$g_1 = \lim_{\eta \to 0} \text{CS}[\mathfrak{p}_1^\eta](X \to Y) < \lim_{\eta \to 0} \text{CS}[\mathfrak{p}_2^\eta](X \to Y) = g_2.$$

Then for any $g \in (g_1, g_2)$, any $\delta > 0$, we can find $\eta < \delta$ small enough such that $\text{CS}[\mathfrak{p}_1^\eta](X \to Y) < g$, $\text{CS}[\mathfrak{p}_2^\eta](X \to Y) > g$. Then we change $\{\alpha_1^j\}$ continuously to $\{\alpha_2^j\}$. During this process CS is always defined, and there exists $\{\alpha_3\}$ such that $\text{CS}[\mathfrak{p}_3^\eta](X \to Y) = g$.

This shows that $\text{CS}(X \to Y)$ is essentially ill-defined.

Since $\text{CS}(X \to Y)$ and $\text{PMI}(X, Y \mid \mathcal{L})$ have the same non-zero terms containing $f(\cdot \mid x, l)$, the same argument shows that $\text{PMI}(X, Y \mid \mathcal{L})$ is not well-defined.

S.4. **Proof of Lemma 3 when X is discrete.** The proofs for discrete and continuous X are different, therefore we state them separately. Whether Y is discrete or continuous does not matter, therefore we assume Y is discrete/continuous when X is discrete/continuous. We impose some restrictions to simplify the proofs. If X is discrete, then U_X is an arbitrary discrete random variable which takes all the values of X with positive probabilities. If X is continuous, then U_X is continuous, and its density function is always positive.

$\textsc{cmi}(X, Y \mid \mathcal{S}_1) = \sum_{s_1} \mathrm{pr}(\mathcal{S}_1 = s_1)\textsc{cmi}(X, Y \mid \mathcal{S}_1 = s_1)$. For a fixed s_1, assume X takes values $1, \ldots, r'$, U_X takes values $1, \ldots, r', \ldots, r$, and Y takes values $1, \ldots, t$ with positive probabilities. Denote $\mathrm{pr}(X = i, Y = j \mid \mathcal{S}_1 = s_1)$ by p_{ij}. Define $p_{-j} = \sum_i p_{ij}$, $p_{i-} = \sum_j p_{ij}$. With ϵ-noise, $p^{\epsilon}_{-j} = p_{-j}$, $p^{\epsilon}_{ij} = (1-\epsilon)p_{ij} + \epsilon q_i p_{-j}$, $p^{\epsilon}_{i-} = (1-\epsilon)p_{i-} + \epsilon q_i$. Here q_i is the density of U_X. Then we have

$$\textsc{cmi}(X, Y \mid \mathcal{S}_1 = s_1) = \sum_{j=1}^{t} \sum_{i=1}^{r'} p_{ij} \log \frac{p_{ij}}{p_{i-}p_{-j}},$$

$$\textsc{cmi}(X^{\epsilon}, Y \mid \mathcal{S}_1 = s_1) = \sum_{j=1}^{t} \sum_{i=1}^{r} \{(1-\epsilon)p_{ij} + \epsilon q_i p_{-j}\} \log \frac{(1-\epsilon)p_{ij} + \epsilon q_i p_{-j}}{\{(1-\epsilon)p_{i-} + \epsilon q_i\}p_{-j}}.$$

$$\begin{aligned} &\textsc{cmi}(X, Y \mid \mathcal{S}_1 = s_1) - \textsc{cmi}(X^{\epsilon}, Y \mid \mathcal{S}_1 = s_1) = \\ &\sum_{j=1}^{t} \sum_{i=1}^{r} \Big[(1 - \epsilon + q_i\epsilon)p_{ij} \log \frac{p_{ij}}{p_{i-}p_{-j}} + \sum_{k \neq i} q_i \epsilon p_{kj} \log \frac{p_{kj}}{p_{k-}p_{-j}} \\ &\quad - \{(1-\epsilon)p_{ij} + \epsilon q_i p_{-j}\} \log \frac{(1-\epsilon)p_{ij} + \epsilon q_i p_{-j}}{p_{-j}\{(1-\epsilon)p_{i-} + \epsilon q_i\}}\Big]. \end{aligned}$$

If $p_{k-} = 0$, namely $k = r' + 1, \ldots, r$, then we stipulate $\frac{p_{kj}}{p_{k-}p_{-j}} = 1$.

For fixed i, j and $k = 1, \ldots, r$, set

$$\begin{aligned} a^k_{ij} &= \frac{p_{kj}}{p_{k-}p_{-j}}, \\ b^k_{ij} &= \frac{\epsilon q_i p_{k-}}{(1-\epsilon)p_{i-} + \epsilon q_i} \quad \text{for} \quad k \neq i, \\ b^i_{ij} &= \frac{(1 - \epsilon + q_i\epsilon)p_{i-}}{(1-\epsilon)p_{i-} + \epsilon q_i}, \\ c_{ij} &= p_{-j}\{(1-\epsilon)p_{i-} + \epsilon q_i\}. \end{aligned}$$

Here we know that $p_{-j} > 0$, $(1-\epsilon)p_{i-} + \epsilon q_i > 0$.

Then we have

$$\begin{aligned} &\textsc{cmi}(X, Y \mid \mathcal{S}_1 = s_1) - \textsc{cmi}(X^{\epsilon}, Y \mid \mathcal{S}_1 = s_1) \\ &= \sum_{j=1}^{t} \sum_{i=1}^{r} c_{ij} \{\sum_{k=1}^{r} b^k_{ij} a^k_{ij} \log a^k_{ij} - (\sum_{k=1}^{r} a^k_{ij} b^k_{ij}) \log(\sum_{k=1}^{r} a^k_{ij} b^k_{ij})\} \geq 0. \end{aligned}$$

The last step is Jensen's inequality, since $a^k_{ij} \geq 0$, $b^k_{ij} \geq 0$, $\sum_{k=1}^{r} b^k_{ij} = 1$, $c_{ij} > 0$, $f(x) = x \log x$ is strictly convex down when $x \geq 0$ (stipulate $0 \log 0 = 0$).

The equality holds if and only if for each i, j, $a^1_{ij} = a^2_{ij} = \cdots = a^{r'}_{ij}$, which means p_{ij}/p_{i-} are equal for all $i \leq r'$. Since $\sum_{i=1}^{r'} p_{i-}(p_{ij}/p_{i-}) = p_{-j}$, $\sum_{i=1}^{r'} p_{i-} = 1$, we have $p_{ij}/p_{i-} = p_{-j}$ for each i, j such that $p_{i-} > 0$ and $p_{-j} > 0$. This is equivalent with that X and Y are independent conditioned on $\mathcal{S}_1 = s_1$.

$\text{CMI}(X, Y \mid \mathcal{S}_1) = 0$ if and only if X and Y are independent conditioned on any possible value of $\mathcal{S}_1$. Therefore, $\text{CMI}(X^\epsilon, Y \mid \mathcal{S}_1) \leq \text{CMI}(X, Y \mid \mathcal{S}_1)$, and the equality holds if and only if $\text{CMI}(X, Y \mid \mathcal{S}_1) = 0$.

S.5. **Proof of Lemma 3 when X is continuous.**

$$\text{CMI}(X, Y \mid \mathcal{S}_1) = \int_{-\infty}^{\infty} \text{CMI}(X, Y \mid \mathcal{S}_1 = s_1) h(s_1) \mathrm{d}s_1,$$

where $h(s_1)$ is the probability density function of $\mathcal{S}_1$. For a fixed s_1, denote the joint probability density function of X, Y conditioned on $\mathcal{S}_1 = s_1$ by $p(x, y)$. Define $p_1(x) = \int_{-\infty}^{\infty} p(x, y)\mathrm{d}y$, $p_2(y) = \int_{-\infty}^{\infty} p(x, y)\mathrm{d}x$. With ϵ-noise, the joint probability density function of X, Y conditioned on $\mathcal{S}_1 = s_1$ is $(1-\epsilon)p(x, y) + \epsilon q(x)p_2(y)$, where $q(x)$ is the density function of U_X. Notice that $\int_{-\infty}^{\infty} q(x)\mathrm{d}x = 1$, $\int_{-\infty}^{\infty}[(1-\epsilon)p(x, y) + \epsilon q(x)p_2(y)]\mathrm{d}x = p_2(y)$. Then we have

$$\begin{aligned}
&\text{CMI}(X, Y \mid \mathcal{S}_1 = s_1) - \text{CMI}(X^\epsilon, Y \mid \mathcal{S}_1 = s_1) \\
&= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} p(x, y) \log \frac{p(x, y)}{p_1(x)p_2(y)} \mathrm{d}x\mathrm{d}y \\
&- \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \{(1-\epsilon)p(x, y) + \epsilon q(x)p_2(y)\} \log \frac{(1-\epsilon)p(x, y) + \epsilon q(x)p_2(y)}{\{(1-\epsilon)p_1(x) + \epsilon q(x)\}p_2(y)} \mathrm{d}x\mathrm{d}y \\
&= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \Big[(1-\epsilon)p(x, y) \log \frac{p(x, y)}{p_1(x)p_2(y)} \\
&+ q(x)\epsilon\Big\{\int_{-\infty}^{\infty} p(x_0, y) \log \frac{p(x_0, y)}{p_1(x_0)p_2(y)} \mathrm{d}x_0\Big\} \\
&- \{(1-\epsilon)p(x, y) + \epsilon q(x)p_2(y)\} \log \frac{(1-\epsilon)p(x, y) + \epsilon q(x)p_2(y)}{\{(1-\epsilon)p_1(x) + \epsilon q(x)\}p_2(y)}\Big] \mathrm{d}x\mathrm{d}y.
\end{aligned}$$

For fixed x, y, we can define a probability measure $\mu_{x,y}(x_0)$ on $\mathbb{R}$, which is a mixture of discrete and continuous type measures. For the discrete component, it has probability $(1-\epsilon)p_1(x)/\{(1-\epsilon)p_1(x) + \epsilon q(x)\}$ to take x. For the continuous component, the probability density function at x_0 is $q(x)\epsilon p_1(x_0)/\{(1-\epsilon)p_1(x) + \epsilon q(x)\}$. Define $F_{x,y}(x_0) = p(x_0, y)/\{p_1(x_0)p_2(y)\}$. If $p_1(x_0) = 0$ or $p_2(y) = 0$, stipulate $F_{x,y}(x_0) = 1$.

Now we have

$$\begin{aligned}
&\text{CMI}(X, Y \mid \mathcal{S}_1 = s_1) - \text{CMI}(X^\epsilon, Y \mid \mathcal{S}_1 = s_1) \\
&= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \{(1-\epsilon)p_1(x) + \epsilon q(x)\}p_2(y)\Big[\int_{-\infty}^{\infty} F_{x,y}(x_0) \log F_{x,y}(x_0) \mathrm{d}\mu_{x,y}(x_0) \\
&- \Big\{\int_{-\infty}^{\infty} F_{x,y}(x_0)\mathrm{d}\mu_{x,y}(x_0)\Big\} \log \Big\{\int_{-\infty}^{\infty} F_{x,y}(x_0)\mathrm{d}\mu_{x,y}(x_0)\Big\}\Big] \mathrm{d}x\mathrm{d}y \geq 0.
\end{aligned}$$

The last step is the probabilistic form of Jensen's inequality, since $F_{x,y}(x_0)$ is non-negative and integrable with probability measure $\mu_{x,y}(x_0)$, $\{(1-\epsilon)p_1(x) + \epsilon q(x)\}p_2(y) > 0$ if $p_2(y) > 0$, and $f(x) = x \log x$ is strictly convex down when $x \geq 0$ (stipulate $0 \log 0 = 0$).

The equality holds if and only if for $p_1(x_0) > 0$ and $p_2(y) > 0$, $F_{x,y}(x_0)$ is a constant with x_0, which means $p(x_0, y)/p_1(x_0)$ is a constant almost surely. Since $\int_{-\infty}^{\infty} p_1(x_0)p(x_0, y)/p_1(x_0)\mathrm{d}x_0 = p_2(y)$, $\int_{-\infty}^{\infty} p_1(x_0) = 1$, we have $p(x_0, y)/p_1(x_0) = p_2(y)$ for almost surely each x_0, y such that $p_1(x_0) > 0$ and $p_2(y) > 0$. This is equivalent with that X and Y are independent conditioned on $\mathcal{S}_1 = s_1$.

$\textsc{cmi}(X, Y \mid \mathcal{S}_1) = 0$ if and only if X and Y are independent conditioned on any possible value of $\mathcal{S}_1$, except a zero-measure set. Therefore, $\textsc{cmi}(X^\epsilon, Y \mid \mathcal{S}_1) \leq \textsc{cmi}(X, Y \mid \mathcal{S}_1)$, and the equality holds if and only if $\textsc{cmi}(X, Y \mid \mathcal{S}_1) = 0$.

S.6. **Proof of Lemma 4.** Set $\mathcal{S} = \{X, Z_1, \ldots, Z_k\}$. Remember that a Markov boundary $\mathcal{M}$ is a minimal subset of $\mathcal{S}$ such that $\textsc{mi}(\mathcal{M}, Y) = \textsc{mi}(\mathcal{S}, Y)$. Denote $\mathcal{S}$ with ϵ-noise on $Z_i \notin \mathcal{M}_0$ by $\mathcal{S}^\epsilon$. Since $\textsc{mi}(\mathcal{M}_0, Y) = \textsc{mi}(\mathcal{S}, Y)$, $\textsc{mi}(\mathcal{M}_0, Y) \leq \textsc{mi}(\mathcal{S}^\epsilon, Y)$, $\textsc{mi}(\mathcal{S}^\epsilon, Y) \leq \textsc{mi}(\mathcal{S}, Y)$, we have $\textsc{mi}(\mathcal{S}^\epsilon, Y) = \textsc{mi}(\mathcal{S}, Y)$. Therefore, $\mathcal{M}_0$ is still a Markov boundary after adding ϵ-noise. Assume in the new distribution, there is another Markov boundary, then it contains a variable with ϵ-noise: Z_i^ϵ. Denote this Markov boundary by $\{Z_i^\epsilon\} \cup \mathcal{S}_1$. Therefore, $\textsc{cmi}(Z_i^\epsilon, Y \mid \mathcal{S}_1) > 0$. However, from Lemma 3, this implies $\textsc{cmi}(Z_i^\epsilon, Y \mid \mathcal{S}_1) < \textsc{cmi}(Z_i, Y \mid \mathcal{S}_1)$, namely $\textsc{mi}(\{Z_i^\epsilon\} \cup \mathcal{S}_1, Y) < \textsc{mi}(\{Z_i\} \cup \mathcal{S}_1, Y)$. But $\textsc{mi}(\{Z_i^\epsilon\} \cup \mathcal{S}_1, Y) = \textsc{mi}(\mathcal{S}^\epsilon, Y) = \textsc{mi}(\mathcal{S}, Y) \geq \textsc{mi}(\{Z_i\} \cup \mathcal{S}_1, Y)$, which is a contradiction.

S.7. **Proof of Lemma 5.** Assume there exists a Markov boundary $\mathcal{M}$ such that $W \in \mathcal{E}$, $W \notin \mathcal{M}$. Then $\mathcal{S} \setminus \{W\} \supset \mathcal{M}$ is a Markov blanket (Proposition S1), and $\textsc{cmi}(Y, \mathcal{S} \mid \mathcal{S} \setminus \{W\}) = 0$, which contradicts to $W \in \mathcal{E}$.

If $W \notin \mathcal{E}$, then $\textsc{cmi}(Y, \mathcal{S} \mid \mathcal{S} \setminus \{W\}) = 0$, and $\mathcal{S} \setminus \{W\}$ is a Markov blanket. This Markov blanket contains a Markov boundary, which does not contain W.

S.8. **Proof of Theorem 2.** If Markov boundary is unique, then $\mathcal{E}$ is just the Markov boundary, therefore $\textsc{cmi}(Y, \mathcal{S} \mid \mathcal{E}) = 0$.

If $\textsc{cmi}(Y, \mathcal{S} \mid \mathcal{E}) = 0$, then $\mathcal{E}$ is a Markov blanket, which means it should contain a Markov boundary. But $\mathcal{E}$ should be contained in every Markov boundary, therefore $\mathcal{E}$ itself is a Markov boundary. $\mathcal{E}$ as a Markov boundary cannot be a proper subset of another Markov boundary, thus the only Markov boundary is $\mathcal{E}$.

S.9. **Proof of Proposition 5.** *Proof that Algorithm 1 is sound and complete.* There exists at least one Markov boundary. The algorithm can always terminate in finite steps and produce an output. It is easy to see that the output $\mathcal{M}_0$ is a Markov blanket. In the last step of Algorithm 1, we have checked that $X_0 \not\perp\!\!\!\perp Y \mid \mathcal{M}_0 \setminus \{X_0\}$. For $X_i \in \mathcal{M}_0$, since $\Delta(X_i, Y \mid \mathcal{M}_0 \setminus \{X_i\}) \geq \Delta(X_0, Y \mid \mathcal{M}_0 \setminus \{X_0\})$, we also have $X_i \not\perp\!\!\!\perp Y \mid \mathcal{M}_0 \setminus \{X_i\}$. Therefore the output of Algorithm 1 is a Markov boundary.

Proof that Algorithm 2 is sound and complete. The algorithm can always terminate in finite steps and produce an output. Markov boundary $\mathcal{M}_0$ is not the unique Markov boundary if and only if there exists variable $X_i \in \mathcal{M}_0$ which is not essential, namely

$$\textsc{mi}(Y, \mathcal{S} \setminus \{X_i\}) = \textsc{mi}(Y, \mathcal{S}).$$

Moreover, since $\textsc{mi}(Y, \mathcal{S} \setminus \{X_i\}) = \textsc{mi}(Y, \mathcal{M}_i)$ and $\textsc{mi}(Y, \mathcal{M}_0) = \textsc{mi}(Y, \mathcal{S})$, we have

$$\textsc{mi}(Y, \mathcal{M}_i) = \textsc{mi}(Y, \mathcal{M}_0),$$

or equivalently,

$$\textsc{cmi}(Y, \mathcal{M}_0 \mid \mathcal{M}_i) = 0.$$

S.10. **Algorithms references in Remarks 5 and 6.** We now describe Algorithms S1 and S2 that were used in the simulation studies.

Algorithm S1 is obtained by replacing step (3) in Algorithm 2 with a direct test of whether X_i is an essential variable.

Algorithm S1: A variant of Algorithm 2 for testing the uniqueness of Markov boundary

---

(1) **Input**
    Joint distribution of $\mathcal{S} = \{X_1, \ldots, X_k\}$ and Y
(2) **Set** $\mathcal{M}_0 = \{X_1, \ldots, X_m\}$ to be the result of Algorithm 1 on $\mathcal{S}$
(3) **For** $i = 1, \ldots, m$,
    If $X_i \perp\!\!\!\perp Y \mid \mathcal{S} \setminus \{X_i\}$
        Output Y has multiple Markov boundaries
        Terminate
(4) **Output** Y has a unique Markov boundary

Proof of correctness of Algorithm S1. For a Markov boundary $\mathcal{M}_0$, it is the unique Markov boundary if and only if it coincides with $\mathcal{E}$. Therefore, we only need to check whether there exists a variable $X_i \in \mathcal{M}_0$ which is not essential, namely $X_i \perp\!\!\!\perp Y \mid \mathcal{S} \setminus \{X_i\}$.

Algorithm S2 is constructed based on Theorem 2 directly.

References

J. Pearl. *Probabilistic Inference in Intelligent Systems.* Morgan Kaufmann, San Mateo, 1988.

Algorithm S2: A benchmark algorithm for testing the uniqueness of Markov boundary based on Theorem 2

(1) **Input**
    Joint distribution of $\mathcal{S} = \{X_1, \ldots, X_k\}$ and Y
(2) **Set** $\hat{\mathcal{E}} = \emptyset$
(3) **For** $i = 1, \ldots, k$,
    If $X_i \not\perp\!\!\!\perp Y \mid \mathcal{S} \setminus \{X_i\}$
        $\hat{\mathcal{E}} = \hat{\mathcal{E}} \cup \{X_i\}$
(4) **If** $Y \perp\!\!\!\perp \mathcal{S} \mid \hat{\mathcal{E}}$
    Output: Y has a unique Markov boundary
  Else
    Output: Y has multiple Markov boundaries