Appendix

A. Proof of Theorem 9

We begin by stating two essential lemmas.

Lemma 3. *For all $\tilde{z}_{[n+m]} \in \mathcal{Z}^{n+m}$ and $w \in \mathcal{W}$, we have*

$$\mathbb{E}_{U_{[n]}}\left[\mathcal{E}\left(w, \tilde{z}_{[n+m]}, U_{[n]}\right)\right] = 0. \tag{26}$$

Proof.

$$\begin{aligned}
&\mathbb{E}_{U_{[n]}}\left[\mathcal{E}\left(w, \tilde{z}_{[n+m]}, U_{[n]}\right)\right] \\
&= \frac{1}{\binom{n+m}{n}} \sum_{u_{[n]}} \left(\frac{1}{m} \sum_{i \in \bar{u}_{[m]}} \ell(w, z_i) - \frac{1}{n} \sum_{i \in u_{[n]}} \ell(w, z_i) \right) \\
&= \frac{1}{\binom{n+m}{n}} \sum_{i=1}^{n+m} \left(\frac{1}{m} \binom{n+m-1}{n} \ell(w, z_i) - \frac{1}{n} \binom{n+m-1}{n-1} \ell(w, z_i) \right) \\
&= 0.
\end{aligned} \tag{27}$$

□

Lemma 4. *If $\ell(w, z) \in [0,1]$ for all $\lambda > 0$, $w \in \mathcal{W}$, $z \in \mathcal{Z}$, then*

$$\log \mathbb{E}_{U_{[n]}}\left[\exp\left(\lambda \mathcal{E}\left(W, \tilde{Z}_{[n+m]}, U_{[n]}\right)\right)\right] \leq \frac{\lambda^2 C_{n,m}(n+m)}{8nm}. \tag{28}$$

Proof. We start by constructing a martingale difference sequence as

$$D_i = \mathbb{E}\left[\mathcal{E}\left(w, \tilde{z}_{[n+m]}, U_{[n]}\right) | U_1, U_2, \ldots, U_i\right] - \mathbb{E}\left[\mathcal{E}\left(w, \tilde{z}_{[n+m]}, U_{[n]}\right) | U_1, U_2, \ldots, U_{i-1}\right], \quad \text{for } i = 1, 2, \ldots, n. \tag{29}$$

Then we define

$$A_i := \inf_u \mathbb{E}\left[\mathcal{E}\left(w, \tilde{z}_{[n+m]}, U_{[n]}\right) | U_1, U_2, \ldots, U_{i-1}, U_i = u\right] - \mathbb{E}\left[\mathcal{E}(w, \tilde{z}_{[n+m]}, U_{[n]}) | U_1, U_2, \ldots, U_{i-1}\right], \tag{30}$$

$$B_i := \sup_u \mathbb{E}\left[\mathcal{E}\left(w, \tilde{z}_{[n+m]}, U_{[n]}\right) | U_1, U_2, \ldots, U_{i-1}, U_i = u\right] - \mathbb{E}\left[\mathcal{E}(w, \tilde{z}_{[n+m]}, U_{[n]}) | U_1, U_2, \ldots, U_{i-1}\right]. \tag{31}$$

Observe that $A_i \leq D_i \leq B_i$ and define $\Delta_i := B_i - A_i$, then we have

$$\begin{aligned}
\Delta_i =& \sup_u \mathbb{E}\left[\mathcal{E}\left(w, \tilde{z}_{[n+m]}, U_{[n]}\right) | U_1, U_2, \ldots, U_{i-1}, U_i = u\right] - \inf_u \mathbb{E}\left[\mathcal{E}\left(w, \tilde{z}_{[n+m]}, U_{[n]}\right) | U_1, U_2, \ldots, U_{i-1}, U_i = u\right] \\
=& \sup_{u,u'} \left(\mathbb{E}\left[\mathcal{E}\left(w, \tilde{z}_{[n+m]}, U_{[n]}\right) | U_1, U_2, \ldots, U_{i-1}, U_i = u\right] - \mathbb{E}\left[\mathcal{E}\left(w, \tilde{z}_{[n+m]}, U_{[n]}\right) | U_1, U_2, \ldots, U_{i-1}, U_i = u'\right]\right) \\
=& \frac{1}{n}\left(\ell(w, z_{u'}) - \ell(w, z_u)\right) + \frac{1}{\binom{n+m-i}{n-i}} \left(\frac{1}{m} \binom{n+m-i-1}{n-i} \ell(w, z_{u'}) - \frac{1}{n} \binom{n+m-i-1}{n-i-1} \ell(w, z_{u'}) \right) \\
&- \frac{1}{\binom{n+m-i}{n-i}} \left(\frac{1}{m} \binom{n+m-i-1}{n-i} \ell(w, z_u) - \frac{1}{n} \binom{n+m-i-1}{n-i-1} \ell(w, z_u) \right) \\
=& \left(\frac{1}{n} + \frac{1}{m} \cdot \frac{m}{n+m-i} - \frac{1}{n} \cdot \frac{n-i}{n+m-i} \right) \left(\ell(w, z_{u'}) - \ell(w, z_u)\right) \\
=& \frac{n+m}{n(n+m-i)} \left(\ell(w, z_{u'}) - \ell(w, z_u)\right) \\
\leq& \frac{n+m}{n(n+m-i)}
\end{aligned} \tag{32}$$

By iteration, we decompose $\mathbb{E}\left[\exp\left(\lambda\mathcal{E}\left(w,\tilde{z}_{[n+m]},U_{[n]}\right)\right)\right]$ and apply Hoeffding's lemma to obtain

$$
\begin{aligned}
\log\mathbb{E}\left[\exp\left(\lambda\mathcal{E}\left(w,\tilde{z}_{[n+m]},U_{[n]}\right)\right)\right] &= \log\mathbb{E}\left[\exp\left(\lambda\left(\mathcal{E}\left(w,\tilde{z}_{[n+m]},U_{[n]}\right)-\mathbb{E}\left[\mathcal{E}\left(w,\tilde{z}_{[n+m]},U_{[n]}\right)\right]\right)\right)\right] && \text{(Lemma 3)}\\
&= \log\mathbb{E}\left[\exp\left(\lambda\sum_{i=1}^{n}D_i\right)\right]\\
&= \log\mathbb{E}\left[\exp\left(\lambda\sum_{i=1}^{n-1}D_i\right)\mathbb{E}\left[\exp\left(\lambda D_n\right)|U_1,U_2,\ldots,U_{n-1}\right]\right] && \text{(tower expectation)}\\
&\leq \log\mathbb{E}\left[\exp\left(\lambda\sum_{i=1}^{n-1}D_i\right)\right]+\frac{\lambda^2\Delta_i^2}{8} && \text{(Hoeffding's lemma)}\\
&\ \ \vdots\\
&\leq \frac{\lambda^2(n+m)^2}{8n^2}\sum_{i=1}^{n}\frac{1}{(n+m-i)^2}\\
&= \frac{\lambda^2(n+m)^2}{8n^2}\sum_{i=m}^{n+m-1}\frac{1}{i^2}.
\end{aligned}
\tag{33}
$$

Now consider an equivalent setup where we use $U_{[m]}$ to pick the test set, then we can also obtain a upper bound that resembles (33). This thus gives a tightened upper bound as

$$
\log\mathbb{E}\left[\exp\left(\lambda\mathcal{E}\left(w,\tilde{z}_{[n+m]},U_{[n]}\right)\right)\right]\leq\min\left(\frac{\lambda^2(n+m)^2}{8n^2}\sum_{i=m}^{n+m-1}\frac{1}{i^2},\frac{\lambda^2(n+m)^2}{8m^2}\sum_{i=n}^{n+m-1}\frac{1}{i^2}\right). \tag{34}
$$

Note that when $m=1$ or $n=1$, the RHS of (34) can be directly computed as

$$
\log\mathbb{E}\left[\exp\left(\lambda\mathcal{E}\left(w,\tilde{z}_{[n+m]},U_{[n]}\right)\right)\right]\leq\frac{\lambda^2(n+m)^2}{8\max(n^2,m^2)}. \tag{35}
$$

Otherwise, (34) is further upper bounded by

$$
\begin{aligned}
\log\mathbb{E}\left[\exp\left(\lambda\mathcal{E}\left(w,\tilde{z}_{[n+m]},U_{[n]}\right)\right)\right] &\leq \frac{\lambda^2(n+m)^2}{8}\min\left(\frac{1}{n^2}\sum_{i=m}^{n+m-1}\frac{1}{i(i-1)},\frac{1}{m^2}\sum_{i=n}^{n+m-1}\frac{1}{i(i-1)}\right)\\
&\leq \frac{\lambda^2(n+m)^2}{8}\min\left(\frac{1}{n^2}\left(\frac{1}{m-1}-\frac{1}{n+m-1}\right),\frac{1}{m^2}\left(\frac{1}{n-1}-\frac{1}{n+m-1}\right)\right)\\
&= \frac{\lambda^2(n+m)^2}{8(n+m-1)}\cdot\frac{1}{nm-\min(n,m)}
\end{aligned}
\tag{36}
$$

Combining (35) and (36) together completes the proof. $\square$

Now, by the Donsker–Varadhan variational representation of KL divergence, for any $\lambda\in\mathbb{R}$ we have

$$
\begin{aligned}
&\mathbb{E}_{W,U_{[n]}|\tilde{z}_{[n+m]}}\left[\lambda\mathcal{E}(W,\tilde{z}_{[n+m]},U_{[n]})\right]\\
&\leq I^{\tilde{z}_{[n+m]}}\left(W;U_{[n]}\right)+\log\mathbb{E}_{W|\tilde{z}_{[n+m]}}\mathbb{E}_{U_{[n]}|\tilde{z}_{[n+m]}}\left[\exp\left(\lambda\mathcal{E}\left(W,\tilde{z}_{[n+m]},U_{[n]}\right)\right)\right] && \text{(Donsker-Varadhan)}\\
&= I^{\tilde{z}_{[n+m]}}\left(W;U_{[n]}\right)+\log\mathbb{E}_{W|\tilde{z}_{[n+m]}}\mathbb{E}_{U_{[n]}}\left[\exp\left(\lambda\mathcal{E}\left(W,\tilde{z}_{[n+m]},U_{[n]}\right)\right)\right] && (U_{[n]}\perp\!\!\!\perp\tilde{Z}_{[n+m]})\\
&\leq I^{\tilde{z}_{[n+m]}}\left(W;U_{[n]}\right)+\frac{\lambda^2C_{n,m}(n+m)}{8mn}. && \text{(Lemma 4)}
\end{aligned}
\tag{37}
$$

After optimizing λ and taking expectations over $\tilde{Z}_{[n+m]}$ on both sides, we obtain

$$
\begin{aligned}
\mathbb{E}_{W,U_{[n]},\tilde{Z}_{[n+m]}}\left[\mathcal{E}\left(W,\tilde{Z}_{[n+m]},U_{[n]}\right)\right] &\leq \mathbb{E}_{\tilde{Z}_{[n+m]}}\left[\inf_{\lambda>0}\frac{1}{\lambda}I^{\tilde{Z}_{[n+m]}}\left(W;U_{[n]}\right)+\frac{\lambda C_{n,m}(n+m)}{8mn}\right]\\
&= \mathbb{E}_{\tilde{Z}_{[n+m]}}\left[\sqrt{\frac{C_{n,m}(n+m)}{2mn}I^{\tilde{Z}_{[n+m]}}\left(W;U_{[n]}\right)}\right]\\
&\leq \sqrt{\frac{C_{n,m}(n+m)}{2mn}I\left(W;U_{[n]}|\tilde{Z}_{[n+m]}\right)}. && \text{(Jensen's inequality)}
\end{aligned}
\tag{38}
$$

This concludes the proof.

B. Proof of Theorem 7

The corresponding expected generalization error with respect to (13) is

$$\begin{aligned}\mathbb{E}_{W,\tilde{Z}_{[n+m]},U_{[n]}}\left[\lambda\mathcal{E}\left(W,\tilde{Z}_{[n+m]},U_{[n]}\right)\right]=&\frac{1}{k}\sum_{i=1}^{k}\mathbb{E}_{W,\tilde{Z}_{[n+m]},U_{[n]}}\left[\lambda\mathcal{E}_i\left(w,\tilde{z}^{(i)}_{[\frac{n+m}{k}]},u^{(i)}_{[\frac{n}{k}]}\right)\right]\\=&\frac{1}{k}\sum_{i=1}^{k}\mathbb{E}_{W,\tilde{Z}_{[n+m]},U^{(i)}_{[\frac{n}{k}]}}\left[\lambda\mathcal{E}_i\left(w,\tilde{z}^{(i)}_{[\frac{n+m}{k}]},u^{(i)}_{[\frac{n}{k}]}\right)\right].\end{aligned}\tag{39}$$

Again, by using the Donsker-Varadhan inequality, we upper bound each summand of the above expression, i.e.,

$$\mathbb{E}_{W,U^{(i)}_{[\frac{n}{k}]}|\tilde{z}_{[n+m]}}\left[\lambda\mathcal{E}_i\left(w,\tilde{z}^{(i)}_{[\frac{n+m}{k}]},u^{(i)}_{[\frac{n}{k}]}\right)\right]\le I^{\tilde{z}_{[n+m]}}\left(W;U^{(i)}_{[\frac{n}{k}]}\right)+\log\mathbb{E}_{W|\tilde{z}_{[n+m]}}\mathbb{E}_{U^{(i)}_{[\frac{n}{k}]}}\left[\exp\left(\lambda\mathcal{E}_i\left(w,\tilde{z}^{(i)}_{[\frac{n+m}{k}]},u^{(i)}_{[\frac{n}{k}]}\right)\right)\right].\tag{40}$$

Notice that Lemma 4 is also applicable to each individual partition, where n and m are scaled to $\frac{n}{k}$ and $\frac{m}{k}$, respectively, we have

$$\begin{aligned}\log\mathbb{E}_{U^{(i)}_{[\frac{n}{k}]}}\left[\exp\left(\lambda\mathcal{E}_i\left(w,\tilde{z}^{(i)}_{[\frac{n+m}{k}]},u^{(i)}_{[\frac{n}{k}]}\right)\right)\right]&\le\min\left(\frac{\lambda^2(n+m)^2}{8n^2}\sum_{i=\frac{m}{k}}^{\frac{n+m}{k}-1}\frac{1}{i^2},\frac{\lambda^2(n+m)^2}{8m^2}\sum_{i=\frac{n}{k}}^{\frac{n+m}{k}-1}\frac{1}{i^2}\right)\\&=\frac{\lambda^2C^k_{n,m}k(n+m)}{8mn}.\end{aligned}\tag{41}$$

The result now follows by optimizing λ and averaging over $\tilde{Z}^{(i)}_{[\frac{n+m}{k}]}$, as done in (38).

C. Proof of Theorem 8

Starting from (39), by marginalizing out partitions that do not appear in each summand of (39), we further simplify the expected generalization error as

$$\mathbb{E}_{W,\tilde{Z}_{[n+m]},U_{[n]}}\left[\lambda\mathcal{E}\left(W,\tilde{Z}_{[n+m]},U_{[n]}\right)\right]=\frac{1}{k}\sum_{i=1}^{k}\mathbb{E}_{W,\tilde{Z}^{(i)}_{[\frac{n+m}{k}]},U^{(i)}_{[\frac{n}{k}]}}\left[\lambda\mathcal{E}_i\left(w,\tilde{z}^{(i)}_{[\frac{n+m}{k}]},u^{(i)}_{[\frac{n}{k}]}\right)\right].\tag{42}$$

Using the Donsker-Varadhan inequality gives

$$\begin{aligned}\mathbb{E}_{W,U^{(i)}_{[\frac{n}{k}]}|\tilde{z}^{(i)}_{[\frac{n+m}{k}]}}\left[\lambda\mathcal{E}_i\left(w,z^{(i)}_{[\frac{n+m}{k}]},u^{(i)}_{[\frac{n}{k}]}\right)\right]&\le I^{\tilde{z}^{(i)}_{[\frac{n+m}{k}]}}\left(W;U^{(i)}_{[\frac{n}{k}]}\right)+\log\mathbb{E}_{W|\tilde{z}^{(i)}_{[\frac{n+m}{k}]}}\mathbb{E}_{U^{(i)}_{[\frac{n}{k}]}}\left[\exp\left(\lambda\mathcal{E}_i\left(w,z^{(i)}_{[\frac{n+m}{k}]},u^{(i)}_{[\frac{n}{k}]}\right)\right)\right]\\&=I^{\tilde{z}^{(i)}_{[\frac{n+m}{k}]}}\left(W;U^{(i)}_{[\frac{n}{k}]}\right)+\frac{\lambda^2C^k_{n,m}k(n+m)}{8mn}.\end{aligned}\tag{43}$$

The rest of the argument also follows from (38).

D. Proof of Lemma 2

For brevity, we omit the subscript of each set (e.g., $U^{(i)}$ for $U^{(i)}_{[\frac{n}{k}]}$) as the set size is clear from the context. To prove $I\left(W;U^{(i)}|\tilde{Z}^{(i)}\right)\le I\left(W;U^{(i)}|\tilde{Z}\right)$, we let $\tilde{Z}^{\backslash(i)}=\tilde{Z}\backslash\tilde{Z}^{(i)}$, then

$$\begin{aligned}&I\left(W;U^{(i)}|\tilde{Z}\right)-I\left(W;U^{(i)}|\tilde{Z}^{(i)}\right)\\&=I\left(W;U^{(i)}|\tilde{Z}^{(i)},\tilde{Z}^{\backslash(i)}\right)-I\left(W;U^{(i)}|\tilde{Z}^{(i)}\right)\\&=I\left(W;U^{(i)}|\tilde{Z}^{(i)},\tilde{Z}^{\backslash(i)}\right)+I\left(U^{(i)};\tilde{Z}^{\backslash(i)}|\tilde{Z}^{(i)}\right)\\&\quad-I\left(W;U^{(i)}|\tilde{Z}^{(i)}\right)\\&=I\left(W,\tilde{Z}^{\backslash(i)};U^{(i)}|\tilde{Z}^{(i)}\right)-I\left(W;U^{(i)}|\tilde{Z}^{(i)}\right)\\&=I\left(U^{(i)};\tilde{Z}^{\backslash(i)}|\tilde{Z}^{(i)},W\right)\\&\ge 0\end{aligned}\tag{44}$$

E. Proof of Theorem 10

Lemma 5. *When m tends to infinity and k, n are constants, we have* $\lim_{m\to\infty} I\left(W; \tilde{Z}_{[n+m]}\right) = \lim_{m\to\infty} I\left(W; \tilde{Z}^{(i)}_{[\frac{n+m}{k}]}\right) = 0$.

Proof. Using the tower property of conditional expectation and the fact that $U_{[n]} \perp\!\!\!\perp \tilde{Z}_{[n+m]}$, we get

$$
\begin{aligned}
P_{W|\tilde{Z}_{[n+m]}} &= \mathbb{E}_{U_{[n]}}\left[P_{W|\tilde{Z}_{[n+m]},U_{[n]}}\right] \\
&= \frac{1}{\binom{\frac{n+m}{k}}{\frac{n}{k}}^k} \sum_{u_{[n]}} P_{W|\tilde{Z}_{[n+m]},U_{[n]}}(w|\tilde{z}_{[n+m]}, u_{[n]}) \\
&= \frac{1}{\binom{\frac{n+m}{k}}{\frac{n}{k}}^k} \sum_{u_{[n]}} P_{W|\tilde{Z}_{U_{[n]}}}(w|\tilde{z}_{u_{[n]}}).
\end{aligned} \tag{45}
$$

Let $g_w : \mathcal{Z}^{n+m} \to [0,1]$ be a function defined as $g_w(\tilde{Z}_{[n+m]}) := P_{W=w|\tilde{Z}_{[n+m]}}$. Given a supersample set $\tilde{z}_{[n+m]} \in \mathcal{Z}^{n+m}$, let $\tilde{z}'_{[n+m]}$ equal $\tilde{z}_{[n+m]}$ for all instances except the i-th, i.e., $\tilde{z}'_{[n+m]} = \tilde{z}_{[n+m]} \backslash \{\tilde{z}_i\} \cup \{\tilde{z}'_i\}$, where $\tilde{z}'_i$ is an independent copy of $\tilde{z}_i$. For each summand in (45), it is shown that $P_{W|\tilde{Z}_{U_{[n]}}}(w|\tilde{z}_{u_{[n]}})$ and $P_{W|\tilde{Z}_{U_{[n]}}}(w|\tilde{z}'_{u_{[n]}})$ only differ when the index i is included by $u_{[n]}$. Thus, the absolute difference between $g_w(\tilde{z}_{[n+m]})$ and $g_w(\tilde{z}'_{[n+m]})$ can be bounds as:

$$
\begin{aligned}
\left|g_w(\tilde{z}_{[n+m]}) - g_w(\tilde{z}'_{[n+m]})\right| &= \left| \frac{1}{\binom{\frac{n+m}{k}}{\frac{n}{k}}^k} \sum_{u_{[n]}:\ i \in u_{[n]}} P_{W|\tilde{Z}_{U_{[n]}}}(w|\tilde{z}_{u_{[n]}}) - P_{W|\tilde{Z}_{U_{[n]}}}(w|\tilde{z}'_{u_{[n]}}) \right| \\
&\leq \frac{1}{\binom{\frac{n+m}{k}}{\frac{n}{k}}^k} \sum_{u_{[n]}:\ i \in u_{[n]}} \left| P_{W|\tilde{Z}_{U_{[n]}}}(w|\tilde{z}_{u_{[n]}}) - P_{W|\tilde{Z}_{U_{[n]}}}(w|\tilde{z}'_{u_{[n]}}) \right| \\
&\leq \frac{1}{\binom{\frac{n+m}{k}}{\frac{n}{k}}^k} \cdot \binom{\frac{n+m}{k}}{\frac{n}{k}}^{k-1} \cdot \binom{\frac{n+m}{k} - 1}{\frac{n}{k} - 1} \\
&= \frac{n}{n+m}.
\end{aligned} \tag{46}
$$

With the difference property in (46), applying McDiarmid's inequality gives

$$
\mathbb{P}\left\{\left|g_w(\tilde{Z}_{[n+m]}) - \mathbb{E}\left[g_w(\tilde{Z}_{[n+m]})\right]\right| \geq \epsilon\right\} \leq \exp\left(-\frac{2(n+m)\epsilon^2}{n^2}\right). \tag{47}
$$

Since $\mathbb{E}\left[g_w(\tilde{Z}_{[n+m]})\right] = P_{W=w}$, in probability we have $\lim_{m\to\infty} P_{W=w|\tilde{Z}_{[n+m]}} = P_{W=w}$. Next, by definition of mutual information, we obtain

$$
\begin{aligned}
\lim_{m\to\infty} I\left(W; \tilde{Z}_{[n+m]}\right) &= \lim_{m\to\infty} \mathbb{E}\left[D\left(P_{W|\tilde{Z}_{[n+m]}} \| P_W\right)\right] \\
&= \lim_{m\to\infty} \sum_w \mathbb{E}\left[P_{W=w|\tilde{Z}_{[n+m]}} \log \frac{P_{W=w|\tilde{Z}_{[n+m]}}}{P_{W=w}}\right] \\
&= 0.
\end{aligned} \tag{48}
$$

By the chain rule of mutual information,

$$
I\left(W; \tilde{Z}_{[n+m]}\right) = I\left(W; \tilde{Z}^{(i)}_{[\frac{n+m}{k}]}\right) + I\left(W; \tilde{Z}^{\backslash(i)}_{[\frac{k-1}{k}(n+m)]} | \tilde{Z}^{(i)}_{[\frac{n+m}{k}]}\right), \tag{49}
$$

where $\tilde{Z}^{\backslash(i)}_{[\frac{k-1}{k}(n+m)]} = \tilde{Z}_{[n+m]} \backslash \tilde{Z}^{(i)}_{[\frac{n+m}{k}]}$. Due to the non-negativity of conditional mutual information and take $\lim_{m\to\infty}$ on both sides, we finally get

$$
\lim_{m\to\infty} I\left(W; \tilde{Z}^{(i)}_{[\frac{n+m}{k}]}\right) = 0. \tag{50}
$$

□

We first provide the proof for the IPCIMI bound. According to the proposed supersample setting, we have

$$
\begin{aligned}
I\left(W;\tilde{Z}^{(i)}_{[\frac{n+m}{k}]},U^{(i)}_{[\frac{n}{k}]}\right) &= I\left(W;\tilde{Z}^{(i)}_{U^{(i)}_{[\frac{n}{k}]}},\tilde{Z}^{(i)}_{\bar{U}^{(i)}_{[\frac{m}{k}]}},U^{(i)}_{[\frac{n}{k}]}\right)\\
&= I\left(W;\tilde{Z}^{(i)}_{U^{(i)}_{[\frac{n}{k}]}}\right)+I\left(W;\tilde{Z}^{(i)}_{\bar{U}^{(i)}_{[\frac{m}{k}]}},U^{(i)}_{[\frac{n}{k}]}|\tilde{Z}^{(i)}_{U^{(i)}_{[\frac{n}{k}]}}\right)\\
&\overset{(a)}{=} I\left(W;\tilde{Z}^{(i)}_{U^{(i)}_{[\frac{n}{k}]}}\right)\\
&\overset{(b)}{=} I\left(W;Z^{(i)}_{[\frac{n}{k}]}\right),
\end{aligned}
\tag{51}
$$

where (a) is due to the fact that W is independent of other random variables given $\tilde{Z}_{U_{[n]}}$, and (b) is according to the supersample set structure where $Z^{(i)}_{[\frac{n}{k}]}$ and $\tilde{Z}^{(i)}_{[\frac{n+m}{k}]}$ lead to the same training set. Next, using the chain rule of mutual information, one can obtain

$$
\begin{aligned}
\lim_{m\to\infty} I\left(W;U^{(i)}_{[\frac{n}{k}]}|\tilde{Z}^{(i)}_{[\frac{n+m}{k}]}\right) &= \lim_{m\to\infty} I\left(W;\tilde{Z}^{(i)}_{[\frac{n+m}{k}]},U^{(i)}_{[\frac{n}{k}]}\right)-I\left(W;\tilde{Z}^{(i)}_{[\frac{n+m}{k}]}\right)\\
&= I\left(W;Z^{(i)}_{[\frac{n}{k}]}\right)-\lim_{m\to\infty} I\left(W;\tilde{Z}^{(i)}_{[\frac{n+m}{k}]}\right)\\
&= I\left(W;Z^{(i)}_{[\frac{n}{k}]}\right),
\end{aligned}
\tag{52}
$$

where the last step is by Lemma 5. Finally, it is simple to verify that

$$
\begin{aligned}
\frac{1}{k}\sum_{i=1}^{k}\lim_{m\to\infty}\sqrt{\frac{C^k_{n,m}k(n+m)}{2nm}I\left(W;U^{(i)}_{[\frac{n}{k}]}|\tilde{Z}^{(i)}_{[\frac{n+m}{k}]}\right)} &= \frac{1}{k}\sum_{i=1}^{k}\sqrt{\lim_{m\to\infty}\frac{C^k_{n,m}k(n+m)}{2nm}\lim_{m\to\infty} I\left(W;U^{(i)}_{[\frac{n}{k}]}|\tilde{Z}^{(i)}_{[\frac{n+m}{k}]}\right)}\\
&= \frac{1}{k}\sum_{i=1}^{k}\sqrt{\frac{k}{2n}I\left(W;Z^{(i)}_{[\frac{n}{k}]}\right)}.
\end{aligned}
\tag{53}
$$

Specially, when $k=1$ and $m\to\infty$, the $(\infty,1)$-IPCIMI bound reduces to the MI bound as

$$
\lim_{m\to\infty}\sqrt{\frac{C^k_{n,m}(n+m)}{2nm}I\left(W;U_{[n]}|\tilde{Z}_{[n+m]}\right)}=\sqrt{\frac{1}{2n}I\left(W;Z_{[n]}\right)},
\tag{54}
$$

while the (∞,n)-IPCIMI bound recovers the IMI bound:

$$
\frac{1}{n}\sum_{i=1}^{n}\lim_{m\to\infty}\sqrt{\frac{C^k_{n,m}(n+m)}{2m}I\left(W;U^{(i)}_{[1]}|\tilde{Z}^{(i)}_{[\frac{m}{n}+1]}\right)}=\frac{1}{n}\sum_{i=1}^{n}\sqrt{\frac{1}{2}I\left(W;Z_i\right)}.
\tag{55}
$$

For the IPCMI bound, with the principle argument derived from the equivalence between the mutual information terms, i.e.,

$$
\begin{aligned}
\lim_{m\to\infty} I\left(W;U^{(i)}_{[\frac{n}{k}]}|\tilde{Z}_{[n+m]}\right) &= \lim_{m\to\infty} I\left(W;\tilde{Z}_{[n+m]},U^{(i)}_{[\frac{n}{k}]}\right)-I\left(W;\tilde{Z}_{[n+m]}\right)\\
&= I\left(W;Z^{(i)}_{[\frac{n}{k}]}\right)-\lim_{m\to\infty} I\left(W;\tilde{Z}^{(i)}_{[\frac{n+m}{k}]}\right)\\
&= I\left(W;Z^{(i)}_{[\frac{n}{k}]}\right),
\end{aligned}
\tag{56}
$$

repeating the same procedure above completes the proof.

F. Calculation Details for the Bernoulli Example

Before presenting the calculation details, we introduce some useful approximations as below.

Lemma 6. *Let X and Y be two independent binomial random variables following $X\sim\mathrm{Bin}(n,p)$ and $Y\sim\mathrm{Bin}(m,p)$, respectively. If $\bar{X}=n-X$, then $\bar{X}\sim\mathrm{Bin}(n,1-p)$. If $Z=X+Y$, then $Z\sim\mathrm{Bin}(n+m,p)$.*

Lemma 7. *If X is a binomial random variable following $X\sim\mathrm{Bin}(n,p)$ and $a>0$, then*

$$
\mathbb{E}\left[\log(X+a)\right]=\log(np+a)-\frac{np(1-p)}{2(np+a)^2\ln 2}+O\left(\frac{1}{n^2}\right).
\tag{57}
$$

Lemma 8. *If X is a binomial random variable following $X \sim \mathrm{Bin}(n, p)$, then*

$$\mathbb{E}\left[X \log X\right] = np \log np + \frac{1-p}{2\ln 2} + \frac{1-p^2}{12np\ln 2} + O\left(\frac{1}{n^2}\right). \tag{58}$$

Lemma 9. *If a, b, c, d are all constants, then $\log \frac{an+b}{cn+d}$ scales as $\log \frac{a}{c} + O\left(\frac{1}{n}\right)$.*

Lemma 10 (Bounds for a binomial coefficient [16]). *If $1 \le k \le n-1$, then*

$$\sqrt{\frac{n}{8k(n-k)}}2^{nH\left(\frac{k}{n}\right)} \le \binom{n}{k} \le \sqrt{\frac{n}{2\pi k(n-k)}}2^{nH\left(\frac{k}{n}\right)}. \tag{59}$$

1) MI Case:

$$I\left(W; Z_{[n]}\right) = -\mathbb{E}\left[\log\binom{n}{X} + X\log p + \bar{X}\log(1-p)\right], \tag{60}$$

where $X \sim \mathrm{Bin}(n, p)$ and $\bar{X} = n - X \sim \mathrm{Bin}(n, 1-p)$.

First, by Lemma 10, the lower bound of the expected binomial coefficient can be approximated as

$$\begin{aligned}
\mathbb{E}\left[\log\binom{n}{X}\right] \ge& \mathbb{E}\left[\frac{1}{2}\left(\log n - \log X - \log\bar{X} - 3\right) + nH(\frac{X}{n})\right] \\
=&\frac{1}{2}\log n - \frac{3}{2} - \mathbb{E}\left[\frac{1}{2}\left(\log X + \log\bar{X}\right) + X\log\frac{X}{n} + \bar{X}\log\frac{\bar{X}}{n}\right] \\
=&\frac{1}{2}\log n - \frac{3}{2} - \mathbb{E}\left[\frac{1}{2}\left(\log X + \log\bar{X}\right) + X\log X + \bar{X}\log\bar{X} - X\log n - \bar{X}\log n\right] \\
=&\frac{1}{2}\log n - \frac{3}{2} - \left(\frac{1}{2}\left(\log np - \frac{1-p}{2np\ln 2} + \log n(1-p) - \frac{p}{2n(1-p)\ln 2} + O\left(\frac{1}{n^2}\right)\right)\right. \\
&+ np\log np + \frac{1-p}{2\ln 2} + \frac{1-p^2}{12np\ln 2} + O\left(\frac{1}{n^2}\right) \\
&+ n(1-p)\log n(1-p) + \frac{p}{2\ln 2} + \frac{1-(1-p)^2}{12n(1-p)\ln 2} + O\left(\frac{1}{n^2}\right) \\
&\left. - np\log n - n(1-p)\log n\right) \\
=& -\frac{1}{2}\log np(1-p) - \frac{1}{2\ln 2} - \frac{3}{2} - np\log p - n(1-p)\log(1-p) + O\left(\frac{1}{n}\right) \\
=&nH(p) - \frac{1}{2}\log n - \frac{\log 8ep(1-p)}{2} + O\left(\frac{1}{n}\right).
\end{aligned} \tag{61}$$

Accordingly, the MI bound is upper bounded by

$$\begin{aligned}
\sqrt{\frac{1}{2n}I\left(W; Z_{[n]}\right)} =& \sqrt{-\frac{1}{2n}\mathbb{E}\left[\log\binom{n}{X} + X\log p + \bar{X}\log(1-p)\right]} \\
\le& \sqrt{\frac{1}{2n}\left(-nH(p) + \frac{1}{2}\log n + \frac{\log 8ep(1-p)}{2} + O\left(\frac{1}{n}\right) - np\log p - n(1-p)\log(1-p)\right)} \\
=& \sqrt{-\frac{1}{2}H(p) + \frac{1}{4}\frac{\log n}{n} + \frac{\log 8ep(1-p)}{4n} + O\left(\frac{1}{n^2}\right) + \frac{1}{2}H(p)} \\
=& \sqrt{\frac{1}{4}\frac{\log n}{n} + O\left(\frac{1}{n}\right)}.
\end{aligned} \tag{62}$$

Likewise, the approximated lower bound can be obtained as

$$\begin{aligned}
\sqrt{\frac{1}{2n}I\left(W; Z_{[n]}\right)} \ge& \sqrt{\frac{1}{4}\frac{\log n}{n} + \frac{\log 2\pi ep(1-p)}{4n} + O\left(\frac{1}{n^2}\right)} \\
=& \sqrt{\frac{1}{4}\frac{\log n}{n} + O\left(\frac{1}{n}\right)}.
\end{aligned} \tag{63}$$

Thus, the MI bound scales as $O\left(\sqrt{\frac{\log n}{n}}\right)$.

2) IMI Case:

$$I\left(W;Z_i\right)=H(p)-\log n+p\mathbb{E}\left[\log(X+1)\right]+(1-p)\mathbb{E}\left[\log(\bar{X}+1)\right], \tag{64}$$

where $X\sim\mathrm{Bin}(n-1,p)$ and $\bar{X}=n-1-X\sim\mathrm{Bin}(n-1,1-p)$.

In order to obtain the order of the above expression, we apply the approximation introduced in Lemma 7 and get

$$\begin{aligned}I\left(W;Z_i\right)=&H(p)-\log n+p\log((n-1)p+1)-\frac{(n-1)(1-p)p}{2\left((n-1)p+1\right)^2\ln 2}\\&+(1-p)\log((n-1)(1-p)+1)-\frac{(n-1)(1-p)p}{2\left((n-1)(1-p)+1\right)^2\ln 2}+O\left(\frac{1}{n^2}\right)\\=&H(p)+p\log\frac{np+1-p}{n}+(1-p)\log\frac{(1-p)n+p}{n}+O\left(\frac{1}{n}\right).\end{aligned}\tag{65}$$

We further apply Lemma 9 to obtain

$$\begin{aligned}I\left(W;Z_i\right)=&H(p)+p\log p+(1-p)\log(1-p)+O\left(\frac{1}{n}\right)\\=&O\left(\frac{1}{n}\right),\end{aligned}\tag{66}$$

Plugging which back to (7) implies that the IMI bound is of order $O\left(\frac{1}{\sqrt{n}}\right)$.

3) IPMI Case:

$$\begin{aligned}I\left(W;Z^{(i)}_{[\frac{n}{k}]}\right)=&\sum_{y=0}^{n-\frac{n}{k}}\binom{n-\frac{n}{k}}{y}p^y(1-p)^{n-\frac{n}{k}-y}\sum_{x=0}^{\frac{n}{k}}\binom{\frac{n}{k}}{x}p^x(1-p)^{\frac{n}{k}-x}\log\frac{\binom{n-\frac{n}{k}}{y}}{\binom{n}{x+y}p^x(1-p)^{\frac{n}{k}-x}}\\=&\mathbb{E}\left[\log\binom{n-\frac{n}{k}}{Y}-\log\binom{n}{X+Y}-X\log p-\bar{X}\log(1-p)\right],\end{aligned}\tag{67}$$

where $X\sim\mathrm{Bin}(\frac{n}{k},p)$, $Y\sim\mathrm{Bin}(n-\frac{n}{k},p)$ and $\bar{X}=\frac{n}{k}-X\sim\mathrm{Bin}(\frac{n}{k},1-p)$.

4) LOO-CMI Case: Under this case, the learner is simplified as $W=\frac{1}{n}(\sum_{i=1}^{n+1}\tilde{Z}_i-\tilde{Z}_U)$, and by Bayes rule we get

$$P_{U|W,\tilde{Z}_{[n+1]}}(u|w,\tilde{z}_{[n+1]})=\frac{P_{W|\tilde{Z}_{[n+1]},U}(w|\tilde{z}_{[n+1]},u)P_{\tilde{Z}_{[n+1]}}(\tilde{z}_{[n+1]})P_U(u)}{\sum_u P_{W,\tilde{Z}_{[n+1]},U}(w,\tilde{z}_{[n+1]},u)}. \tag{68}$$

For $w=\frac{k}{n}$, it is easy to verify that the number of 1's in $\tilde{z}_{[n+1]}$ is either k when $\tilde{z}_u=0$ or $k+1$ when $\tilde{z}_u=1$ since otherwise $P_{W|\tilde{Z}_{[n+1]},U}$ is zero. When $\tilde{z}_u=0$ and $\sum_{i=1}^{n+1}\tilde{z}_i=k$, we have

$$\begin{aligned}P_{U|W,\tilde{Z}_{[n+1]}}(u|w,\tilde{z}_{[n+1]})=&\frac{\frac{1}{n+1}p^k(1-p)^{n+1-k}}{\frac{n+1-k}{n+1}p^k(1-p)^{n+1-k}}\\=&\frac{1}{n+1-k}.\end{aligned}\tag{69}$$

Likewise, it holds when $\tilde{z}_u=1$ and $\sum_{i=1}^{n+1}\tilde{z}_i=k+1$:

$$P_{U|W,\tilde{Z}_{[n+1]}}(u|w,\tilde{z}_{[n+1]})=\frac{1}{k+1}. \tag{70}$$

For $w=\frac{k}{n}$, combining (69) and (70) gives

$$P_{U|W,\tilde{Z}_{[n+1]}}(u|w,\tilde{z}_{[n+1]})=\begin{cases}\frac{1}{n+1-k} & \text{If } \tilde{z}_u=0 \text{ and } \sum_{i=1}^{n+1}\tilde{z}_i=k\\ \frac{1}{k+1} & \text{If } \tilde{z}_u=1 \text{ and } \sum_{i=1}^{n+1}\tilde{z}_i=k+1\\ 0 & \text{Otherwise}\end{cases}. \tag{71}$$

By definition, we compute the mutual information as

$$
\begin{aligned}
I\left(W;U|\tilde{Z}_{[n+1]}\right) &= \mathbb{E}_{W,\tilde{Z}_{[n+1]},U}\left[\log\frac{P_{U|W,\tilde{Z}_{[n+1]}}}{P_U}\right] \\
&= \sum_{k=0}^{n}\frac{n+1-k}{n+1}\binom{n+1}{k}p^k(1-p)^{n+1-k}\log\frac{n+1}{n+1-k} + \sum_{k=0}^{n}\frac{k+1}{n+1}\binom{n+1}{k+1}p^{k+1}(1-p)^{n-k}\log\frac{n+1}{k+1} \\
&= \log(n+1) - \sum_{k=0}^{n}\binom{n}{k}p^k(1-p)^{n-k}\left((1-p)\log(n+1-k)+p\log(k+1)\right).
\end{aligned} \tag{72}
$$

Let X be a binomial random variable $\mathrm{Bin}(n,p)$, then the above mutual information can be rewritten as

$$
I\left(W;U|\tilde{Z}_{[n+1]}\right) = \log(n+1) - \mathbb{E}\left[(1-p)\log(n+1-X)+p\log(X+1)\right]. \tag{73}
$$

Using the approximations in Lemma 7, we get

$$
\begin{aligned}
I\left(W;U|\tilde{Z}_{[n+1]}\right) &= \log(n+1) - p\left(\log(np+1)-\frac{np(1-p)}{2(np+1)^2\ln 2}\right) - (1-p)\left(\log(n+1-np)-\frac{np(1-p)}{2(n+1-np)^2\ln 2}\right) \\
&= p\log\frac{n+1}{np+1} + (1-p)\log\frac{n+1}{(1-p)n+1} + O\left(\frac{1}{n}\right) \\
&= -p\log p - (1-p)\log(1-p) + O\left(\frac{1}{n}\right) \\
&= H(p) + O\left(\frac{1}{n}\right),
\end{aligned} \tag{74}
$$

where we apply Lemma 9 in the third step. Finally, it is shown that the LOO bound converges to a constant $\sqrt{H(p)}$.

5) ICIMI Case: Akin to the LOO case, the mutual information can be computed as

$$
\begin{aligned}
I\left(W;R_i|\tilde{Z}_i,\tilde{Z}_{i+n}\right) =& \mathbb{E}_{W,\tilde{Z}_i,\tilde{Z}_{i+n},R_i}\left[\log\frac{P_{W|R_i,\tilde{Z}_i,\tilde{Z}_{i+n}}}{P_{W|\tilde{Z}_i,\tilde{Z}_{i+n}}}\right] \\
=& \mathbb{E}_{W,\tilde{Z}_i,\tilde{Z}_{i+n},R_i}\left[\log\frac{\binom{n-1}{k-\tilde{Z}_{i+R_in}}p^{k-\tilde{Z}_{i+R_in}}(1-p)^{n-1-k+\tilde{Z}_{i+R_in}}}{\frac{1}{2}\binom{n-1}{k-\tilde{Z}_i}p^{k-\tilde{Z}_i}(1-p)^{n-1-k+\tilde{Z}_i}+\frac{1}{2}\binom{n-1}{k-\tilde{Z}_{i+n}}p^{k-\tilde{Z}_{i+n}}(1-p)^{n-1-k+\tilde{Z}_{i+n}}}\right] \\
=& \mathbb{E}_{W,\tilde{Z}_i,\tilde{Z}_{i+n},R_i}\left[\log\frac{2\binom{n-1}{k-\tilde{Z}_{i+R_in}}p^{-\tilde{Z}_{i+R_in}}(1-p)^{\tilde{Z}_{i+R_in}}}{\binom{n-1}{k-\tilde{Z}_i}p^{-\tilde{Z}_i}(1-p)^{\tilde{Z}_i}+\binom{n-1}{k-\tilde{Z}_{i+n}}p^{-\tilde{Z}_{i+n}}(1-p)^{\tilde{Z}_{i+n}}}\right] \\
=& \frac{1}{2}\sum_{k=0}^{n-1}\binom{n-1}{k}p^{k+1}(1-p)^{n-k}\log\frac{2}{\frac{k}{n-k}\cdot\frac{1-p}{p}+1} + \frac{1}{2}\sum_{k=1}^{n}\binom{n-1}{k-1}p^k(1-p)^{n+1-k}\log\frac{2}{\frac{n-k}{k}\cdot\frac{p}{1-p}+1} \\
&+ \frac{1}{2}\sum_{k=1}^{n}\binom{n-1}{k-1}p^k(1-p)^{n+1-k}\log\frac{2}{\frac{n-k}{k}\cdot\frac{p}{1-p}+1} + \frac{1}{2}\sum_{k=0}^{n-1}\binom{n-1}{k}p^{k+1}(1-p)^{n-k}\log\frac{2}{\frac{k}{n-k}\cdot\frac{1-p}{p}+1} \\
=& 2p(1-p)\log 2 - \sum_{k=0}^{n-1}\binom{n-1}{k}p^{k+1}(1-p)^{n-k}\log\left(\frac{k}{n-k}\cdot\frac{1-p}{p}+1\right)\left(\frac{n-k-1}{k+1}\cdot\frac{p}{1-p}+1\right) \\
=& p(1-p)\left(2-\mathbb{E}\left[\log\left(\frac{X}{n-X}\cdot\frac{1-p}{p}+1\right)\left(\frac{n-X-1}{X+1}\cdot\frac{p}{1-p}+1\right)\right]\right) \\
=& p(1-p)\left(2-\mathbb{E}\left[\log\left(\frac{\alpha X}{\bar{X}+1}+1\right)\right]-\mathbb{E}\left[\left(\frac{\beta\bar{X}}{X+1}+1\right)\right]\right),
\end{aligned} \tag{75}
$$

where $X \sim \mathrm{Bin}(n-1,p)$, $\alpha = \frac{1-p}{p}$ and $\beta = \frac{p}{1-p}$. Without generality, we assume $p < \frac{1}{2}$ and $\alpha > 1$, $\beta < 1$. Using Lemma 7, one can approximate the above expression as

$$
\begin{aligned}
I\left(W;R_i|\tilde{Z}_i,\tilde{Z}_{i+n}\right) =& p(1-p)\left(2-\mathbb{E}\left[\log\frac{(\alpha-1)X+n}{\bar{X}+1}+\log\frac{n-(1-\beta)\bar{X}}{X+1}\right]\right)\\
=& p(1-p)\left(2-\mathbb{E}\left[\log(\alpha-1)+\log\left(X+\frac{n}{\alpha-1}\right)-\log\left(\bar{X}+1\right)\right.\right.\\
&\left.\left.+\log(1-\beta)+\log\left(\frac{n}{1-\beta}-\bar{X}\right)-\log\left(X+1\right)\right]\right)\\
=& p(1-p)\left(2-\log(\alpha-1)(1-\beta)-\log\left((n-1)p+\frac{n}{\alpha-1}\right)+\frac{(n-1)(1-p)p}{2\left((n-1)p+\frac{n}{\alpha-1}\right)^2\ln 2}\right.\\
&+\log\left((n-1)(1-p)+1\right)-\frac{(n-1)(1-p)p}{2\left((n-1)(1-p)+1\right)^2\ln 2}\\
&-\log\left(\frac{n}{1-\beta}-(n-1)(1-p)\right)+\frac{(n-1)(1-p)p}{2\left(\frac{n}{1-\beta}-(n-1)(1-p)\right)^2\ln 2}\\
&\left.+\log\left((n-1)p+1\right)-\frac{(n-1)(1-p)p}{2\left((n-1)p+1\right)^2\ln 2}+O\left(\frac{1}{n^2}\right)\right)\\
=& p(1-p)\left(2-\log(\alpha-1)(1-\beta)+\log\frac{(n-1)(1-p)+1}{(n-1)p+\frac{n}{\alpha-1}}+\log\frac{(n-1)p+1}{\frac{n}{1-\beta}-(n-1)(1-p)}+O\left(\frac{1}{n}\right)\right)\\
=& p(1-p)\left(2-\log(\alpha-1)(1-\beta)+\log\frac{1-p}{p+\frac{1}{\alpha-1}}+\log\frac{p}{\frac{1}{1-\beta}-(1-p)}+O\left(\frac{1}{n}\right)\right)\\
=& p(1-p)\left(2+\log\frac{1-p}{p(\alpha-1)+1}+\frac{p}{1-(1-p)(1-\beta)}+O\left(\frac{1}{n}\right)\right)\\
=& p(1-p)\left(2+\log\frac{1-p}{2-2p}+\log\frac{p}{2p}+O\left(\frac{1}{n}\right)\right)\\
=& O\left(\frac{1}{n}\right).
\end{aligned}
\tag{76}
$$

Thus, the ICIMI bound under the Bernoulli case is of order $O\left(\frac{1}{\sqrt{n}}\right)$.

6) LmO-CMI Case: Generalized from the LOO case, the learner in the LmO case is $W = \frac{1}{n}\sum_{i=1}^n \tilde{Z}_{U_i}$. Consider realizations $w = \frac{x}{n}$ and $\sum_{i=1}^{n+m}\tilde{z}_i - \sum_{i=1}^{n}\tilde{z}_{u_i} = y$, then the mutual information is given by

$$
\begin{aligned}
I\left(W;U_{[n]}|\tilde{Z}_{[n+m]}\right) &= \mathbb{E}_{W,\tilde{Z}_{[n+m]},U_{[n]}}\left[\log\frac{P_{W|\tilde{Z}_{[n+m]},U_{[n]}}}{P_{W|\tilde{Z}_{[n+m]}}}\right]\\
&= \sum_{x=0}^{n}\sum_{y=0}^{m}\frac{\binom{x+y}{x}\binom{n+m-x-y}{n-x}}{\binom{n+m}{n}}\binom{n+m}{x+y}p^{x+y}(1-p)^{n+m-x-y}\log\frac{\binom{n+m}{n}}{\binom{x+y}{x}\binom{n+m-x-y}{n-x}}\\
&= \log\binom{n+m}{n}-\sum_{x=0}^{n}\sum_{y=0}^{m}\binom{n}{x}p^x(1-p)^{n-x}\binom{m}{y}p^y(1-p)^{m-y}\log\binom{x+y}{x}\binom{n+m-x-y}{n-x}.
\end{aligned}
\tag{77}
$$

Let X and Y be two binomial random variables following $X \sim \mathrm{Bin}(n,p)$ and $Y \sim \mathrm{Bin}(m,p)$, respectively, then the above mutual information can be rewritten as

$$
I\left(W;U_{[n]}|\tilde{Z}_{[n+m]}\right) = \log\binom{n+m}{n} - \mathbb{E}\left[\log\binom{X+Y}{X}\binom{\bar{X}+\bar{Y}}{\bar{X}}\right], \tag{78}
$$

where $\bar{X} = n - X$ and $\bar{Y} = m - Y$.

Next, we will approximate each binomial coefficient term with a slight abuse of Lemma 10. Specifically, we have

$$\log\binom{n+m}{n} \leq \frac{1}{2}\log\frac{n+m}{nm} - \frac{1}{2}\log 2\pi + (n+m)H\left(\frac{n}{n+m}\right). \tag{79}$$

Then, for the expected binomial coefficients, we get

$$\begin{aligned}
\mathbb{E}\left[\log\binom{X+Y}{X}\right] \geq & \mathbb{E}\left[\frac{1}{2}\left(\log(X+Y) - \log X - \log Y - c\right) + X\log\frac{X+Y}{X} + Y\log\frac{X+Y}{Y}\right] \\
= & \mathbb{E}\left[\frac{1}{2}\left(\log Z - \log X - \log Y - c\right) + Z\log Z - X\log X - Y\log Y\right] \\
= & \frac{1}{2}\left(\log(n+m)p - \frac{1-p}{2(n+m)p\ln 2} + O\left(\frac{1}{(n+m)^2}\right)\right. \\
& \left. - \log np + \frac{1-p}{2np\ln 2} + O\left(\frac{1}{n^2}\right) - \log mp + \frac{1-p}{2mp\ln 2} + O\left(\frac{1}{m^2}\right)\right) \\
& + (n+m)p\log(n+m)p + \frac{1-p}{2\ln 2} + \frac{1-p^2}{12(n+m)p\ln 2} + O\left(\frac{1}{(n+m)^2}\right) \\
& - np\log np - \frac{1-p}{2\ln 2} - \frac{1-p^2}{12np\ln 2} + O\left(\frac{1}{n^2}\right) - mp\log np - \frac{1-p}{2\ln 2} - \frac{1-p^2}{12mp\ln 2} + O\left(\frac{1}{m^2}\right) \\
= & \frac{1}{2}\log\frac{n+m}{nm} + np\log\frac{n+m}{n} + mp\log\frac{n+m}{m} - \frac{\log p + (1-p)\log e}{2} \\
& + \frac{(p-1)(p-2)}{12p\ln 2}\left(\frac{1}{n} + \frac{1}{m} - \frac{1}{n+m}\right) + O\left(\frac{1}{(n+m)^2}\right) + O\left(\frac{1}{n^2}\right) + O\left(\frac{1}{m^2}\right).
\end{aligned} \tag{80}$$

Likewise, it holds that

$$\begin{aligned}
\mathbb{E}\left[\log\binom{\bar{X}+\bar{Y}}{\bar{X}}\right] \geq & \frac{1}{2}\log\frac{n+m}{nm} + n(1-p)\log\frac{n+m}{n} + m(1-p)\log\frac{n+m}{m} - \frac{\log(1-p) + p\log e}{2} \\
& + \frac{p(p+1)}{12(1-p)\ln 2}\left(\frac{1}{n} + \frac{1}{m} - \frac{1}{n+m}\right) + O\left(\frac{1}{(n+m)^2}\right) + O\left(\frac{1}{n^2}\right) + O\left(\frac{1}{m^2}\right).
\end{aligned} \tag{81}$$

Then combine all the results, it follows that

$$\begin{aligned}
I\left(W; U_{[n]}|\tilde{Z}_{[n+m]}\right) \leq & \frac{1}{2}\log\frac{n+m}{nm} - \frac{1}{2}\log 2\pi + (n+m)H\left(\frac{n}{n+m}\right) \\
& - \frac{1}{2}\log\frac{n+m}{nm} - np\log\frac{n+m}{n} - mp\log\frac{n+m}{m} + \frac{\log p + (1-p)\log e}{2} \\
& - \frac{(p-1)(p-2)}{12p\ln 2}\left(\frac{1}{n} + \frac{1}{m} - \frac{1}{n+m}\right) + O\left(\frac{1}{(n+m)^2}\right) + O\left(\frac{1}{n^2}\right) + O\left(\frac{1}{m^2}\right) \\
& - \frac{1}{2}\log\frac{n+m}{nm} - n(1-p)\log\frac{n+m}{n} - m(1-p)\log\frac{n+m}{m} + \frac{\log(1-p) + p\log e}{2} \\
& - \frac{p(p+1)}{12(1-p)\ln 2}\left(\frac{1}{n} + \frac{1}{m} - \frac{1}{n+m}\right) + O\left(\frac{1}{(n+m)^2}\right) + O\left(\frac{1}{n^2}\right) + O\left(\frac{1}{m^2}\right) \\
= & \frac{1}{2}\log\frac{nm}{n+m} + \frac{\log\frac{p(1-p)e}{\pi} - 1}{2} - \frac{5p^2 - 5p + 2}{12p(1-p)\ln 2}\left(\frac{1}{n} + \frac{1}{m} - \frac{1}{n+m}\right) + O\left(\frac{1}{(n+m)^2}\right) + O\left(\frac{1}{n^2}\right) + O\left(\frac{1}{m^2}\right)
\end{aligned} \tag{82}$$

7) LOFO-IPCMI Case: Let X be the number of 1's in the test batch, i.e., $X = \sum_{i=1}^{\frac{n}{k}} \tilde{Z}_i^{(U)}$, where $0 \leq x \leq k$. Since each batch sample is Bernoulli distributed, then X is a binomial random variable and its PMF at $X = x$, denoted by q_x, is a Bernstein polynomial expressed as

$$q_x = \binom{k}{x} p^x (1-p)^{k-x}. \tag{83}$$

Furthermore, let y be the number of training batches, in each of which the number of 1's is x, where $0 \leq y \leq \frac{n}{k}$. We witness each such training batch as a "successful trial", and thus the joint probability is also a binomial PMF. Then, we can compute the mutual information as follows:

$$\begin{aligned}
I\left(W; U|\tilde{Z}_{[n+k]}\right) = & \sum_{x=0}^{k} q_x \sum_{y=0}^{\frac{n}{k}} \binom{\frac{n}{k}}{y} q_x^y (1-q_x)^{\frac{n}{k}-y} \log\frac{\frac{n}{k}+1}{y+1} \\
= & \log\left(\frac{n}{k}+1\right) - \mathbb{E}\left[\mathbb{E}\left[\log(Y+1)|X\right]\right]
\end{aligned} \tag{84}$$

where $X \sim \mathrm{Bin}(k,p)$ and $Y \sim \mathrm{Bin}\left(\frac{n}{k}, q_x\right)$.

8) LOSFO-IPCIMI Case:

$$I\left(W; U^{(i)}|\tilde{Z}^{(i)}_{[\frac{n}{m}+1]}\right) = \log\left(\frac{n}{m}+1\right) - p\mathbb{E}\left[\log\left(Y+1+\frac{\alpha X\bar{Y}}{\bar{X}+1}\right)\right] - (1-p)\mathbb{E}\left[\log\left(\bar{Y}+1+\frac{\beta\bar{X}Y}{X+1}\right)\right], \tag{85}$$

where $X \sim \mathrm{Bin}\left(n-\frac{n}{m},p\right)$, $Y \sim \mathrm{Bin}\left(\frac{n}{m},p\right)$, $\bar{X} = n - \frac{n}{m} - X \sim \mathrm{Bin}\left(n-\frac{n}{m},1-p\right)$, $\bar{Y} = \frac{n}{m} - Y \sim \mathrm{Bin}\left(\frac{n}{m},1-p\right)$, $\alpha = \frac{1-p}{p}$ and $\beta = \frac{p}{1-p}$. Through the above expression, One can validate the argument that the PLOO-CMI bound is a general case of the LOO-CMI and ICIMI bounds. With detailed calculations omitted, we note that letting $m=1$ reduces (85) to the LOO-CMI bound in (73), whereas setting $m=n$ leads to the ICIMI bound shown in (75). In the following, we continue to derive the order of (85).

Let $f(X,Y)$ be any deterministic function of X and Y. By multivariate Taylor expansion, we approximate $f(X,Y)$ at points $X=\mathbb{E}[X]$ and $Y=\mathbb{E}[Y]$:

$$\begin{aligned} f(X,Y) \approx & f\left(\mathbb{E}[X],\mathbb{E}[Y]\right) + \frac{\partial f}{\partial X}\left(\mathbb{E}[X],\mathbb{E}[Y]\right)\left(X-\mathbb{E}[X]\right) + \frac{\partial f}{\partial Y}\left(\mathbb{E}[X],\mathbb{E}[Y]\right)\left(Y-\mathbb{E}[Y]\right) + \frac{\partial^2 f}{\partial X^2}\left(\mathbb{E}[X],\mathbb{E}[Y]\right)\frac{\left(X-\mathbb{E}[X]\right)^2}{2} \\ & + \frac{\partial^2 f}{\partial X \partial Y}\left(\mathbb{E}[X],\mathbb{E}[Y]\right)\left(X-\mathbb{E}[X]\right)\left(Y-\mathbb{E}[Y]\right) + \frac{\partial^2 f}{\partial Y^2}\left(\mathbb{E}[X],\mathbb{E}[Y]\right)\frac{\left(Y-\mathbb{E}[Y]\right)^2}{2}. \end{aligned} \tag{86}$$

Take expectation on both sides and use the independence of X and Y, it follows that

$$\begin{aligned} \mathbb{E}\left[f(X,Y)\right] \approx & f\left(\mathbb{E}[X],\mathbb{E}[Y]\right) + \frac{\partial^2 f}{\partial X^2}\left(\mathbb{E}[X],\mathbb{E}[Y]\right)\frac{\mathbb{E}\left[\left(X-\mathbb{E}[X]\right)^2\right]}{2} + \frac{\partial^2 f}{\partial Y^2}\left(\mathbb{E}[X],\mathbb{E}[Y]\right)\frac{\mathbb{E}\left[\left(Y-\mathbb{E}[Y]\right)^2\right]}{2} \\ = & f\left(\mathbb{E}[X],\mathbb{E}[Y]\right) + \frac{\partial^2 f}{\partial X^2}\left(\mathbb{E}[X],\mathbb{E}[Y]\right)\frac{m-1}{2m}np(1-p) + \frac{\partial^2 f}{\partial Y^2}\left(\mathbb{E}[X],\mathbb{E}[Y]\right)\frac{n}{2m}p(1-p). \end{aligned} \tag{87}$$

On the one hand, by letting $f_1(X,Y) = \log\left(Y+1+\frac{\alpha X\bar{Y}}{\bar{X}+1}\right)$ (where $\bar{X}$ and $\bar{Y}$ are deterministic functions of X and Y, respectively), we now compute $f_1\left(\mathbb{E}[X],\mathbb{E}[Y]\right)$ and prove that both $\frac{\partial^2 f_1}{\partial X^2}\left(\mathbb{E}[X],\mathbb{E}[Y]\right)$ and $\frac{\partial^2 f_1}{\partial Y^2}\left(\mathbb{E}[X],\mathbb{E}[Y]\right)$ are of order $O\left(\frac{1}{n^2}\right)$. First, plugging $X=\mathbb{E}[X]=\frac{m-1}{m}np$ and $Y=\mathbb{E}[Y]=\frac{n}{m}p$ into $f_1\left(X,Y\right)$ gives

$$\begin{aligned} f_1\left(\mathbb{E}[X],\mathbb{E}[Y]\right) = & \log\left(\frac{n}{m}p+1+\frac{(1-p)^2\frac{n}{m}\frac{m-1}{m}n}{(1-p)\frac{m-1}{m}n+1}\right) \\ = & \log\frac{(1-p)(m-1)n^2+((1-p)(m-1)+p)mn+m^2}{(1-p)(m-1)mn+k^2}. \end{aligned} \tag{88}$$

Next, we compute the following partial derivatives of f_1 as

$$\frac{\partial f_1}{\partial X}\left(X,Y\right) = \frac{\alpha\left(\frac{m-1}{m}n+1\right)Y}{\left(\bar{X}+1\right)^2\left(Y+1\right)+\alpha\left(\bar{X}+1\right)XY}, \tag{89}$$

$$\frac{\partial^2 f_1}{\partial X^2}\left(X,Y\right) = \frac{\alpha\left(\frac{m-1}{m}n+1\right)\left(\alpha XY+(2-\alpha)\bar{X}Y+2\bar{X}+(2-\alpha)Y+2\right)Y}{\left(\left(\bar{X}+1\right)^2\left(Y+1\right)+\alpha\left(\bar{X}+1\right)XY\right)^2}, \tag{90}$$

$$\frac{\partial f_1}{\partial Y}\left(X,Y\right) = \frac{\bar{X}-\alpha X+1}{\left(\bar{X}+1\right)\left(Y+1\right)+\alpha X\bar{Y}}, \tag{91}$$

$$\frac{\partial^2 f_1}{\partial Y^2}\left(X,Y\right) = -\frac{\left(\bar{X}-\alpha X+1\right)^2}{\left(\left(\bar{X}+1\right)\left(Y+1\right)+\alpha X\bar{Y}\right)^2}. \tag{92}$$

At points $X=\frac{m-1}{m}np$, $\bar{X}=\frac{m-1}{m}n(1-p)$, $Y=\frac{n}{m}p$ and $\bar{Y}=\frac{n}{m}(1-p)$, one can verify that

$$\frac{\partial f_1}{\partial X}\left(\mathbb{E}[X],\mathbb{E}[Y]\right) = O\left(\frac{1}{n}\right), \qquad \frac{\partial^2 f_1}{\partial X^2}\left(\mathbb{E}[X],\mathbb{E}[Y]\right) = O\left(\frac{1}{n^2}\right), \tag{93}$$

$$\frac{\partial f_1}{\partial X}\left(\mathbb{E}[X],\mathbb{E}[Y]\right) = O\left(\frac{1}{n}\right), \qquad \frac{\partial^2 f_1}{\partial X^2}\left(\mathbb{E}[X],\mathbb{E}[Y]\right) = O\left(\frac{1}{n^2}\right). \tag{94}$$

Plugging (88), (93) and (94) back into (87) yields

$$\mathbb{E}\left[f_1(X,Y)\right] = \log\frac{(1-p)(m-1)n^2+((1-p)(m-1)+p)mn+m^2}{(1-p)(m-1)mn+m^2} + O\left(\frac{1}{n}\right). \tag{95}$$

On the other, we let $f_2(X,Y)=\log\left(\bar{Y}+1+\frac{\beta\bar{X}Y}{X+1}\right)$ and follow a similar procedure to obtain

$$\mathbb{E}\left[f_2(X,Y)\right]=\log\frac{p(m-1)n^2+(p(m-1)+1-p)mn+m^2}{p(m-1)mn+m^2}+O\left(\frac{1}{n}\right). \tag{96}$$

Combining all the results yields

$$\begin{aligned}I\left(W;U^{(i)}|\tilde{Z}^{(i)}_{[\frac{n}{m}+1]}\right)=&\log\left(\frac{n}{m}+1\right)-p\mathbb{E}\left[f_1(X,Y)\right]-(1-p)\mathbb{E}\left[f_2(X,Y)\right]\\ =&p\left(\log\left(\frac{n}{m}+1\right)-\log\frac{(1-p)(m-1)n^2+((1-p)(m-1)+p)mn+m^2}{(1-p)(m-1)mn+m^2}\right)\\ &+(1-p)\left(\log\left(\frac{n}{m}+1\right)-\log\frac{p(m-1)n^2+(p(m-1)+1-p)mn+m^2}{p(m-1)mn+m^2}\right)+O\left(\frac{1}{n}\right)\\ =&p\log\left(1+\frac{(1-p)mn}{(1-p)(m-1)n^2+((1-p)(m-1)+p)mn+m^2}\right)\\ &+(1-p)\log\left(1+\frac{pmn}{p(m-1)n^2+(p(m-1)+1-p)mn+m^2}\right)+O\left(\frac{1}{n}\right)\\ =&\frac{p(1-p)mn}{(1-p)(m-1)n^2+((1-p)(m-1)+p)mn+m^2}+\frac{p(1-p)mn}{p(m-1)n^2+(p(m-1)+1-p)mn+m^2}+O\left(\frac{1}{n}\right)\\ =&O\left(\frac{1}{n}\right),\end{aligned} \tag{97}$$

plugging which into (23) implies that PLOO-CMI bound scales as $O\left(\frac{1}{\sqrt{n}}\right)$.

9) (m,n)-IPCIMI Case:

$$I\left(W;U^{(i)}|\tilde{Z}^{(i)}_{[\frac{m}{n}+1]}\right)=\log\left(\frac{m}{n}+1\right)-p\mathbb{E}\left[\log\left(Y+1+\frac{\beta\bar{X}\bar{Y}}{X+1}\right)\right]-(1-p)\mathbb{E}\left[\log\left(\bar{Y}+1+\frac{\alpha XY}{\bar{X}+1}\right)\right], \tag{98}$$

where $X\sim\mathrm{Bin}\left(n-1,p\right)$, $Y\sim\mathrm{Bin}\left(\frac{m}{n},p\right)$, $\bar{X}=n-1-X\sim\mathrm{Bin}\left(n-1,1-p\right)$, $\bar{Y}=\frac{m}{n}-Y\sim\mathrm{Bin}\left(\frac{m}{n},1-p\right)$, $\alpha=\frac{1-p}{p}$ and $\beta=\frac{p}{1-p}$.