VI. Appendix

A. Proof for Lemma 2

By definition of Kullback-Leibler divergence we have

$$
\begin{aligned}
&D(Q\|R)\\
=&-\sum_{t\in T} Q(t)\ln\frac{R(t)}{Q(t)} && (25)\\
=&-\sum_{t\in T} Q(t)\cdot 2\ln\sqrt{\frac{R(t)}{Q(t)}} && (26)\\
=&-2\sum_{t\in T} Q_t\Big[\Big(\sqrt{\frac{R(t)}{Q(t)}}-1\Big)-\frac{1}{2}\Big(\sqrt{\frac{R(t)}{Q(t)}}-1\Big)^2\\
&\qquad+o\Big(\Big(\sqrt{\frac{R(t)}{Q(t)}}-1\Big)^2\Big)\Big] && (27)\\
=&2-2\sum_{t\in T}\sqrt{Q(t)R(t)}+\sum_{t\in T}\left(\sqrt{R(t)}-\sqrt{Q(t)}\right)^2\\
&\qquad+o\left(H^2(Q_T,R_T)\right) && (28)\\
=&4H^2(Q_T,R_T)+o\left(H^2(Q_T,R_T)\right) && (29)
\end{aligned}
$$

where (27) follows from $\ln(1+x)=x-\frac{1}{2}x^2+o(x^2)$, while (28) follows from $\sum_{t\in T}Q(t)=\sum_{t\in T}R(t)=1$. The last line follows from the definition of Hellinger distance.

B. Proof for Lemma 4

By definition (10) the coefficient is given by $\alpha = F(8n_z D\rho)\frac{D}{\rho}$ where

$$F(x)\triangleq\frac{I_{\frac{|\mathcal{X}||\mathcal{Y}|}{2}}(x)}{I_{\frac{|\mathcal{X}||\mathcal{Y}|}{2}-1}(x)}. \quad (30)$$

For the bound to be hold, it suffices to prove that the transcendental function $F(x)$ can be bounded by the following quadratic rational functions

$$L(x)<F(x)<U(x) \quad (31)$$

where

$$L(x)=\frac{|\mathcal{X}||\mathcal{Y}|x+2x^2}{|\mathcal{X}|^2|\mathcal{Y}|^2+2|\mathcal{X}||\mathcal{Y}|x+2x^2}$$

and

$$U(x)=\frac{2|\mathcal{X}||\mathcal{Y}|x+x^2}{2|\mathcal{X}|^2|\mathcal{Y}|^2+|\mathcal{X}||\mathcal{Y}|x+x^2}$$

In order to bound

$$F(x)=\frac{I_{\frac{|\mathcal{X}||\mathcal{Y}|}{2}}(x)}{I_{\frac{|\mathcal{X}||\mathcal{Y}|}{2}-1}(x)},$$

we generalize the problem with bounding $f_p(x)=\frac{I_p(x)}{I_{p-1}(x)}$ and letting $p=\frac{|\mathcal{X}||\mathcal{Y}|}{2}$ to give the result.

By definition of first type of modified Bessel function, we have $I_p(x)=\sum_{n=0}^{\infty}\frac{x^{2n+p}}{2^{2n+p}n!\Gamma(p+n+1)}$. Thus

$$
\begin{aligned}
f_p'(x)&=\frac{I_p'(x)I_{p-1}(x)-I_p(x)I_{p-1}'(x)}{I_{p-1}^2(x)} && (32)\\
&=\frac{\sum_{n=0}^{\infty}\frac{(2n+p)x^{2n+p-1}}{2^{2n+p}n!\Gamma(p+n+1)}}{I_{p-1}(x)}\\
&\quad-\frac{I_p(x)\sum_{n=0}^{\infty}\frac{(2n+p-1)x^{2n+p-2}}{2^{2n+p-1}n!\Gamma(p+n)}}{I_{p-1}^2(x)} && (33)\\
&=\frac{(I_{p-1}(x)-\frac{p}{x}I_p(x))}{I_{p-1}(x)} && (34)\\
&\quad-\frac{I_p(x)(I_p(x)+\frac{p-1}{x}I_{p-1}(x))}{I_{p-1}^2(x)} && (35)\\
&=1-\frac{2p-1}{x}f_p(x)-f_p^2(x) && (36)
\end{aligned}
$$

Define

$$L_p(x)=\frac{px+x^2}{2p^2+2px+x^2},$$

and

$$U_p(x)=\frac{4px+x^2}{8p^2+2px+x^2}$$

We have $U_p(0)=0=f_p(0)$ and

$$
\begin{aligned}
&U_p'(x)-(1-\frac{2p-1}{x}U_p(x)-U_p^2(x))\\
=&\frac{x^2(8p(p-1)+(6p-1)x)}{(8p^2+2px+x^2)^2}>0
\end{aligned}
$$

thus $f_p(x)<U_p(X)$.

We also have

$$L_p(0)=0=f_p(0)$$

and

$$
\begin{aligned}
&L_p'(x)-(1-\frac{2p-1}{x}L_p(x)-L_p^2(x))\\
=&-\frac{x^2(p^2+2p+x^2)}{(2p^2+2px+x^2)^2}<0
\end{aligned}
$$

thus $L_p(x)<f_p(x)$.

C. Proof for Lemma 6

In the following proof, we abuse the notation of $P_{XY|Z}(x,y|z)$ as $P(x,y|z)$. To prove equation (14), we first illustrate that

$$
\begin{aligned}
&\left|\mathbb{E}[[H(\tilde{P}^*_{XY|Z},P_{XY|Z})]^2]-\left(\frac{\alpha_0^2|\mathcal{X}||\mathcal{Y}|}{8n_z}+(1-\alpha_0)^2\rho_0^2\right)\right|\\
\le&\sqrt{\mathbb{E}\left[H^2\left(\tilde{P}^*_{XY|Z},P_{XY|Z}\right)\right]}\frac{8\sqrt{n_z|\mathcal{X}||\mathcal{Y}|}D^2}{|\mathcal{X}||\mathcal{Y}|}\\
&+\left(\frac{8\sqrt{n_z|\mathcal{X}||\mathcal{Y}|}D^2}{|\mathcal{X}||\mathcal{Y}|}\right)^2, && (37)
\end{aligned}
$$

where $\rho_0 \triangleq H(P_{XY|Z}, P^{(\mathrm{M})}_{XY|Z})$ and $\alpha_0 \triangleq F(8n_z D\rho_0)\frac{D}{\rho_0}$. To this end, note that from the definitions of Hellinger distance and the proposed estimator $\tilde{P}^*_{XY|Z}$, we have

$$
\begin{aligned}
&\mathbb{E}\big[[H(\tilde{P}^*_{XY|Z}, P_{XY|Z})]^2\big] \\
=&\frac{1}{2}\mathbb{E}\Big[\sum_{x,y}\Big(\sqrt{\tilde{P}^*(x,y|z)}-\sqrt{P(x,y|z)}\Big)^2\Big] \qquad (38)\\
=&\frac{1}{2}\mathbb{E}\Big[\sum_{x,y}\Big(\alpha_0\sqrt{\hat{P}(x,y|z)}+(1-\alpha_0)\sqrt{P^{(\mathrm{M})}(x,y|z)} \\
&\qquad -\sqrt{P(x,y|z)}\Big)^2\Big] \qquad (39)\\
&+\frac{1}{2}\mathbb{E}\Big[\sum_{x,y}\Big(\alpha\sqrt{\hat{P}(x,y|z)}+(1-\alpha)\sqrt{P^{(\mathrm{M})}(x,y|z)} \\
&\qquad -\sqrt{P(x,y|z)}\Big)^2\Big] \qquad (40)\\
&-\frac{1}{2}\mathbb{E}\Big[\sum_{x,y}\Big(\alpha_0\sqrt{\hat{P}(x,y|z)}+(1-\alpha_0)\sqrt{P^{(\mathrm{M})}(x,y|z)} \\
&\qquad -\sqrt{P(x,y|z)}\Big)^2\Big]. \qquad (41)
\end{aligned}
$$

For (39), we have

$$
\begin{aligned}
&\frac{1}{2}\mathbb{E}\Big[\sum_{x,y}\Big(\alpha_0\sqrt{\hat{P}(x,y|z)}+(1-\alpha_0)\sqrt{P^{(\mathrm{M})}(x,y|z)} \\
&\qquad -\sqrt{P(x,y|z)}\Big)^2\Big] \\
=&\frac{\alpha_0^2}{2}\mathbb{E}\Big[\sum_{x,y}\Big(\sqrt{\hat{P}(x,y|z)}-\sqrt{P(x,y|z)}\Big)^2\Big] \qquad (42)\\
&+\alpha_0(1-\alpha_0)\mathbb{E}\Big[\sum_{x,y}\Big(\sqrt{\hat{P}(x,y|z)}-\sqrt{P(x,y|z)}\Big) \\
&\quad\cdot\Big(\sqrt{P^{(\mathrm{M})}(x,y|z)}-\sqrt{P(x,y|z)}\Big)\Big] \qquad (43)\\
&+\frac{(1-\alpha_0)^2}{2}\mathbb{E}\Big[\sum_{x,y}\Big(\sqrt{P^{(\mathrm{M})}(x,y|z)}-\sqrt{P(x,y|z)}\Big)^2\Big]. \qquad (44)
\end{aligned}
$$

Then we compute (42)-(44). For (42) and (43), since $\sqrt{\hat{P}_{XY|Z}}$ follows Gaussian distribution, we can immediately have that

$$
\mathbb{E}\Big[\sum_{x,y}\Big(\sqrt{\hat{P}(x,y|z)}-\sqrt{P(x,y|z)}\Big)^2\Big]=\frac{|\mathcal{X}||\mathcal{Y}|}{4n_z}, \qquad (45)
$$

$$
\begin{aligned}
&\mathbb{E}\Big[\sum_{x,y}\Big(\sqrt{\hat{P}(x,y|z)}-\sqrt{P(x,y|z)}\Big) \\
&\quad\cdot\Big(\sqrt{P^{(\mathrm{M})}(x,y|z)}-\sqrt{P(x,y|z)}\Big)\Big]=0. \qquad (46)
\end{aligned}
$$

And for (44), we have that

$$
\frac{1}{2}\mathbb{E}\Big[\sum_{x,y}\Big(\sqrt{P^{(\mathrm{M})}(x,y|z)}-\sqrt{P(x,y|z)}\Big)^2\Big]=\rho_0^2. \qquad (47)
$$

While for (40) and (41), we have

$$
\begin{aligned}
&\Big|\frac{1}{2}\mathbb{E}\Big[\sum_{x,y}\Big(\alpha\sqrt{\hat{P}(x,y|z)}+(1-\alpha)\sqrt{P^{(\mathrm{M})}(x,y|z)} \\
&\qquad -\sqrt{P(x,y|z)}\Big)^2\Big] \\
&-\frac{1}{2}\mathbb{E}\Big[\sum_{x,y}\Big(\alpha_0\sqrt{\hat{P}(x,y|z)}+(1-\alpha_0)\sqrt{P^{(\mathrm{M})}(x,y|z)} \\
&\qquad -\sqrt{P(x,y|z)}\Big)^2\Big]\Big| \\
=&\Big|\frac{1}{2}\mathbb{E}\Big[\sum_{x,y}\Big((\alpha+\alpha_0)\sqrt{\hat{P}(x,y|z)} \\
&\qquad +(2-\alpha-\alpha_0)\sqrt{P^{(\mathrm{M})}(x,y|z)} \\
&\qquad -2\sqrt{P(x,y|z)}\Big)\cdot\Big(\alpha-\alpha_0\Big) \\
&\quad\Big(\sqrt{\hat{P}(x,y|z)}-\sqrt{P^{(\mathrm{M})}(x,y|z)}\Big)\Big]\Big| \qquad (48)\\
=&\frac{1}{2}\Big|\mathbb{E}\Big[\sum_{x,y}\Big[\Big(2\alpha\sqrt{\hat{P}(x,y|z)}+(2-2\alpha)\sqrt{P^{(\mathrm{M})}(x,y|z)} \\
&\quad -2\sqrt{P(x,y|z)}\Big) \\
&\quad -\Big(\alpha-\alpha_0\Big)\Big(\sqrt{\hat{P}(x,y|z)}-\sqrt{P^{(\mathrm{M})}(x,y|z)}\Big)\Big] \\
&\quad\cdot\Big(\alpha-\alpha_0\Big)\Big(\sqrt{\hat{P}(x,y|z)}-\sqrt{P^{(\mathrm{M})}(x,y|z)}\Big)\Big]\Big| \qquad (49)\\
\le&\Big|\mathbb{E}\Big[\sum_{x,y}\Big(\sqrt{\tilde{P}^*(x,y|z)}-\sqrt{P(x,y|z)}\Big) \\
&\quad\cdot\Big(\alpha-\alpha_0\Big)\Big(\sqrt{\hat{P}(x,y|z)}-\sqrt{P^{(\mathrm{M})}(x,y|z)}\Big)\Big]\Big| \qquad (50)\\
&+\frac{1}{2}\mathbb{E}\Big[\Big(\alpha-\alpha_0\Big)^2\Big(\sqrt{\hat{P}(x,y|z)}-\sqrt{P^{(\mathrm{M})}(x,y|z)}\Big)^2\Big] \qquad (51)
\end{aligned}
$$

As $0<\alpha_0<\frac{8n_z D^2}{|\mathcal{X}||\mathcal{Y}|}$ and so does α, as well as $\mathbb{E}\Big[\sum_{x,y}\Big(\sqrt{\hat{P}(x,y|z)}-\sqrt{P^{(\mathrm{M})}(x,y|z)}\Big)^2\Big]=2D^2+\frac{|\mathcal{X}||\mathcal{Y}|}{4n_z}$, we have

$$
\begin{aligned}
&\Big|\mathbb{E}\Big[\sum_{x,y}\Big(\sqrt{\tilde{P}^*(x,y|z)}-\sqrt{P(x,y|z)}\Big) \\
&\quad\cdot\Big(\alpha-\alpha_0\Big)\Big(\sqrt{\hat{P}(x,y|z)}-\sqrt{P^{(\mathrm{M})}(x,y|z)}\Big)\Big]\Big| \\
\le&\sqrt{\mathbb{E}\Big[H^2\Big(\tilde{P}^*_{XY|Z},P_{XY|Z}\Big)\Big]}\frac{8\sqrt{n_z|\mathcal{X}||\mathcal{Y}|}D^2}{|\mathcal{X}||\mathcal{Y}|}. \qquad (52)
\end{aligned}
$$

and

$$
\begin{aligned}
&\Big|\mathbb{E}\Big[\sum_{x,y}\Big(\alpha-\alpha_0\Big)^2\Big(\sqrt{\hat{P}(x,y|z)}-\sqrt{P^{(\mathrm{M})}(x,y|z)}\Big)^2\Big]\Big| \\
\le&\Big(\frac{8\sqrt{n_z|\mathcal{X}||\mathcal{Y}|}D^2}{|\mathcal{X}||\mathcal{Y}|}\Big)^2. \qquad (53)
\end{aligned}
$$

Finally, we have inequality (37) from (39)-(47) and (50)-(53).

As the maximum of $\mathbb{E}[H^2(\tilde{P}^*_{XY|Z}, P_{XY|Z})]$ is achieved if and only if $\rho_0 = D$. This implies $\alpha_0 = F(8n_z D^2)$in which the definition of F is referred to definition (30), we have

$$\begin{aligned}
&\Big|\max_{P_{XY|Z}\in U}\mathbb{E}[H^2(\tilde{P}^*_{XY|Z}, P_{XY|Z})] \\
&\quad -\Big(\frac{A^2|\mathcal{X}||\mathcal{Y}|}{8n_z}+(1-A)^2D^2\Big)\Big| \\
&\le\sqrt{\max_{P_{XY|Z}\in U}\mathbb{E}\Big[H^2\Big(\tilde{P}^*_{XY|Z}, P_{XY|Z}\Big)\Big]}\frac{8\sqrt{n_z|\mathcal{X}||\mathcal{Y}|}D^2}{|\mathcal{X}||\mathcal{Y}|} \\
&\quad +\Big(\frac{8\sqrt{n_z|\mathcal{X}||\mathcal{Y}|}D^2}{|\mathcal{X}||\mathcal{Y}|}\Big)^2, \qquad (54)
\end{aligned}$$

where $U \triangleq \{P_{XY|Z}|H(P_{XY|Z}, P^{(\mathrm{M})}_{XY|Z}) \le D\}$.

It follows that

$$\begin{aligned}
&\Big|\frac{1}{D^2}\max_{P_{XY|Z}\in U}\mathbb{E}\Big[H^2(\tilde{P}^*_{XY|Z}, P_{XY|Z})\Big] \\
&\quad -\Big(\frac{|\mathcal{X}||\mathcal{Y}|A^2}{8n_zD^2}+(1-A)^2\Big)\Big| \\
&\le\frac{8\sqrt{n_z|\mathcal{X}||\mathcal{Y}|}D}{|\mathcal{X}||\mathcal{Y}|}\frac{\sqrt{\max_{P_{XY|Z}\in U}\mathbb{E}\Big[H^2(\tilde{P}^*_{XY|Z}, P_{XY|Z})\Big]}}{D} \\
&\quad +\Big(\frac{8\sqrt{n_z|\mathcal{X}||\mathcal{Y}|}D}{|\mathcal{X}||\mathcal{Y}|}\Big)^2.
\end{aligned}$$

Let

$$\frac{1}{D}\sqrt{\max_{P_{XY|Z}\in U}\mathbb{E}\Big[H^2(\tilde{P}^*_{XY|Z}, P_{XY|Z})\Big]} = t$$

we have

$$|t^2-k_1| \le k_2 t + k_3$$

where

$$k_1 = \frac{|\mathcal{X}||\mathcal{Y}|A^2}{8n_zD^2}+(1-\alpha)^2,$$

$$k_2 = \frac{8\sqrt{n_z|\mathcal{X}||\mathcal{Y}|}D}{|\mathcal{X}||\mathcal{Y}|}$$

and

$$k_3 = \Big(\frac{8\sqrt{n_z|\mathcal{X}||\mathcal{Y}|}D}{|\mathcal{X}||\mathcal{Y}|}\Big)^2.$$

Note that $t > 0$ and $k_i > 0, i = 1, 2, 3$. From

$$t^2 - k_1 \le k_2 t + k_3$$

we have

$$\begin{aligned}
t &\le \frac{k_2+\sqrt{k_2^2+4(k_1+k_3)}}{2} && (55)\\
&= k_1+k_3+\frac{k_2}{2}+\frac{\sqrt{k_2^2+4(k_1+k_3)}-2\sqrt{k_1+k_3}}{2} && (56)\\
&= k_1+k_3+\frac{k_2}{2}+\frac{k_2^2}{2(\sqrt{k_2^2+4(k_1+k_3)}+2\sqrt{k_1+k_3})} && (57)\\
&\le k_1+k_3+\frac{k_2}{2}+\frac{k_2^2}{8\sqrt{k_1+k_3}} && (58)
\end{aligned}$$

As $1-16\eta \le k_1 \le 1+8\eta$, $0 \le k_2 \le 8\sqrt{\eta}$ and $0 \le k_3 \le 64\eta$ we have $\exists C > 0$ s.t. $t \le 1 + C\sqrt{\eta}$ While from $t^2 - k_1 \ge -k_2 t - k_3$ we have

$$\begin{aligned}
t &\ge \frac{-k_2+\sqrt{k_2^2+4(k_1-k_3)}}{2} && (59)\\
&= k_1-k_3-\frac{k_2}{2}+\frac{\sqrt{k_2^2+4(k_1-k_3)}-2(k_1-k_3)}{2} && (60)\\
&\ge k_1-k_3-\frac{k_2}{2} && (61)
\end{aligned}$$

As $1-16\eta \le k_1 \le 1+8\eta$, $0 \le k_2 \le 8\sqrt{\eta}$ and $0 \le k_3 \le 64\eta$ we have $\exists C > 0$ s.t. $t \ge 1 - C\sqrt{\eta}$ i.e.

$$\Big|\frac{1}{D^2}\max_{P_{XY|Z}\in U}\mathbb{E}[H^2(\tilde{P}^*_{XY|Z}, P_{XY|Z})]-1\Big| \le C\sqrt{\eta} \qquad (62)$$

D. Proof for Lemma 7

It suffices to decompose the gap $\mathbb{E}[\big[H(\bar{P}_{XY|Z}, P_{XY|Z})\big]^2] - \mathbb{E}[\big[H(\tilde{P}^*_{XY|Z}, P_{XY|Z})\big]^2]$ into (65)-(68), which is bounded as (75) and (99) respectively.

$$\begin{aligned}
&\mathbb{E}[\big[H(\bar{P}_{XY|Z}, P_{XY|Z})\big]^2] - \mathbb{E}[\big[H(\tilde{P}^*_{XY|Z}, P_{XY|Z})\big]^2] \\
=&\frac{1}{2}\mathbb{E}\Big[\sum_{x,y}\Big(\sqrt{\bar{P}(x,y|z)}-\sqrt{P(x,y|z)}\Big)^2\Big] && (63)\\
&-\frac{1}{2}\mathbb{E}\Big[\sum_{x,y}\Big(\sqrt{\tilde{P}^*(x,y|z)}-\sqrt{P(x,y|z)}\Big)^2\Big] && (64)\\
=&\frac{1}{2}\mathbb{E}\Big[\sum_{x,y}\Big(\alpha\sqrt{\hat{P}(x,y|z)}+(1-\alpha)\sqrt{P^{(\mathrm{M})}(x,y|z)} \\
&\qquad +(1-\alpha)(\hat{P}^{(\mathrm{M})}(x.y|z)-P^{(\mathrm{M})}(x.y|z)) \\
&\qquad -\sqrt{P(x,y|z)}\Big)^2\Big] && (65)\\
&-\frac{1}{2}\mathbb{E}\Big[\sum_{x,y}\Big(\alpha\sqrt{\hat{P}(x,y|z)}+(1-\alpha)\sqrt{P^{(\mathrm{M})}(x,y|z)} \\
&\qquad -\sqrt{P(x,y|z)}\Big)^2\Big] && (66)\\
&+\frac{1}{2}\mathbb{E}\Big[\sum_{x,y}\Big(\alpha'\sqrt{\hat{P}(x,y|z)}+(1-\alpha')\sqrt{\hat{P}^{(\mathrm{M})}(x,y|z)} \\
&\qquad -\sqrt{P(x,y|z)}\Big)^2\Big] && (67)\\
&-\frac{1}{2}\mathbb{E}\Big[\sum_{x,y}\Big(\alpha\sqrt{\hat{P}(x,y|z)}+(1-\alpha)\sqrt{\hat{P}^{(\mathrm{M})}(x,y|z)} \\
&\qquad -\sqrt{P(x,y|z)}\Big)^2\Big] && (68)
\end{aligned}$$

For (65) and (66), we have

$$
\begin{aligned}
&\frac{1}{2}\mathbb{E}\Big[\sum_{x,y}\Big(\alpha\sqrt{\hat{P}(x,y|z)}+(1-\alpha)\sqrt{P^{(\mathrm{M})}(x,y|z)}\\
&\qquad\quad+(1-\alpha)(\hat{P}^{(\mathrm{M})}(x.y|z)-P^{(\mathrm{M})}(x.y|z))\\
&\qquad\quad-\sqrt{P(x,y|z)}\Big)^2\Big]\\
&-\frac{1}{2}\mathbb{E}\Big[\sum_{x,y}\Big(\alpha\sqrt{\hat{P}(x,y|z)}+(1-\alpha)\sqrt{P^{(\mathrm{M})}(x,y|z)}\\
&\qquad\quad-\sqrt{P(x,y|z)}\Big)^2\Big]\\
=&\mathbb{E}\Big[\alpha(1-\alpha)\sum_{x,y}\Big(\sqrt{\hat{P}(x,y|z)}-\sqrt{P(x,y|z)}\Big)\\
&\quad\cdot\Big(\sqrt{\hat{P}^{(\mathrm{M})}(x,y|z)}-\sqrt{P^{(\mathrm{M})}(x,y|z)}\Big)\Big] \qquad (69)\\
&+\mathbb{E}\Big[(1-\alpha)^2\sum_{x,y}\Big(\sqrt{P^{(\mathrm{M})}(x,y|z)}-\sqrt{P(x,y|z)}\Big)\\
&\quad\cdot\Big(\sqrt{\hat{P}^{(\mathrm{M})}(x,y|z)}-\sqrt{P^{(\mathrm{M})}(x,y|z)}\Big)\Big] \qquad (70)\\
&+\frac{1}{2}\mathbb{E}\Big[(1-\alpha)^2\sum_{x,y}\Big(\sqrt{\hat{P}^{(\mathrm{M})}(x,y|z)}-\sqrt{P^{(\mathrm{M})}(x,y|z)}\Big)^2\Big] \qquad (71)
\end{aligned}
$$

Nothe the Corollary 5 gives that $0<\alpha<\frac{8n_zD^2}{|\mathcal{X}||\mathcal{Y}|}$, to give bound for (69)-(71) it suffices to prove

$$
\begin{aligned}
&\mathbb{E}\Big[\sum_{x,y}\Big(\sqrt{\hat{P}(x,y|z)}-\sqrt{P(x,y|z)}\Big)\\
&\quad\cdot\Big(\sqrt{\hat{P}^{(\mathrm{M})}(x,y|z)}-\sqrt{P^{(\mathrm{M})}(x,y|z)}\Big)\Big]\le\frac{(|\mathcal{X}||\mathcal{Y}|)^{\frac{3}{2}}}{\sqrt{2}n}, \qquad (72)
\end{aligned}
$$

$$
\begin{aligned}
&\mathbb{E}\Big[\sum_{x,y}\Big(\sqrt{P^{(\mathrm{M})}(x,y|z)}-\sqrt{P(x,y|z)}\Big)\\
&\quad\cdot\Big(\sqrt{\hat{P}^{(\mathrm{M})}(x,y|z)}-\sqrt{P^{(\mathrm{M})}(x,y|z)}\Big)\Big]\le\frac{\sqrt{2|\mathcal{X}||\mathcal{Y}|}D}{2n} \qquad (73)
\end{aligned}
$$

and

$$
\mathbb{E}\Big[\sum_{x,y}\Big(\sqrt{\hat{P}^{(\mathrm{M})}(x,y|z)}-\sqrt{P^{(\mathrm{M})}(x,y|z)}\Big)^2\Big]\le\frac{\sqrt{|\mathcal{X}||\mathcal{Y}|}}{2n}, \qquad (74)
$$

then it follows that

$$
\begin{aligned}
&\frac{1}{2}\mathbb{E}\Big[\sum_{x,y}\Big(\alpha\sqrt{\hat{P}(x,y|z)}+(1-\alpha)\sqrt{P^{(\mathrm{M})}(x,y|z)}\\
&\qquad+(1-\alpha)(\hat{P}^{(\mathrm{M})}(x.y|z)-P^{(\mathrm{M})}(x.y|z))\\
&\qquad-\sqrt{P(x,y|z)}\Big)^2\Big]\\
&-\frac{1}{2}\mathbb{E}\Big[\sum_{x,y}\Big(\alpha\sqrt{\hat{P}(x,y|z)}+(1-\alpha)\sqrt{P^{(\mathrm{M})}(x,y|z)}\\
&\qquad-\sqrt{P(x,y|z)}\Big)^2\Big]\\
&\le\frac{4\sqrt{2}D^2}{\sqrt[4]{|\mathcal{X}||\mathcal{Y}|}}+\Big(\frac{\sqrt{2}}{2n_z}D+\frac{1}{2n_z}\Big)\sqrt{|\mathcal{X}||\mathcal{Y}|} \qquad (75)
\end{aligned}
$$

To prove inequatlities (72)-(74), we start from (74). Then the result is used to prove (72) utilizing Cauchy inequality.

Since $\sqrt{\hat{P}_{XY|Z}}$ follows Gaussian distribution, we can immediately have that

$$
\mathbb{E}\Big[\sum_{x,y}\Big(\sqrt{\hat{P}(x,y|z)}-\sqrt{P(x,y|z)}\Big)^2\Big]=\frac{|\mathcal{X}||\mathcal{Y}|}{4n_z},. \qquad (76)
$$

$$
\mathbb{E}\Big[\sum_{x}\Big(\sqrt{\hat{P}(x|z)}-\sqrt{P(x|z)}\Big)^2\Big]=\frac{|\mathcal{X}|}{4n_z},. \qquad (77)
$$

$$
\mathbb{E}\Big[\sum_{y}\Big(\sqrt{\hat{P}(y|z)}-\sqrt{P(y|z)}\Big)^2\Big]=\frac{|\mathcal{Y}|}{4n_z}. \qquad (78)
$$

Further we have

$$
\begin{aligned}
&\mathbb{E}\Big[\sum_{x,y}\Big(\sqrt{\hat{P}^{(\mathrm{M})}(x,y|z)}-\sqrt{P^{(\mathrm{M})}(x,y|z)}\Big)^2\Big]\\
=&\mathbb{E}\Big[\sum_{x,y}\Big(\hat{P}^{(\mathrm{M})}(x,y|z)+P^{(\mathrm{M})}(x,y|z)\\
&\qquad-2\sqrt{\hat{P}^{(\mathrm{M})}(x,y|z)}\sqrt{P^{(\mathrm{M})}(x,y|z)}\Big)\Big] \qquad (79)\\
=&2-2\mathbb{E}\Big[\sum_{x,y}\sqrt{P(x|z)P(y|z)}\sqrt{\hat{P}(x|z)\hat{P}(y|z)}\Big] \qquad (80)\\
=&2-2\mathbb{E}\Big\{\sum_{x,y}\sqrt{P(x|z)P(y|z)}\\
&\qquad\Big[\Big(\sqrt{\hat{P}(x|z)}-\sqrt{P(x|z)}\Big)\\
&\qquad\quad\cdot\Big(\sqrt{\hat{P}(y|z)}-\sqrt{P(y|z)}\Big)\\
&\qquad+\Big(\sqrt{\hat{P}(x|z)}-\sqrt{P(x|z)}\Big)\cdot\sqrt{P(y|z)}\\
&\qquad+\sqrt{P(x|z)}\cdot\Big(\sqrt{\hat{P}(y|z)}-\sqrt{P(y|z)}\Big)\\
&\qquad+\sqrt{P(x|z)}\cdot\sqrt{P(y|z)}\Big]\Big\} \qquad (81)
\end{aligned}
$$

$$
\begin{aligned}
&= -2\sum_{x,y}\mathbb{E}\Big[\Big(\sqrt{\hat{P}(x|z)}-\sqrt{P(x|z)}\Big) \\
&\qquad \cdot\Big(\sqrt{\hat{P}(y|z)}-\sqrt{P(y|z)}\Big)\Big]\cdot\sqrt{P(x|z)P(y|z)} \qquad (82)\\
&\le 2\sqrt{\sum_{x,y}\Big\{\mathbb{E}\Big[\Big(\sqrt{\hat{P}(x|z)}-\sqrt{P(x|z)}\Big)\cdot\Big(\sqrt{\hat{P}(y|z)}-\sqrt{P(y|z)}\Big)\Big]\Big\}^2}\cdot\sqrt{\sum_{x,y}P(x|z)P(y|z)} \qquad (83)\\
&= 2\sqrt{\sum_{x,y}\Big\{\mathbb{E}\Big[\Big(\sqrt{\hat{P}(x|z)}-\sqrt{P(x|z)}\Big)\cdot\Big(\sqrt{\hat{P}(y|z)}-\sqrt{P(y|z)}\Big)\Big]\Big\}^2} \qquad (84)
\end{aligned}
$$

Where (80) follows from $\sum_{x,y}\hat{P}^{(\mathrm{M})}(x,y|z) = \sum_{x,y}P^{(\mathrm{M})}(x,y|z) = 1$, (82) follows from $\mathbb{E}[\hat{P}(x|z)] = P(x|z)$,$\mathbb{E}[\hat{P}(y|z)] = P(y|z)$ and $\sum_{x,y}P(x|z)P(y|z) = 1$, while by Cauchy inequality we have (83). As $\sum_{x,y}P(x|z)P(y|z) = 1$ we have (84).

By Cauchy inequality we have

$$
\begin{aligned}
&\mathbb{E}\Big[\Big(\sqrt{\hat{P}(x|z)}-\sqrt{P(x|z)}\Big)\Big(\sqrt{\hat{P}(y|z)}-\sqrt{P(y|z)}\Big)\Big] \qquad (85)\\
&\le \sqrt{\mathbb{E}\Big[\Big(\sqrt{\hat{P}(x|z)}-\sqrt{P(x|z)}\Big)^2\Big]}\cdot\sqrt{\mathbb{E}\Big[\Big(\sqrt{\hat{P}(y|z)}-\sqrt{P(y|z)}\Big)^2\Big]} \qquad (86)\\
&\le \frac{1}{4n_z}, \qquad \forall(x,y)\in\mathcal{X}\times\mathcal{Y} \qquad (87)
\end{aligned}
$$

We have bound of (84) as

$$
\begin{aligned}
&\sum_{x,y}\Big\{\mathbb{E}\Big[\Big(\sqrt{\hat{P}(x|z)}-\sqrt{P(x|z)}\Big)\cdot\Big(\sqrt{\hat{P}(y|z)}-\sqrt{P(y|z)}\Big)\Big]\Big\}^2\\
&\le \frac{|\mathcal{X}||\mathcal{Y}|}{16n_z^2} \qquad (88)
\end{aligned}
$$

thus

$$
\mathbb{E}\Big[\sum_{x,y}\Big(\sqrt{\hat{P}^{(\mathrm{M})}(x,y|z)}-\sqrt{P^{(\mathrm{M})}(x,y|z)}\Big)^2\Big] \le \frac{\sqrt{|\mathcal{X}||\mathcal{Y}|}}{2n_z} \qquad (89)
$$

To prove inequality (72), by Cauchy inequality:

$$
\begin{aligned}
&\mathbb{E}\Big[\sum_{x,y}\Big(\sqrt{\hat{P}(x,y|z)}-\sqrt{P(x,y|z)}\Big)\cdot\Big(\sqrt{\hat{P}^{(\mathrm{M})}(x,y|z)}-\sqrt{P^{(\mathrm{M})}(x,y|z)}\Big)\Big]\\
&\le \sqrt{\mathbb{E}\Big[\sum_{x,y}\Big(\sqrt{\hat{P}(x,y|z)}-\sqrt{P(x,y|z)}\Big)^2\Big]\cdot\mathbb{E}\Big[\sum_{x,y}\Big(\sqrt{\hat{P}^{(\mathrm{M})}(x,y|z)}-\sqrt{P^{(\mathrm{M})}(x,y|z)}\Big)^2\Big]} \qquad (90)\\
&\le \frac{(|\mathcal{X}||\mathcal{Y}|)^{\frac{3}{2}}}{2\sqrt{2}n_z}. \qquad (91)
\end{aligned}
$$

For inequality (73), we have

$$
\begin{aligned}
&\mathbb{E}\Big[\sum_{x,y}\Big(\sqrt{P^{(\mathrm{M})}(x,y|z)}-\sqrt{P(x,y|z)}\Big)\cdot\Big(\sqrt{\hat{P}^{(\mathrm{M})}(x,y|z)}-\sqrt{P^{(\mathrm{M})}(x,y|z)}\Big)\Big]\\
&\le \sqrt{\sum_{x,y}\Big(\sqrt{P^{(\mathrm{M})}(x,y|z)}-\sqrt{P(x,y|z)}\Big)^2}\cdot\sqrt{\sum_{x,y}\Big\{\mathbb{E}\Big[\sqrt{\hat{P}^{(\mathrm{M})}(x,y|z)}-\sqrt{P^{(\mathrm{M})}(x,y|z)}\Big]\Big\}^2} \qquad (92)\\
&\le \sqrt{2}D\sqrt{\sum_{x,y}\mathbb{E}\Big[\Big(\sqrt{\hat{P}(x|z)}-\sqrt{P(x|z)}\Big)^2\Big]\cdot\mathbb{E}\Big[\Big(\sqrt{\hat{P}(y|z)}-\sqrt{P(y|z)}\Big)^2\Big]} \qquad (93)\\
&\le \frac{\sqrt{2|\mathcal{X}||\mathcal{Y}|}D}{4n_z} \qquad (94)
\end{aligned}
$$

While for (67) and (68), we have

$$
\begin{aligned}
&\frac{1}{2}\mathbb{E}\Big[\sum_{x,y}\Big(\alpha'\sqrt{\hat{P}(x,y|z)}+(1-\alpha')\sqrt{\hat{P}^{(\mathrm{M})}(x,y|z)}-\sqrt{P(x,y|z)}\Big)^2\Big]\\
&-\frac{1}{2}\mathbb{E}\Big[\sum_{x,y}\Big(\alpha\sqrt{\hat{P}(x,y|z)}+(1-\alpha)\sqrt{\hat{P}^{(\mathrm{M})}(x,y|z)}-\sqrt{P(x,y|z)}\Big)^2\Big]\\
&=\frac{1}{2}\mathbb{E}\Big[\sum_{x,y}\Big((\alpha'+\alpha)\sqrt{\hat{P}(x,y|z)}+(2-\alpha'-\alpha))\sqrt{\hat{P}^{(\mathrm{M})}(x,y|z)}-2\sqrt{P(x,y|z)}\Big)\Big(\alpha'-\alpha\Big)\\
&\qquad\Big(\sqrt{\hat{P}(x,y|z)}-\sqrt{\hat{P}^{(\mathrm{M})}(x,y|z)}\Big)\Big] \qquad (95)\\
&\le \mathbb{E}\Big[\sum_{x,y}\Big(\sqrt{\bar{P}(x,y|z)}-\sqrt{P(x,y|z)}\Big)\cdot\Big(\alpha'-\alpha\Big)\Big(\sqrt{\hat{P}(x,y|z)}-\sqrt{\hat{P}^{(\mathrm{M})}(x,y|z)}\Big)\Big] \qquad (96)
\end{aligned}
$$

For bound of $\mathbb{E}\Big[\Big(\sqrt{\hat{P}(x,y|z)}-\sqrt{\hat{P}^{(\mathrm{M})}(x,y|z)}\Big)^2\Big]$, we have

$$\begin{aligned}
&\mathbb{E}\Big[\Big(\sqrt{\hat{P}(x,y|z)}-\sqrt{\hat{P}^{(\mathrm{M})}(x,y|z)}\Big)^2\Big] \\
&\leq 2\Big(\mathbb{E}\Big[\Big(\sqrt{\hat{P}(x,y|z)}-\sqrt{P^{(\mathrm{M})}(x,y|z)}\Big)^2\Big] \\
&\qquad + \mathbb{E}\Big[\Big(\sqrt{P^{(\mathrm{M})}(x,y|z)}-\sqrt{\hat{P}^{(\mathrm{M})}(x,y|z)}\Big)^2\Big]\Big) \qquad (97) \\
&\leq 2\Big(D^2+\frac{|\mathcal{X}||\mathcal{Y}|}{4n_z}+\frac{\sqrt{|\mathcal{X}||\mathcal{Y}|}}{2n_z}\Big) \leq \frac{|\mathcal{X}||\mathcal{Y}|}{n_z}, \qquad (98)
\end{aligned}$$

in which we assume $|\mathcal{X}||\mathcal{Y}| > 16$. The assumption is valid as $|\mathcal{X}||\mathcal{Y}| > 64$ follows from the condition $\frac{n_z D^2}{|\mathcal{X}||\mathcal{Y}|} < \frac{1}{8}$ and $\frac{\sqrt{|\mathcal{X}||\mathcal{Y}|}}{n_z D^2} < 1$. As $0 < \alpha < \frac{8 n_z D^2}{|\mathcal{X}||\mathcal{X}|}$ and so does α', we have

$$\begin{aligned}
&\Big|\mathbb{E}\Big[\sum_{x,y}\Big(\sqrt{\bar{P}(x,y|z)}-\sqrt{P(x,y|z)}\Big) \\
&\qquad \cdot\Big(\alpha'-\alpha\Big)\Big(\sqrt{\hat{P}(x,y|z)}-\sqrt{\hat{P}^{(\mathrm{M})}(x,y|z)}\Big)\Big]\Big| \\
&\leq \sqrt{\mathbb{E}\Big[H^2\Big(\bar{P}_{XY|Z},P_{XY|Z}\Big)\Big]}\,\frac{8\sqrt{n_z|\mathcal{X}||\mathcal{Y}|}D^3}{|\mathcal{X}||\mathcal{Y}|} \qquad (99)
\end{aligned}$$

The result of Lemma 7 follows from (65)-(68) utilizing (75) and (99).