VI. Appendix

A. Proof for Lemma 2

By definition of Kullback-Leibler divergence we have

$$\begin{aligned}
&D(Q\|R)\\
=&-\sum_{t\in T} Q(t)\ln\frac{R(t)}{Q(t)} \quad (24)\\
=&-\sum_{t\in T} Q(t)\cdot 2\ln\sqrt{\frac{R(t)}{Q(t)}} \quad (25)\\
=&-2\sum_{t\in T} Q_t\Big[\Big(\sqrt{\frac{R(t)}{Q(t)}}-1\Big)-\frac{1}{2}\Big(\sqrt{\frac{R(t)}{Q(t)}}-1\Big)^2\\
&\quad+o\Big(\Big(\sqrt{\frac{R(t)}{Q(t)}}-1\Big)^2\Big)\Big] \quad (26)\\
=&2-2\sum_{t\in T}\sqrt{Q(t)R(t)}+\sum_{t\in T}\left(\sqrt{R(t)}-\sqrt{Q(t)}\right)^2\\
&\quad+o\left(H^2(Q_T,R_T)\right) \quad (27)\\
=&4H^2(Q_T,R_T)+o\left(H^2(Q_T,R_T)\right) \quad (28)
\end{aligned}$$

where (26) follows from $\ln(1+x)=x-\frac{1}{2}x^2+o(x^2)$, while (27) 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

$$F(x)\triangleq\frac{I_{\frac{|\mathcal{X}||\mathcal{Y}|}{2}}(x)}{I_{\frac{|\mathcal{X}||\mathcal{Y}|}{2}-1}(x)}. \quad (29)$$

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

$$L(x)<F(x)<U(x) \quad (30)$$

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

$$\begin{aligned}
f_p'(x)=&\frac{I_p'(x)I_{p-1}(x)-I_p(x)I_{p-1}'(x)}{I_{p-1}^2(x)} \quad (31)\\
=&\frac{\sum_{n=0}^{\infty}\frac{(2n+p)x^{2n+p-1}}{2^{2n+p}n!\Gamma(p+n+1)}}{I_{p-1}(x)}\\
&-\frac{I_p(x)\sum_{n=0}^{\infty}\frac{(2n+p-1)x^{2n+p-2}}{2^{2n+p-1}n!\Gamma(p+n)}}{I_{p-1}^2(x)} \quad (32)\\
=&\frac{(I_{p-1}(x)-\frac{p}{x}I_p(x))}{I_{p-1}(x)} \quad (33)\\
&-\frac{I_p(x)(I_p(x)+\frac{p-1}{x}I_{p-1}(x))}{I_{p-1}^2(x)} \quad (34)\\
=&1-\frac{2p-1}{x}f_p(x)-f_p^2(x) \quad (35)
\end{aligned}$$

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

$$\begin{aligned}
&\left|\mathbb{E}\big[[H(\tilde{P}^*_{XY|Z},P_{XY|Z})]^2\big]-\Big(\frac{\alpha_0^2|\mathcal{X}||\mathcal{Y}|}{8n_z}+(1-\alpha_0)^2\rho_0^2\Big)\right|\\
\le&\sqrt{\mathbb{E}\Big[H^2\Big(\tilde{P}^*_{XY|Z},P_{XY|Z}\Big)\Big]}\frac{8\sqrt{n_z|\mathcal{X}||\mathcal{Y}|}D^2}{|\mathcal{X}||\mathcal{Y}|}\\
&+\Big(\frac{8\sqrt{n_z|\mathcal{X}||\mathcal{Y}|}D^2}{|\mathcal{X}||\mathcal{Y}|}\Big)^2, \quad (36)
\end{aligned}$$

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

$$\begin{aligned}
&\mathbb{E}[[H(\tilde{P}^*_{XY|Z}, P_{XY|Z})]^2] \\
=&\frac{1}{2}\mathbb{E}\Big[\sum_{x,y}\Big(\sqrt{\tilde{P}^*(x,y|z)} - \sqrt{P(x,y|z)}\Big)^2\Big] \qquad (37)\\
=&\frac{1}{2}\mathbb{E}\Big[\sum_{x,y}\Big(\alpha_0\sqrt{\hat{P}(x,y|z)} + (1-\alpha_0)\sqrt{P^{(\mathrm{M})}(x,y|z)} \\
&\qquad - \sqrt{P(x,y|z)}\Big)^2\Big] \qquad (38)\\
&+\frac{1}{2}\mathbb{E}\Big[\sum_{x,y}\Big(\alpha\sqrt{\hat{P}(x,y|z)} + (1-\alpha)\sqrt{P^{(\mathrm{M})}(x,y|z)} \\
&\qquad - \sqrt{P(x,y|z)}\Big)^2\Big] \qquad (39)\\
&-\frac{1}{2}\mathbb{E}\Big[\sum_{x,y}\Big(\alpha_0\sqrt{\hat{P}(x,y|z)} + (1-\alpha_0)\sqrt{P^{(\mathrm{M})}(x,y|z)} \\
&\qquad - \sqrt{P(x,y|z)}\Big)^2\Big]. \qquad (40)
\end{aligned}$$

For (38), we have

$$\begin{aligned}
&\frac{1}{2}\mathbb{E}\Big[\sum_{x,y}\Big(\alpha_0\sqrt{\hat{P}(x,y|z)} + (1-\alpha_0)\sqrt{P^{(\mathrm{M})}(x,y|z)} \\
&\qquad - \sqrt{P(x,y|z)}\Big)^2\Big] \\
=&\frac{\alpha_0^2}{2}\mathbb{E}\Big[\sum_{x,y}\Big(\sqrt{\hat{P}(x,y|z)} - \sqrt{P(x,y|z)}\Big)^2\Big] \qquad (41)\\
&+\alpha_0(1-\alpha_0)\mathbb{E}\Big[\sum_{x,y}\Big(\sqrt{\hat{P}(x,y|z)} - \sqrt{P(x,y|z)}\Big) \\
&\quad\cdot\Big(\sqrt{P^{(\mathrm{M})}(x,y|z)} - \sqrt{P(x,y|z)}\Big)\Big] \qquad (42)\\
&+\frac{(1-\alpha_0)^2}{2}\mathbb{E}\Big[\sum_{x,y}\Big(\sqrt{P^{(\mathrm{M})}(x,y|z)} - \sqrt{P(x,y|z)}\Big)^2\Big]. \qquad (43)
\end{aligned}$$

Then we compute (41)-(43). For (41) and (42), since $\sqrt{\hat{P}_{XY|Z}}$ follows Gaussian distribution, we can immediately have that

$$\mathbb{E}\Big[\sum_{x,y}\Big(\sqrt{\hat{P}(x,y|z)} - \sqrt{P(x,y|z)}\Big)^2\Big] = \frac{|\mathcal{X}||\mathcal{Y}|}{4n_z}, \qquad (44)$$

$$\begin{aligned}
&\mathbb{E}\Big[\sum_{x,y}\Big(\sqrt{\hat{P}(x,y|z)} - \sqrt{P(x,y|z)}\Big) \\
&\quad\cdot\Big(\sqrt{P^{(\mathrm{M})}(x,y|z)} - \sqrt{P(x,y|z)}\Big)\Big] = 0. \qquad (45)
\end{aligned}$$

And for (43), we have that

$$\frac{1}{2}\mathbb{E}\Big[\sum_{x,y}\Big(\sqrt{P^{(\mathrm{M})}(x,y|z)} - \sqrt{P(x,y|z)}\Big)^2\Big] = \rho_0^2. \qquad (46)$$

While for (39) and (40), we have

$$\begin{aligned}
&\Big|\frac{1}{2}\mathbb{E}\Big[\sum_{x,y}\Big(\alpha\sqrt{\hat{P}(x,y|z)} + (1-\alpha)\sqrt{P^{(\mathrm{M})}(x,y|z)} \\
&\qquad - \sqrt{P(x,y|z)}\Big)^2\Big] \\
&-\frac{1}{2}\mathbb{E}\Big[\sum_{x,y}\Big(\alpha_0\sqrt{\hat{P}(x,y|z)} + (1-\alpha_0)\sqrt{P^{(\mathrm{M})}(x,y|z)} \\
&\qquad - \sqrt{P(x,y|z)}\Big)^2\Big]\Big| \\
=&\Big|\frac{1}{2}\mathbb{E}\Big[\sum_{x,y}\Big((\alpha+\alpha_0)\sqrt{\hat{P}(x,y|z)} \\
&\qquad + (2-\alpha-\alpha_0)\sqrt{P^{(\mathrm{M})}(x,y|z)} \\
&\qquad - 2\sqrt{P(x,y|z)}\Big)\cdot\Big(\alpha-\alpha_0\Big) \\
&\quad\Big(\sqrt{\hat{P}(x,y|z)} - \sqrt{P^{(\mathrm{M})}(x,y|z)}\Big)\Big]\Big| \qquad (47)\\
=&\frac{1}{2}\Big|\mathbb{E}\Big[\sum_{x,y}\Big[\Big(2\alpha\sqrt{\hat{P}(x,y|z)} + (2-2\alpha)\sqrt{P^{(\mathrm{M})}(x,y|z)} \\
&\qquad - 2\sqrt{P(x,y|z)}\Big) \\
&\qquad - \Big(\alpha-\alpha_0\Big)\Big(\sqrt{\hat{P}(x,y|z)} - \sqrt{P^{(\mathrm{M})}(x,y|z)}\Big)\Big] \\
&\quad\cdot\Big(\alpha-\alpha_0\Big)\Big(\sqrt{\hat{P}(x,y|z)} - \sqrt{P^{(\mathrm{M})}(x,y|z)}\Big)\Big]\Big| \qquad (48)\\
\le&\Big|\mathbb{E}\Big[\sum_{x,y}\Big(\sqrt{\tilde{P}^*(x,y|z)} - \sqrt{P(x,y|z)}\Big) \\
&\quad\cdot\Big(\alpha-\alpha_0\Big)\Big(\sqrt{\hat{P}(x,y|z)} - \sqrt{P^{(\mathrm{M})}(x,y|z)}\Big)\Big]\Big| \qquad (49)\\
&+\frac{1}{2}\mathbb{E}\Big[\Big(\alpha-\alpha_0\Big)^2\Big(\sqrt{\hat{P}(x,y|z)} - \sqrt{P^{(\mathrm{M})}(x,y|z)}\Big)^2\Big] \qquad (50)
\end{aligned}$$

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

$$\begin{aligned}
&\Big|\mathbb{E}\Big[\sum_{x,y}\Big(\sqrt{\tilde{P}^*(x,y|z)} - \sqrt{P(x,y|z)}\Big) \\
&\quad\cdot\Big(\alpha-\alpha_0\Big)\Big(\sqrt{\hat{P}(x,y|z)} - \sqrt{P^{(\mathrm{M})}(x,y|z)}\Big)\Big]\Big| \\
&\le\sqrt{\mathbb{E}\Big[H^2\Big(\tilde{P}^*_{XY|Z}, P_{XY|Z}\Big)\Big]}\frac{8\sqrt{n_z|\mathcal{X}||\mathcal{Y}|}D^2}{|\mathcal{X}||\mathcal{Y}|}. \qquad (51)
\end{aligned}$$

and

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

Finally, we have inequality (36) from (38)-(46) and (49)-(52).

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 (29), we have

$$\begin{aligned}&\Big|\max_{P_{XY|Z}\in U}\mathbb{E}[H^2(\tilde{P}^*_{XY|Z}, P_{XY|Z})]\\&\quad-\Big(\frac{A^2|\mathcal{X}||\mathcal{Y}|}{8n_z}+(1-A)^2D^2\Big)\Big|\\&\leq\sqrt{\max_{P_{XY|Z}\in U}\mathbb{E}\Big[H^2\Big(\tilde{P}^*_{XY|Z}, P_{XY|Z}\Big)\Big]}\frac{8\sqrt{n_z|\mathcal{X}||\mathcal{Y}|}D^2}{|\mathcal{X}||\mathcal{Y}|}\\&\quad+\Big(\frac{8\sqrt{n_z|\mathcal{X}||\mathcal{Y}|}D^2}{|\mathcal{X}||\mathcal{Y}|}\Big)^2, \end{aligned} \tag{53}$$

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

It follows that

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

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

$$|t^2-k_1|\leq k_2t+k_3$$

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

$$t^2-k_1\leq k_2t+k_3$$

we have

$$\begin{aligned}t&\leq\frac{k_2+\sqrt{k_2^2+4(k_1+k_3)}}{2} &(54)\\&=k_1+k_3+\frac{k_2}{2}+\frac{\sqrt{k_2^2+4(k_1+k_3)}-2\sqrt{k_1+k_3}}{2} &(55)\\&=k_1+k_3+\frac{k_2}{2}+\frac{k_2^2}{2(\sqrt{k_2^2+4(k_1+k_3)}+2\sqrt{k_1+k_3})} &(56)\\&\leq k_1+k_3+\frac{k_2}{2}+\frac{k_2^2}{8\sqrt{k_1+k_3}} &(57)\end{aligned}$$

As $1-16\eta\leq k_1\leq 1+8\eta$, $0\leq k_2\leq 8\sqrt{\eta}$ and $0\leq k_3\leq 64\eta$ we have $\exists C>0$ s.t. $t\leq 1+C\sqrt{\eta}$ While from $t^2-k_1\geq -k_2t-k_3$ we have

$$\begin{aligned}t&\geq\frac{-k_2+\sqrt{k_2^2+4(k_1-k_3)}}{2} &(58)\\&=k_1-k_3-\frac{k_2}{2}+\frac{\sqrt{k_2^2+4(k_1-k_3)}-2(k_1-k_3)}{2} &(59)\\&\geq k_1-k_3-\frac{k_2}{2} &(60)\end{aligned}$$

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

$$\Big|\frac{1}{D^2}\max_{P_{XY|Z}\in U}\mathbb{E}[H^2(\tilde{P}^*_{XY|Z}, P_{XY|Z})]-1\Big|\leq C\sqrt{\eta} \tag{61}$$

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 (64)-(67), which is bounded as (74) and (98) respectively.

$$\begin{aligned}&\mathbb{E}[\big[H(\bar{P}_{XY|Z}, P_{XY|Z})\big]^2]-\mathbb{E}[\big[H(\tilde{P}^*_{XY|Z}, P_{XY|Z})\big]^2]\\&=\frac{1}{2}\mathbb{E}\Big[\sum_{x,y}\Big(\sqrt{\bar{P}(x,y|z)}-\sqrt{P(x,y|z)}\Big)^2\Big] &(62)\\&\quad-\frac{1}{2}\mathbb{E}\Big[\sum_{x,y}\Big(\sqrt{\tilde{P}^*(x,y|z)}-\sqrt{P(x,y|z)}\Big)^2\Big] &(63)\\&=\frac{1}{2}\mathbb{E}\Big[\sum_{x,y}\Big(\alpha\sqrt{\hat{P}(x,y|z)}+(1-\alpha)\sqrt{P^{(\mathrm{M})}(x,y|z)}\\&\qquad+(1-\alpha)(\hat{P}^{(\mathrm{M})}(x.y|z)-P^{(\mathrm{M})}(x.y|z))\\&\qquad-\sqrt{P(x,y|z)}\Big)^2\Big] &(64)\\&\quad-\frac{1}{2}\mathbb{E}\Big[\sum_{x,y}\Big(\alpha\sqrt{\hat{P}(x,y|z)}+(1-\alpha)\sqrt{P^{(\mathrm{M})}(x,y|z)}\\&\qquad-\sqrt{P(x,y|z)}\Big)^2\Big] &(65)\\&\quad+\frac{1}{2}\mathbb{E}\Big[\sum_{x,y}\Big(\alpha'\sqrt{\hat{P}(x,y|z)}+(1-\alpha')\sqrt{\hat{P}^{(\mathrm{M})}(x,y|z)}\\&\qquad-\sqrt{P(x,y|z)}\Big)^2\Big] &(66)\\&\quad-\frac{1}{2}\mathbb{E}\Big[\sum_{x,y}\Big(\alpha\sqrt{\hat{P}(x,y|z)}+(1-\alpha)\sqrt{\hat{P}^{(\mathrm{M})}(x,y|z)}\\&\qquad-\sqrt{P(x,y|z)}\Big)^2\Big] &(67)\end{aligned}$$

For (64) and (65), we have

$$\begin{aligned}
&\frac{1}{2}\mathbb{E}\Big[\sum_{x,y}\Big(\alpha\sqrt{\hat{P}(x,y|z)}+(1-\alpha)\sqrt{P^{(\mathrm{M})}(x,y|z)}\\
&\qquad\qquad+(1-\alpha)(\hat{P}^{(\mathrm{M})}(x.y|z)-P^{(\mathrm{M})}(x.y|z))\\
&\qquad\qquad-\sqrt{P(x,y|z)}\Big)^2\Big]\\
&-\frac{1}{2}\mathbb{E}\Big[\sum_{x,y}\Big(\alpha\sqrt{\hat{P}(x,y|z)}+(1-\alpha)\sqrt{P^{(\mathrm{M})}(x,y|z)}\\
&\qquad\qquad-\sqrt{P(x,y|z)}\Big)^2\Big]\\
=&\mathbb{E}\Big[\alpha(1-\alpha)\sum_{x,y}\Big(\sqrt{\hat{P}(x,y|z)}-\sqrt{P(x,y|z)}\Big)\\
&\qquad\cdot\Big(\sqrt{\hat{P}^{(\mathrm{M})}(x,y|z)}-\sqrt{P^{(\mathrm{M})}(x,y|z)}\Big)\Big] \qquad (68)\\
&+\mathbb{E}\Big[(1-\alpha)^2\sum_{x,y}\Big(\sqrt{P^{(\mathrm{M})}(x,y|z)}-\sqrt{P(x,y|z)}\Big)\\
&\qquad\cdot\Big(\sqrt{\hat{P}^{(\mathrm{M})}(x,y|z)}-\sqrt{P^{(\mathrm{M})}(x,y|z)}\Big)\Big] \qquad (69)\\
&+\frac{1}{2}\mathbb{E}\Big[(1-\alpha)^2\sum_{x,y}\Big(\sqrt{\hat{P}^{(\mathrm{M})}(x,y|z)}-\sqrt{P^{(\mathrm{M})}(x,y|z)}\Big)^2\Big] \qquad (70)
\end{aligned}$$

Nothe the Corollary 5 gives that $0<\alpha<\frac{8n_zD^2}{|\mathcal{X}||\mathcal{Y}|}$, to give bound for (68)-(70) it suffices to prove

$$\begin{aligned}
&\mathbb{E}\Big[\sum_{x,y}\Big(\sqrt{\hat{P}(x,y|z)}-\sqrt{P(x,y|z)}\Big)\\
&\qquad\cdot\Big(\sqrt{\hat{P}^{(\mathrm{M})}(x,y|z)}-\sqrt{P^{(\mathrm{M})}(x,y|z)}\Big)\Big]\le\frac{(|\mathcal{X}||\mathcal{Y}|)^{\frac{3}{2}}}{\sqrt{2}n}, \qquad (71)
\end{aligned}$$

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

and

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

then it follows that

$$\begin{aligned}
&\frac{1}{2}\mathbb{E}\Big[\sum_{x,y}\Big(\alpha\sqrt{\hat{P}(x,y|z)}+(1-\alpha)\sqrt{P^{(\mathrm{M})}(x,y|z)}\\
&\qquad\qquad+(1-\alpha)(\hat{P}^{(\mathrm{M})}(x.y|z)-P^{(\mathrm{M})}(x.y|z))\\
&\qquad\qquad-\sqrt{P(x,y|z)}\Big)^2\Big]\\
&-\frac{1}{2}\mathbb{E}\Big[\sum_{x,y}\Big(\alpha\sqrt{\hat{P}(x,y|z)}+(1-\alpha)\sqrt{P^{(\mathrm{M})}(x,y|z)}\\
&\qquad\qquad-\sqrt{P(x,y|z)}\Big)^2\Big]\\
\le&\frac{4\sqrt{2}D^2}{\sqrt[4]{|\mathcal{X}||\mathcal{Y}|}}+\Big(\frac{\sqrt{2}}{2n_z}D+\frac{1}{2n_z}\Big)\sqrt{|\mathcal{X}||\mathcal{Y}|} \qquad (74)
\end{aligned}$$

To prove inequatlities (71)-(73), we start from (73). Then the result is used to prove (71) utilizing Cauchy inequality.

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

$$\mathbb{E}\Big[\sum_{x,y}\Big(\sqrt{\hat{P}(x,y|z)}-\sqrt{P(x,y|z)}\Big)^2\Big]=\frac{|\mathcal{X}||\mathcal{Y}|}{4n_z},. \qquad (75)$$

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

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

Further we have

$$\begin{aligned}
&\mathbb{E}\Big[\sum_{x,y}\Big(\sqrt{\hat{P}^{(\mathrm{M})}(x,y|z)}-\sqrt{P^{(\mathrm{M})}(x,y|z)}\Big)^2\Big]\\
=&\mathbb{E}\Big[\sum_{x,y}\Big(\hat{P}^{(\mathrm{M})}(x,y|z)+P^{(\mathrm{M})}(x,y|z)\\
&\qquad-2\sqrt{\hat{P}^{(\mathrm{M})}(x,y|z)}\sqrt{P^{(\mathrm{M})}(x,y|z)}\Big)\Big] \qquad (78)\\
=&2-2\mathbb{E}\Big[\sum_{x,y}\sqrt{P(x|z)P(y|z)}\sqrt{\hat{P}(x|z)\hat{P}(y|z)}\Big] \qquad (79)\\
=&2-2\mathbb{E}\Big\{\sum_{x,y}\sqrt{P(x|z)P(y|z)}\\
&\qquad\Big[\Big(\sqrt{\hat{P}(x|z)}-\sqrt{P(x|z)}\Big)\\
&\qquad\quad\cdot\Big(\sqrt{\hat{P}(y|z)}-\sqrt{P(y|z)}\Big)\\
&\qquad+\Big(\sqrt{\hat{P}(x|z)}-\sqrt{P(x|z)}\Big)\cdot\sqrt{P(y|z)}\\
&\qquad+\sqrt{P(x|z)}\cdot\Big(\sqrt{\hat{P}(y|z)}-\sqrt{P(y|z)}\Big)\\
&\qquad+\sqrt{P(x|z)}\cdot\sqrt{P(y|z)}\Big]\Big\} \qquad (80)
\end{aligned}$$

$$=-2\sum_{x,y}\mathbb{E}\Big[\Big(\sqrt{\hat{P}(x|z)}-\sqrt{P(x|z)}\Big) \cdot\Big(\sqrt{\hat{P}(y|z)}-\sqrt{P(y|z)}\Big)\Big]\cdot\sqrt{P(x|z)P(y|z)} \tag{81}$$

$$\leq 2\sqrt{\sum_{x,y}\Big\{\mathbb{E}\Big[\Big(\sqrt{\hat{P}(x|z)}-\sqrt{P(x|z)}\Big) \cdot\Big(\sqrt{\hat{P}(y|z)}-\sqrt{P(y|z)}\Big)\Big]\Big\}^2}\cdot\sqrt{\sum_{x,y}P(x|z)P(y|z)} \tag{82}$$

$$=2\sqrt{\sum_{x,y}\Big\{\mathbb{E}\Big[\Big(\sqrt{\hat{P}(x|z)}-\sqrt{P(x|z)}\Big) \cdot\Big(\sqrt{\hat{P}(y|z)}-\sqrt{P(y|z)}\Big)\Big]\Big\}^2} \tag{83}$$

Where (79) follows from $\sum_{x,y}\hat{P}^{(\mathrm{M})}(x,y|z) = \sum_{x,y}P^{(\mathrm{M})}(x,y|z) = 1$, (81) 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 (82). As $\sum_{x,y}P(x|z)P(y|z)=1$ we have (83).

By Cauchy inequality we have

$$\mathbb{E}\Big[\Big(\sqrt{\hat{P}(x|z)}-\sqrt{P(x|z)}\Big)\Big(\sqrt{\hat{P}(y|z)}-\sqrt{P(y|z)}\Big)\Big] \tag{84}$$

$$\leq\sqrt{\mathbb{E}\Big[\Big(\sqrt{\hat{P}(x|z)}-\sqrt{P(x|z)}\Big)^2\Big]} \cdot\sqrt{\mathbb{E}\Big[\Big(\sqrt{\hat{P}(y|z)}-\sqrt{P(y|z)}\Big)^2\Big]} \tag{85}$$

$$\leq\frac{1}{4n_z},\qquad \forall(x,y)\in\mathcal{X}\times\mathcal{Y} \tag{86}$$

We have bound of (83) as

$$\sum_{x,y}\Big\{\mathbb{E}\Big[\Big(\sqrt{\hat{P}(x|z)}-\sqrt{P(x|z)}\Big) \cdot\Big(\sqrt{\hat{P}(y|z)}-\sqrt{P(y|z)}\Big)\Big]\Big\}^2 \leq\frac{|\mathcal{X}||\mathcal{Y}|}{16n_z^2} \tag{87}$$

thus

$$\mathbb{E}\Big[\sum_{x,y}\Big(\sqrt{\hat{P}^{(\mathrm{M})}(x,y|z)}-\sqrt{P^{(\mathrm{M})}(x,y|z)}\Big)^2\Big]\leq\frac{\sqrt{|\mathcal{X}||\mathcal{Y}|}}{2n_z} \tag{88}$$

To prove inequality (71), by Cauchy inequality:

$$\mathbb{E}\Big[\sum_{x,y}\Big(\sqrt{\hat{P}(x,y|z)}-\sqrt{P(x,y|z)}\Big) \cdot\Big(\sqrt{\hat{P}^{(\mathrm{M})}(x,y|z)}-\sqrt{P^{(\mathrm{M})}(x,y|z)}\Big)\Big]$$

$$\leq\sqrt{\mathbb{E}\Big[\sum_{x,y}\Big(\sqrt{\hat{P}(x,y|z)}-\sqrt{P(x,y|z)}\Big)^2\Big] \cdot\mathbb{E}\Big[\sum_{x,y}\Big(\sqrt{\hat{P}^{(\mathrm{M})}(x,y|z)}-\sqrt{P^{(\mathrm{M})}(x,y|z)}\Big)^2\Big]} \tag{89}$$

$$\leq\frac{(|\mathcal{X}||\mathcal{Y}|)^{\frac{3}{2}}}{2\sqrt{2}n_z}. \tag{90}$$

For inequality (72), we have

$$\mathbb{E}\Big[\sum_{x,y}\Big(\sqrt{P^{(\mathrm{M})}(x,y|z)}-\sqrt{P(x,y|z)}\Big) \cdot\Big(\sqrt{\hat{P}^{(\mathrm{M})}(x,y|z)}-\sqrt{P^{(\mathrm{M})}(x,y|z)}\Big)\Big]$$

$$\leq\sqrt{\sum_{x,y}\Big(\sqrt{P^{(\mathrm{M})}(x,y|z)}-\sqrt{P(x,y|z)}\Big)^2} \cdot\sqrt{\sum_{x,y}\Big\{\mathbb{E}\Big[\sqrt{\hat{P}^{(\mathrm{M})}(x,y|z)}-\sqrt{P^{(\mathrm{M})}(x,y|z)}\Big]\Big\}^2} \tag{91}$$

$$\leq\sqrt{2}D\sqrt{\sum_{x,y}\mathbb{E}\Big[\Big(\sqrt{\hat{P}(x|z)}-\sqrt{P(x|z)}\Big)^2\Big] \cdot\mathbb{E}\Big[\Big(\sqrt{\hat{P}(y|z)}-\sqrt{P(y|z)}\Big)^2\Big]} \tag{92}$$

$$\leq\frac{\sqrt{2|\mathcal{X}||\mathcal{Y}|}D}{4n_z} \tag{93}$$

While for (66) and (67), we have

$$\frac{1}{2}\mathbb{E}\Big[\sum_{x,y}\Big(\alpha'\sqrt{\hat{P}(x,y|z)}+(1-\alpha')\sqrt{\hat{P}^{(\mathrm{M})}(x,y|z)} -\sqrt{P(x,y|z)}\Big)^2\Big]$$

$$-\frac{1}{2}\mathbb{E}\Big[\sum_{x,y}\Big(\alpha\sqrt{\hat{P}(x,y|z)}+(1-\alpha)\sqrt{\hat{P}^{(\mathrm{M})}(x,y|z)} -\sqrt{P(x,y|z)}\Big)^2\Big]$$

$$=\frac{1}{2}\mathbb{E}\Big[\sum_{x,y}\Big((\alpha'+\alpha)\sqrt{\hat{P}(x,y|z)} +(2-\alpha'-\alpha))\sqrt{\hat{P}^{(\mathrm{M})}(x,y|z)} -2\sqrt{P(x,y|z)}\Big)\Big(\alpha'-\alpha\Big) \Big(\sqrt{\hat{P}(x,y|z)}-\sqrt{\hat{P}^{(\mathrm{M})}(x,y|z)}\Big)\Big] \tag{94}$$

$$\leq\mathbb{E}\Big[\sum_{x,y}\Big(\sqrt{\bar{P}(x,y|z)}-\sqrt{P(x,y|z)}\Big) \cdot\Big(\alpha'-\alpha\Big)\Big(\sqrt{\hat{P}(x,y|z)}-\sqrt{\hat{P}^{(\mathrm{M})}(x,y|z)}\Big)\Big] \tag{95}$$

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

$$\begin{aligned}
&\mathbb{E}\Big[\Big(\sqrt{\hat{P}(x,y|z)}-\sqrt{\hat{P}^{(\mathrm{M})}(x,y|z)}\Big)^2\Big]\\
&\leq 2\Big(\mathbb{E}\Big[\Big(\sqrt{\hat{P}(x,y|z)}-\sqrt{P^{(\mathrm{M})}(x,y|z)}\Big)^2\Big]\\
&\qquad+\mathbb{E}\Big[\Big(\sqrt{P^{(\mathrm{M})}(x,y|z)}-\sqrt{\hat{P}^{(\mathrm{M})}(x,y|z)}\Big)^2\Big]\Big) \qquad (96)\\
&\leq 2\Big(D^2+\frac{|\mathcal{X}||\mathcal{Y}|}{4n_z}+\frac{\sqrt{|\mathcal{X}||\mathcal{Y}|}}{2n_z}\Big)\leq\frac{|\mathcal{X}||\mathcal{Y}|}{n_z}, \qquad (97)
\end{aligned}$$

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 $0<\alpha<\frac{8n_z D^2}{|\mathcal{X}||\mathcal{X}|}$ and so does α', we have

$$\begin{aligned}
&\Big|\mathbb{E}\Big[\sum_{x,y}\Big(\sqrt{\bar{P}(x,y|z)}-\sqrt{P(x,y|z)}\Big)\\
&\qquad\cdot\Big(\alpha'-\alpha\Big)\Big(\sqrt{\hat{P}(x,y|z)}-\sqrt{\hat{P}^{(\mathrm{M})}(x,y|z)}\Big)\Big]\Big|\\
&\leq\sqrt{\mathbb{E}\Big[H^2\Big(\bar{P}_{XY|Z},P_{XY|Z}\Big)\Big]}\frac{8\sqrt{n_z|\mathcal{X}||\mathcal{Y}|}D^3}{|\mathcal{X}||\mathcal{Y}|} \qquad (98)
\end{aligned}$$

The result of Lemma 7 follows from (64)-(67) utilizing (74) and (98).