A Runtime of ERS.

We provide an analysis of ERS runtime. Let $\omega = \max_{x,y} P_{Y|X}(y|x)/Q(y)$ and $\omega_x = \max_y P_{Y|X}(y|x)/Q(y)$, where $P_{Y|X=x}$ is the target distribution and Q_Y is the proposal distribution. For the batch size N and input x, we have the following bound on the average batch acceptance probability Δ_x, which we will show in Appendix E.1:

$$\Delta_x \geq \frac{N}{N-1+\omega_x} \geq \frac{N}{N-1+\omega}, \tag{14}$$

Thus, the expected number of batches in ERS is:

$$\text{Expected Number of Batches} = \frac{1}{\Delta_x} \leq \frac{N-1+\omega}{N}, \tag{15}$$

which leads to the runtime, i.e. the expected number of proposals as:

$$\text{Expected Runtime} = \frac{N}{\Delta_x} \leq N-1+\omega. \tag{16}$$

In practice, since we typically choose $N = O(\omega)$, the expected runtime is also $O(\omega)$.

B Coding Cost of Standard Rejection Sampling

For the proof, we generalize and use $P(.)$ and $Q(.)$ as the target and proposal distributions. This allows shorthand the notations while also generalizing the results for arbitrary distributions.

B.1 Extension of Braverman and Garg [4]'s Method for Continuous Setting

This method is an extension of the work by Braverman and Garg [4] to the continuous setting. The core idea is to divide the acceptance region into smaller bins, visualized in Figure 7. Specifically, for each pair (U_i, Y_i) from W, we denote $\tilde{U}_i = \omega U_i Q(Y_i)$. The encoder selects the index K according to rejection sampling rule, which is 7 in Figure 7. It then sends the bin index of the first accepted sample, where the bin corresponds to the smallest scaled region that $\tilde{U}_K$ belongs to. In Figure 7, this corresponds to the orange region and the content of the message is 3. Then the encoder sends another message which indicates the rank of the selected sample within that bin, which is 1. The decoder then K accordingly. Formally, the two steps are as follow:

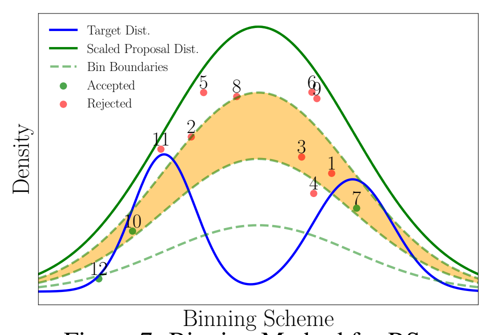

Figure 7: Binning Method for RS.

- *Binning*: The encoder sends to the decoder the ceiling $T = \lceil \frac{\tilde{U}_K}{Q(Y_K)} \rceil$. Upon receiving T, the decoder collects the set:

$$\mathcal{S}_T = \{i | (T-1)Q(Y_i) \leq \tilde{U}_i \leq TQ(Y_i)\}, \tag{17}$$

- *Index Selection*: The encoder locates the original chosen index K within $\mathcal{S}_T$, says G, and send G to the receiver. We have $\mathbb{E}[\log G] \leq 1$.

Binning Step. We will show the $\mathbb{E}[\log T] \leq D_{KL}(P||Q) + \log(e)$., adapting the proof for the discrete case presented in [33]. First, we note that:

$$Y_K \sim P(.), \quad U_K | Y_K \sim \mathcal{U}\left(0, \frac{P(Y_K)}{\omega Q(Y_K)}\right) \tag{18}$$

We then have:

$$
\begin{aligned}
\mathbb{E}[\log T] &= \mathbb{E}\left[\log\left(\left\lceil \frac{\tilde{U}_K}{Q(Y_K)} \right\rceil\right)\right] && (19)\\
&\leq \mathbb{E}\left[\log\left(1+\frac{\tilde{U}_K}{Q(Y_K)}\right)\right] && (20)\\
&= \mathbb{E}\left[\log\left(1+\omega U_K\right)\right] && (21)\\
&= \mathbb{E}\left[\mathbb{E}\left[\log\left(1+\omega U_K\right)|Y_K\right]\right] && (22)\\
&= \int_{-\infty}^{+\infty} P(y)\left[\frac{\omega Q(y)}{P(y)}\int_0^{\omega^{-1}P(y)/Q(y)} \log\left(1+\omega u\right)du\right]dy \ \ \text{(Due to (18))} && (23)\\
&\leq \int_{-\infty}^{+\infty} P(y)\left[\frac{\omega Q(y)}{P(y)}\int_0^{\omega^{-1}P(y)/Q(y)} \log\left(1+\frac{P(y)}{Q(y)}\right)du\right]dy && (24)\\
&\leq \int_{-\infty}^{+\infty} P(y)\left[\frac{\omega Q(y)}{P(y)}\int_0^{\omega^{-1}P(y)/Q(y)} \log\left(\frac{P(y)}{Q(y)}\right)+\frac{Q(y)\log(e)}{P(y)}du\right]dy && (25)\\
&= \int_{-\infty}^{+\infty} P(y)\log\left(\frac{P(y)}{Q(y)}\right)dy + \int_{-\infty}^{+\infty} Q(y)\log(e)dy && (26)\\
&= D_{KL}(P||Q)+\log(e), && (27)
\end{aligned}
$$

where we use the following results for the last inequality:

$$
\log(1+x) \leq \log(x) + \frac{\log(e)}{x} \quad \text{(for all } x > -1). \tag{28}
$$

Index Selection Step. We first show that $\mathbb{E}[G] \leq 2$ by using recursion. We define $\mathcal{A}$ as an event where the first samples is accepted, i.e. $U_1 \leq \frac{P(Y_1)}{\omega Q(Y_1)}$. Then, if $\mathcal{A}$ happens then we have $G = 1$, i.e. $\mathbb{E}[G|A] = 1$, since it is also the first sample in $\mathcal{S}_T$.

Before proceeding to the case where $\mathcal{A}$ does not happen, i.e. $\bar{\mathcal{A}}$, we define the following random variable $M = \mathbb{1}[1 \in \mathcal{S}_T]$, i.e. $M = 1$ if the first proposed sample from W stays within the ceiling $(T-1)Q(Y_1) \leq \tilde{U}_1 \leq TQ(Y_1)$ and $M = 0$ otherwise.

Then we have the two following recursion identities:

$$
\begin{cases}
\mathbb{E}[G|\bar{A}, M=0] = \mathbb{E}[G] \\
\mathbb{E}[G|\bar{A}, M=1] = 1 + \mathbb{E}[G]
\end{cases} \tag{29}
$$

For the first equality, given that the first sample U_1, Y_1 does not stay within $\mathcal{S}_T$ does not implies any information about G, since all the samples are i.i.d. For the second equality takes into account the fact that we now accept the first sample (U_1, Y_1) and repeat the counting process. Hence, we have:

$$
\begin{aligned}
\mathbb{E}[G|\bar{A}] &= \Pr(M=0|\bar{A})\mathbb{E}[G|\bar{A}, M=0] + \Pr(M=1|\bar{A})\mathbb{E}[G|\bar{A}, M=1] && (30)\\
&= \mathbb{E}[G] + \Pr(M=1|\bar{A}) && (31)
\end{aligned}
$$

We now express $\mathbb{E}[G]$ as follows:

$$
\begin{aligned}
\mathbb{E}[G] &= \Pr(A)\mathbb{E}[G|A] + \Pr(\bar{A})\mathbb{E}[G|\bar{A}] && (32)\\
&= \Pr(A) + \Pr(\bar{A})(\mathbb{E}[G] + \Pr(M=1|\bar{A})) && (33)
\end{aligned}
$$

Rearranging the terms, we obtain:

$$
\mathbb{E}[G] = 1 + \frac{\Pr(M=1, \bar{A})}{\Pr(A)} \tag{34}
$$

We have $\Pr(A) = \int_{-\infty}^{\infty} \omega^{-1} P(y) Q^{-1}(y) Q(y) dy = \omega^{-1}$. For $\Pr(M = 1, \bar{A})$, we have:

$$
\begin{aligned}
\Pr(M = 1, \bar{A}) &\leq \Pr(M = 1) && (35)\\
&= \sum_{t=0}^{\infty} \Pr((T-1)Q(Y_1) \leq \tilde{U}_1 \leq (T-1)Q(y), T = t) && (36)\\
&= \sum_{t=0}^{\infty} \Pr((t-1)Q(Y_1) \leq \tilde{U}_1 \leq (t-1)Q(y)) \Pr(T = t) && (37)\\
&= \sum_{t=0}^{\infty} \omega^{-1} \Pr(T = t) && (38)\\
&= \omega^{-1} && (39)
\end{aligned}
$$

Thus, we obtain $\mathbb{E}[G] \leq 2$ and hence $\mathbb{E}[\log G] \leq 1$.

B.2 The Sorting Method

The encoding process is as follows:

- *Grouping*: the encoder sends the ceiling $L = \lceil \frac{K}{\lfloor \omega \rfloor} \rceil$ to the decoder. The decoder then knows $(L-1)\omega + 1 \leq K \leq L\omega$, i.e. K is in range L. We have $\mathrm{E}[\log L] = 1$ bit.
- *Sorting*: The encoder and decoder both sort the uniform random variables U_i within the selected range $(L-1)\lfloor \omega \rfloor + 1 \leq i \leq L\lfloor \omega \rfloor$. Let the sorted list be $U_{\pi(1)} \leq U_{\pi(2)} \leq ... \leq U_{\pi(\lfloor \omega \rfloor)}$ where $\pi(.)$ is the mapping between the sorted index and the original unsorted one. The encoder sends the rank of U_K within this list, i.e. sends the value $\hat{K}$ such that $K = \pi(\hat{K})$. The decoder receive $\hat{K}$ and retrieve Y_K accordingly. The coding cost for this step is $D_{KL}(P||Q) + \log(e)$.

We provide detail analysis for each step below:

Grouping Step. Since each proposal is accepted with probability ω^{-1}, this means:

$$\Pr(K > \ell \lfloor \omega \rfloor) = \left(1 - \omega^{-1}\right)^{\ell \lfloor \omega \rfloor} < \left(\frac{1}{2}\right)^{\ell}, \tag{40}$$

where we will prove the RHS inequality in Appendix B.3. Hence, we have $\Pr(L > \ell) < \left(\frac{1}{2}\right)^{-\ell}$ and:

$$\mathbb{E}[L] = \sum_{\ell=0}^{\infty} \Pr(L > \ell) < 1 + 0.5^{-1} + 0.5^{-2} + ... = 2. \tag{41}$$

Finaly, using Jensen's inequality, we have:

$$\mathbb{E}[\log L] \leq \log(\mathbb{E}[L]) = 1. \tag{42}$$

Sorting Step. To bound the coding cost in step 2, we first express $\mathbb{E}[\log \hat{K}]$ with the rule of conditional expectation as follows:

$$
\begin{aligned}
\mathbb{E}[\log \hat{K}] &= \int_{-\infty}^{\infty} P(y) \mathbb{E}[\log \hat{K} | Y_K = y] dy && (43)\\
&= \int_{-\infty}^{\infty} P(y) \left(\int_{-\infty}^{\infty} \mathbb{E}[\log \hat{K} | Y_K = y, U_K = u] P(U_K = u | Y_K = y) du \right) dy && (44)\\
&= \int_{-\infty}^{\infty} P(y) \left(\int_{0}^{\frac{P(y)}{\omega Q(y)}} \mathbb{E}[\log \hat{K} | Y_K = y, U_K = u] \frac{\omega Q(y)}{P(y)} du \right) dy, && (45)
\end{aligned}
$$

where the last step, $P(U_K = u | Y_K = y) = \frac{\omega Q(y)}{P(y)}$ for $0 \leq u \leq \frac{P(y)}{\omega Q(y)}$ is due to the acceptance condition in rejection sampling. We will show in Section B.3.1 that:

$$\mathbb{E}[\log \hat{K} | Y_K = y, U_K = u] \leq \log(\omega u + 1) \tag{46}$$

Then, combining this with Equation (45), we obtain:

$$\begin{aligned}
\mathbb{E}[\log \hat{K}] &\leq \int_{-\infty}^{\infty} P(y) \left(\int_0^{\frac{P(y)}{\omega Q(y)}} \frac{\omega Q(y)}{P(y)} \log(\omega u + 1) du \right) dy && (47)\\
&\leq \int_{-\infty}^{\infty} P(y) \left(\int_0^{\frac{P(y)}{\omega Q(y)}} \frac{\omega Q(y)}{P(y)} \log\left(\frac{P(y)}{Q(y)} + 1\right) du \right) dy && (48)\\
&= \int_{-\infty}^{\infty} P(y) \left[\frac{P(y)}{\omega Q(y)} \frac{\omega Q(y)}{P(y)} \log\left(\frac{P(y)}{Q(y)} + 1\right) \right] dy && (49)\\
&= \int_{-\infty}^{\infty} P(y) \log\left(\frac{P(y)}{Q(y)} + 1\right) dy && (50)\\
&\leq \int_{-\infty}^{\infty} P(y) \left[\log\left(\frac{P(y)}{Q(y)}\right) + \frac{\log(e) Q(y)}{P(y)} \right] dy && (51)\\
&= D_{KL}(P||Q) + \log(e) && (52)
\end{aligned}$$

Hence, we have $\mathbb{E}[\log \hat{K}] \leq D_{KL}(P||Q) + \log(e)$ on average.

B.3 Proof for Inequality (40)

The proof for this inequality is self-contained. We want to prove that for any $\omega \geq 1$, we have:

$$f(\omega) = (1 - \omega^{-1})^{\lfloor \omega \rfloor} \leq \frac{1}{2}. \tag{53}$$

Consider the behavior of $f(\omega)$ at every interval $[n, n+1)$ where $n \in \mathbb{Z}^+, n \geq 1$. Since $\omega \geq 1$, the function $f_n(\omega) = \left(1 - \omega^{-1}\right)^n$ is increasing and hence:

$$\sup_\omega f_n(\omega) = \left(1 - \frac{1}{n+1}\right)^n = \left(\frac{n}{n+1}\right)^n$$

for every interval $[n, n+1)$. We will show that $\sup_\omega f_n(\omega)$ is decreasing for $n \geq 1$ and thus we have $\sup_\omega f(\omega) = \sup_\omega f_1(\omega) = \frac{1}{2}$.

Consider the function $g(x) = \left(\frac{x}{x+1}\right)^n$ for $x \geq 1, x \in \mathbb{R}$. Let $h(x) = \ln(g(x)) = x \ln(\frac{x}{x+1})$, then we simply need to show $h(x)$ is decreasing. Consider its first derivative:

$$h'(x) = \ln\left(\frac{x}{x+1}\right) + \frac{1}{x+1} \leq 0, \tag{54}$$

since:

$$\ln\left(\frac{x}{x+1}\right) = \ln\left(1 - \frac{1}{x+1}\right) \leq -\frac{1}{x+1} \tag{55}$$

due to the inquality $\ln(1+y) < y$ for all y.

B.3.1 Proof for Inequality (46)

We begin by applying Jensen's inequality for concave function $\log(x)$:

$$\begin{aligned}
\mathbb{E}[\log \hat{K} | Y_K = y, U_K = u] &\leq \log \mathbb{E}[\hat{K} | Y_K = y, U_K = u] \quad \text{(by Jensen's Inequality)} && (56)\\
&= \log \mathbb{E}_L[\mathbb{E}[\hat{K} | Y_K = y, U_K = u, L = \ell]] && (57)
\end{aligned}$$

Given K is within the range $L = \ell$ and $U_K = u$, we can express $\hat{K}$ as follows:

$$\begin{aligned}
\hat{K} &= |\{U_i < u, (\ell - 1)\lfloor \omega \rfloor + 1 \leq i \leq \ell \lfloor \omega \rfloor\}| + 1, && (58)\\
&= \Omega(u, \ell) + 1 && (59)
\end{aligned}$$

i.e. the number of U_i (plus 1 for the ranking) within the range L that has value lesser than u.

We can see that the the index i within the range L satisfying $U_i < u$ are from the index that are either (1) rejected, i.e. index $i < K$ or (2) not examined by the algorithm, i.e. index $i > K$. The rest of this proof will show the following upperbound:

$$\mathbb{E}[\Omega(u,\ell)|Y_K = y, U_K = u, L = \ell] \leq \omega u, \text{ for any } \ell \tag{60}$$

For readability, we split the proof into different proof steps.

Proof Step 1: We condition on the mapped index of $\pi(\hat{K})$ on the original array:

$$\begin{aligned}
&\mathbb{E}[\hat{K}|Y_K = y, U_K = u, L = \ell] && (61)\\
&= \mathbb{E}_{\pi(\hat{K})}\left[\mathbb{E}[\hat{K} \mid Y_K = y, U_K = u, L = \ell, \pi(\hat{K}) = k]\right] && (62)\\
&= \mathbb{E}_{\pi(\hat{K})}\left[\mathbb{E}[\Omega(u,\ell) + 1 \mid Y_K = y, U_K = u, L = \ell, \pi(\hat{K}) = k]\right] && (63)\\
&= \mathbb{E}_{\pi(\hat{K})}\left[\mathbb{E}[\Omega(u,\ell) \mid Y_K = y, U_K = u, L = \ell, \pi(\hat{K}) = k]\right] + 1 && (64)\\
&= \mathbb{E}_{\pi(\hat{K})}\left[\mathbb{E}[\Omega_1(u,\ell,k) + \Omega_2(u,\ell,k) \mid Y_K = y, U_K = u, L = \ell, \pi(\hat{K}) = k]\right] + 1, && (65)
\end{aligned}$$

where $\Omega_1(u,\ell,k), \Omega_2(u,\ell,k)$ are the number of $U_i < u$ within the range $L = \ell$ that occurs before and after the selected index k respectively. Specifically:

$$\Omega_1(u,\ell,k) = |\{U_i < u, (\ell-1)\lfloor\omega\rfloor + 1 \leq i < (\ell-1)\lfloor\omega\rfloor + k\}| \tag{66}$$
$$\Omega_2(u,\ell,k) = |\{U_i < u, (\ell-1)\lfloor\omega\rfloor + k + 1 \leq i \leq \ell\lfloor\omega\rfloor\}|, \tag{67}$$

which also naturally gives $\Omega(u,\ell) = \Omega_1(u,\ell,k) + \Omega_2(u,\ell,k)$.

Proof Step 2: Consider $\Omega_2(u,\ell,k)$, since each proposal (Y_i, U_i) is i.i.d distributed and the fact that k is the index of the accepted sample, for every $i > K$, we have:

$$\Pr(U_i < u \mid Y_K = y, U_K = u, L = \ell, \pi(\hat{K}) = k) = \Pr(U_i < u)$$

This gives us:

$$\begin{aligned}
\mathbb{E}[\Omega_2(u,\ell,k) \mid Y_K = y, U_K = u, L = \ell, \pi(\hat{K}) = k] &= (\lfloor\omega\rfloor - k)\Pr(U < u) && (68)\\
&= (\lfloor\omega\rfloor - k)u && (69)\\
&\leq \frac{(\lfloor\omega\rfloor - k)u}{\Pr(\text{reject a sample})} && (70)\\
&\leq \frac{(\lfloor\omega\rfloor - k)u}{1 - \omega^{-1}} && (71)
\end{aligned}$$

Proof Step 3: For $\Omega_1(u,\ell,k)$, we do not have such independent property since for every sample with index $i < K$, we know that they are rejected samples, and hence for $i < k$:

$$\begin{aligned}
\Pr(U_i < u \mid Y_K = y, U_K = u, L = \ell, \pi(\hat{K}) = k) &= \Pr(U_i < u | Y_i \text{ is rejected}) && (72)\\
&= \frac{\Pr(U_i < u, Y_i \text{ is rejected})}{\Pr(Y_i \text{ is rejected})} && (73)\\
&\leq \frac{\Pr(U_i < u)}{\Pr(Y_i \text{ is rejected})} && (74)\\
&= \frac{u}{1 - \omega^{-1}}, && (75)
\end{aligned}$$

which gives us:

$$\mathbb{E}[\Omega_2(u,\ell,k) \mid Y_K = y, U_K = u, L = \ell, \pi(\hat{K}) = k] \leq \frac{(k-1)u}{1 - \omega^{-1}} \tag{76}$$

To prove Equation (72), note that the following events are equivalent:

$$\begin{aligned}
\{Y_K = y, U_K = u, L = \ell, \pi(\hat{K}) = k\} &= \{Y_k = y, U_k = u, Y_{1\ldots k-1} \text{ are rejected}\} && (77)\\
&\triangleq \Lambda(u, y, k) && (78)
\end{aligned}$$

Here, we note that Y_k, U_k denote the value at index k within W, which is different from Y_K, U_K, the value selected by the rejection sampler. Hence:

$$\Pr(U_i < u|\Lambda(u,y,k)) = \frac{\Pr(U_i < u, Y_{1\ldots k-1} \text{ are rejected}|Y_k = y, U_k = u)}{\Pr(Y_{1\ldots k-1} \text{ are rejected}|Y_k = y, U_k = u)} \tag{79}$$

$$= \frac{\Pr(U_i < u,\ Y_{1\ldots k-1} \text{ are rejected})}{\Pr(Y_{1\ldots k-1} \text{ are rejected})} \quad \text{(Since } (Y_i, U_i) \text{ are i.i.d)} \tag{80}$$

$$= \Pr(U_i < u|Y_i \text{ is rejected}), \tag{81}$$

Proof Step 4: From the above result from Step 2 and 3, we have $\Omega(u,\ell) = \Omega_1(u,\ell,k) + \Omega_2(u,\ell,k) \leq \omega u$ and as a result:

$$\mathbb{E}[K|Y_K = y, U_K = u, L = \ell] \leq \frac{(\lfloor\omega\rfloor - 1)u}{1-\omega^{-1}} + 1 \tag{82}$$

$$\leq \frac{(\omega - 1)u}{1-\omega^{-1}} + 1 \quad (\text{Since} \lfloor\omega\rfloor \leq \omega) \tag{83}$$

$$= \omega u + 1 \tag{84}$$

which completes the proof.

B.4 Overall Coding Cost.

We now provide the upperbound on $H[K]$ for our *Sorting Method*. Since the message in the *Binning Method* also consists of two parts, the results are the same. For each part of the message, namely L and K, we encode it with a prefix-code from Zipf distribution [25]. For $H[L]$, we have:

$$H[L] \leq \mathrm{E}_X[\mathrm{E}[\log L|X = x]] + \log(\mathrm{E}_X[\mathrm{E}[\log L|X = x]] + 1) + 1 \tag{85}$$

$$= 3 \text{ bits} \tag{86}$$

Hence, the rate for the first message is $R_1 \leq H[L] + 1 = 4$bits.

Similarly, for $H[\hat{K}]$:

$$H[\hat{K}] \leq \mathbb{E}_X[\mathbb{E}[\log \hat{K}|X = x] \ + \log(\mathbb{E}_X[\mathbb{E}[\log \hat{K}|X = x] + 1) + 1 \tag{87}$$

$$= I(X;Y) + \log(e) + \log(I(X;Y) + \log(e) + 1) + 1 \tag{88}$$

$$\leq I(X;Y) + \log(I(X;Y) + 1) + 2\log(e) + 1 \tag{89}$$

Hence, the rate for the second message is $R_2 \leq H[\hat{K}] + 1 = I(X;Y) + \log(I(X;Y) + 1) + 2\log(e) + 2$bits . Also note that:

$$H[K|W] = H[L, \hat{K}|W] \quad (\text{Given } W, K \text{ and } (L, \hat{K}) \text{ are bijective }) \tag{90}$$

$$\leq H[L|W] + H[\hat{K}|W] \tag{91}$$

$$\leq H[L] + H[\hat{K}] \tag{92}$$

$$\leq I(X;Y) + \log(I(X;Y) + 1) + 7 \text{ (bits)} \tag{93}$$

Since we are compressing two messages separately, we have: $R \leq R_1 + R_2 = I(X;Y) + \log(I(X;Y) + 1) + 9$ (bits)

C Matching Probability of Rejection Sampling

C.1 Distributed Matching Probabilities of RS

Follow the setup in Section 3.2.1, each party independently performs RS using the proposal distribution $Q_Y(\cdot)$ to select indices K_A and K_B and set $(Y_A, Y_B) = (Y_{K_A}, Y_{K_B})$. We assume the bounding condition holds for both parties, i.e. $\max_y \left(P_Y^A(y)Q_Y^{-1}(y), P_Y^B(y)Q_Y^{-1}(y)\right) \leq \omega$, Proposition C.1 shows the probability that they select the same index, given that $Y_{K_A} = y$.

Proposition C.1. *Let $W, Q(.), P_Y^A(.)$ and $P_Y^B(.)$ defined as above. Then we have:*

$$\Pr(Y_A = Y_B | Y_A = y) = \frac{\min(1, P_Y^B(y)/P_Y^A(y))}{1 + \mathrm{TV}(P_Y^A, P_Y^B)} \geq \frac{1}{2\left(1 + P_Y^A(y)/P_Y^B(y)\right)} \tag{94}$$

Furthermore, we have:

$$\Pr(Y_{K_A} = Y_{K_B}) = \frac{1 - \mathrm{TV}(P_Y^A, P_Y^B)}{1 + \mathrm{TV}(P_Y^A, P_Y^B)}. \tag{95}$$

where $\mathrm{TV}(P_Y^A, P_Y^B)$ is the total variation distance between two distribution P_Y^A and P_Y^B.

This matching probability is not as strong, compared to PML as well as IML, details in Appendix C.2[1]. In the case of GRS, we provide an analysis via a non-trivial example in Appendix D.2, where we demonstrate that it is possible to construct target and proposal distributions such that $\Pr(K_A{=}K_B \mid Y_{K_A}{=}y) \to 0.0$, even when $P_Y^A(y) = P_Y^B(y)$. In contrast, this probability is greater than $1/4$ for standard RS. In summary, while GRS and RS can achieve a coding cost in (1), its matching probability remains lower than that attainable by PML and IML.

C.1.1 Proof.

We denote by K_A, K_B the index selected by parties A and B, respectively. We first note that the event $\{K_A = K_B = i, Y_i = y\}$ is equivalent to the event $\{K_A = K_B = i, Y_{K_A} = y\}$, thus:

$$\Pr(K_A = K_B = i | Y_{K_A} = y) = \frac{\Pr(K_A = K_B = i | Y_i = y) Q_Y(y)}{P_Y^A(y)}, \tag{96}$$

where the denominator is due to $Y_{K_A} \sim P_Y^A(.)$. Since:

$$\begin{aligned} \Pr(K_A = K_B | Y_{K_A} = y) &= \sum_{i=1}^{\infty} \Pr(K_A = K_B = i | Y_{K_A} = y) & (97)\\ &= \frac{Q_Y(y)}{P_Y^A(y)} \sum_{i=1}^{\infty} P(K_A = K_B = i | Y_i = y) & (98) \end{aligned}$$

We will later show that:

$$\Pr(K_A = K_B = i | Y_i = y) = \frac{\min(P_Y^A(y), P_Y^B(y))}{\omega Q_Y(y)} \left[1 - \frac{1}{\omega} \int \max(P_Y^A(y), P_Y^B(y)) dy\right]^{i-1}, \tag{99}$$

which gives us:

$$\begin{aligned} &\Pr(K_A{=}K_B | Y_{K_A}{=}y) & (100)\\ &= \frac{Q_Y(y)}{P_Y^A(y)} \cdot \frac{\min(P_Y^A(y), P_Y^B(y))}{\omega Q_Y(y)} \sum_{i=1}^{\infty} \left[1 - \frac{1}{\omega} \int \max(P_Y^A(y), P_Y^B(y)) dy\right]^{i-1} & (101)\\ &= \frac{\min(P_Y^A(y), P_Y^B(y))}{\omega P_Y^A(y)} \sum_{i=0}^{\infty} \left[1 - \frac{1}{\omega} \int \max(P_Y^A(y), P_Y^B(y)) dy\right]^{i} & (102)\\ &= \frac{\min(P_Y^A(y), P_Y^B(y))}{\omega P_Y^A(y)} \frac{\omega}{\int \max(P_Y^A(y), P_Y^B(y)) dy} & (103)\\ &= \frac{\min(1, P_Y^B(y)/P_Y^A(y))}{\int \max(P_Y^A(y), P_Y^B(y)) dy} & (104)\\ &= \frac{\min(1, P_Y^B(y)/P_Y^A(y))}{1 + \mathrm{TV}(P_Y^A, P_Y^B)}, & (105) \end{aligned}$$

where $TV(P_Y^A, P_Y^B)$ is the total variation distance between $P_Y^A(.)$ and $P_Y^B(.)$. Using the inequality $\min(u, v) \geq \frac{uv}{u+v}$ and the fact that $TV(P_Y^A, P_Y^B) \leq 1$ gives us the latter inequality.

[1] Daliri et al. [7] also arrives to a similar conclusion but for discrete case, targeting a different problem.

To show (99), we first compute the following probabilities where A and B both accept/terminate a given sample $Y = y$:

$$\gamma(y) = \Pr(A \text{ and } B \text{ accepts } Y|Y = y) \tag{106}$$
$$= \Pr(U \leq \min(P_Y^A(y), P_Y^B(y)))|Y = y) \tag{107}$$
$$= \frac{\min(P_Y^A(y), P_Y^B(y))}{\omega Q_Y(y)} \tag{108}$$

and,

$$\hat{\gamma}(y) = \Pr(A \text{ and } B \text{ rejects } Y|Y = y) \tag{109}$$
$$= \Pr(U > \max(P_Y^A(y), P_Y^B(y)))|Y = y) \tag{110}$$
$$= 1 - \frac{\max(P_Y^A(y), P_Y^B(y))}{\omega Q_Y(y)} \tag{111}$$

Then we have:

$$\Pr(K_A = K_B = i|Y_i = y_i) \tag{112}$$
$$= \int \Pr(K_A = K_B = i|Y_{1:i} = y_{1:i}) Q_Y(Y_{1:i-1} = y_{1:i-1}|Y_i = y) dy_{1:i-1} \tag{113}$$
$$= \int \Pr(K_A = K_B = i|Y_{1:i} = y_{1:i}) Q_Y(Y_{1:i-1} = y_{1:i-1}) dy_{1:i-1} \tag{114}$$
$$= \int \Pr(K_A = K_B = i|Y_{1:i} = y_{1:i}) Q_Y(Y_{1:i-1} = y_{1:i-1}) dy_{1:i-1} \tag{115}$$
$$= \gamma(y_i) \int \prod_{j=1}^{i-1} \hat{\gamma}(y_j) Q_Y(y_j) dy_{1:i-1} \tag{116}$$
$$= \gamma(y_i) \prod_{j=1}^{i-1} \int \hat{\gamma}(y) Q_Y(y) dy \tag{117}$$
$$= \frac{\min(P_Y^A(y), P_Y^B(y))}{\omega Q_Y(y)} \left[\int \left(1 - \frac{\max(P_Y^A(y), P_Y^B(y))}{\omega Q_Y(y)} \right) Q_Y(y) dy \right]^{i-1} \tag{118}$$
$$= \frac{\min(P_Y^A(y), P_Y^B(y))}{\omega Q_Y(y)} \left[1 - \frac{1}{\omega} \int \max(P_Y^A(y), P_Y^B(y) dy \right]^{i-1} \tag{119}$$

Finally, we note that:

$$\Pr(B \text{ outputs } y|A \text{ outputs } y) \tag{120}$$
$$= \Pr(K_B = K_A|Y_{K_{P_A}=y}) + \Pr(\text{party } B \text{ outputs } y, K_B \neq K_A|Y_{K_A=y}) \tag{121}$$

Finally, note that in the case where $P_A(.), P_B(.)$ are continuous distribution, we have:

$$\Pr(\text{party } B \text{ outputs } y, K_{P_B} \neq K_{P_A}|Y_{K_{P_A}=y}) = 0.0 \tag{122}$$

This completes the proof.

C.2 Comparision with Poisson Matching Lemma

We will compare the average matching probability $\Pr(K_A = K_B)$ between RS and PML in the continuous case. Starting from equation (30) in [24] and assume $P_Y^A(y) \leq P_Y^B(y)$, we have:

$$
\begin{aligned}
&P(Y_A = Y_B = y) && (123)\\
&= \Pr(K_A = K_B | Y_A = y) P(Y_A = y) && (124)\\
&= \frac{1}{\int_{-\infty}^{\infty} \max\left\{\frac{P_Y^A(v)}{P_Y^A(y)}, \frac{P_Y^B(v)}{P_Y^B(y)}\right\} dv} && (125)\\
&= \frac{P_Y^A(y)}{\int_{-\infty}^{\infty} \max\left\{P_Y^A(v), \frac{P_Y^B(v)}{P_Y^B(y)} P_Y^A(y)\right\} dv} && (126)\\
&\geq \frac{P_Y^A(y)}{\int_{-\infty}^{\infty} \max\left\{P_Y^A(v), P_Y^B(v)\right\} dv} \quad \text{(Since we assume } P_Y^A(y) \leq P_Y^B(y)) && (127)\\
&= \frac{P_Y^A(y)}{1 + \mathrm{TV}(P_Y^A, P_Y^B)} && (128)
\end{aligned}
$$

Repeating the same step for $P_Y^A(y) \geq P_Y^B(y)$, we have:

$$
P(Y_A = Y_B = y) \geq \frac{\min(P_Y^A(y), P_Y^B(y))}{1 + \mathrm{TV}(P_Y^A, P_Y^B)} \tag{129}
$$

Taking the integral with respect to y for both sides gives us the desired inequality where the RHS expression is the average matching probability of RS. Finally, the same conclusion holds for IML since the matching probability of IML converges to that of PML.

D Greedy Rejection Sampling.

D.1 Coding Cost

Compared to the standard RS approach described above, GRS is a more well-known tool for channel simulation [12, 16], as its runtime entropy, i.e., $H[K]$, is significantly lower than that of standard RS. Unlike standard RS, where the acceptance probability remains the same on average at each step, GRS greedily accepts samples from high-density regions as early as possible (see [12] for more details). Using these properties, Flamich and Theis [12] provide the following upper bound on $H[K]$, which generalizes the discrete version established by Harsha et al. [16]:

$$
H[K] \leq I[X;Y] + \log(I[X;Y] + 1) + 4, \tag{130}
$$

which has a smaller constant compared to the bound for standard RS. We conclude with a note on the coding cost of GRS, highlighting that, unlike standard RS, which is relatively easy to implement in practice, GRS can be more challenging to deploy as it requires repeatedly computing a complex and potentially intractable integral.

D.2 Matching Probability in Greedy Rejection Sampling

Setup. Let the proposal distribution Q_Y be a discrete uniform $\text{Unif}[1, n]$, i.e. $Q_Y(y) = q = 1/n$ and $U \sim \mathcal{U}(0, 1)$ as in standard RS. Then, we define W as follow:

$$
W = \{(Y_1, U_1), (Y_2, U_2), ...\} \tag{131}
$$

Our goal is to show that, for this proposal distribution Q_Y, there exists the target distributions $P_Y^A(.)$ and $P_Y^B(.)$ such that the GRS matching probability $\Pr(Y_A = Y_B | Y_A = y) \to 0.0$ even when $P_Y^A(y) = P_Y^B(y)$. Let $n = 2k + 1$, we construct the following P_Y^A and P_Y^B:

$$
P_Y^A(Y = 1) = \frac{k+1}{2k+1}, \quad P_Y^A(Y = i) = \begin{cases} \frac{1}{2k+1}, & \text{for } 1 < i \leq k+1 \\ 0.0, & \text{for } i > k+1 \end{cases}, \tag{132}
$$

$$
P_Y^B(Y = 1) = \frac{k+1}{2k+1}, \quad P_Y^B(Y = i) = \begin{cases} \frac{1}{2k+1}, & \text{for } i > k+1 \\ 0.0, & \text{for } 1 < i \leq k+1 \end{cases}, \tag{133}
$$

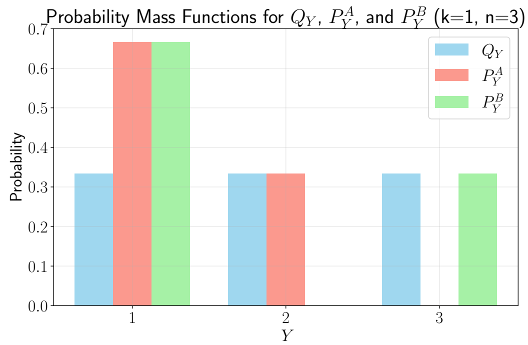

Figure 8: Visualization of example distributions in Section D.2 for $k = 1$.

where we visualize this in Figure 8.

GRS Matching Probability. Given party A has target distribution $P_Y^A(.)$ and party B has target distribution $P_Y^B(.)$, with each running the GRS procedure to obtain their samples Y_A, Y_B respectively. We want to characterize the probability that party A and party B outputs the same value, give party A's output. We denote K_A and K_B as the index within W that party A and party B select respectively, i.e., $Y_{K_A} = Y_A$ and $Y_{K_B} = Y_B$.

Consider the event $Y_A = 1$, with the construction above, we have the following properties:

- If party A and party B both see the first proposal $Y_1 = 1$, they will greedily accept it, since $P_Y^A(Y = 1) = P_Y^B(Y = 1) \geq Q_Y(Y = 1)$. So in this case:
$$\Pr(K_A = K_B = 1, Y_A = 1) = Q_Y(Y = 1) = \frac{1}{2k + 1}$$
- On the other hand, if the first proposal $Y_1 \neq 1$ then either party A or B must accept and output $Y_1 \neq 1$ since for $y \neq 1$, the probability distribution complement each other and equal to $Q_Y(y) = \frac{1}{2k+1}$. For example, for $n = 3$ and $Y_2 = 2$, then party A will accept it while party B must reject it. Therefore, we have:
$$\Pr(K_A = K_B > 1, Y_A = 1) = 0.0.$$
- Finally, from the previous analysis, for any positive integers $i \neq j$, we have
$$\Pr(K_A = i, K_B = j, Y_A = 1, Y_B = 1) = 0.0,$$
Indeed, consider $i = 1$ then $\Pr(K_A = 1, K_B = j, Y_A = 1, Y_B = 1) = 0.0$ since both of them must accept the first proposal $Y_1 = 1$. On the other hand, if $i > 1$ then we must have $j = 1$ since we know that $Y_1 \neq 1$ in this case and thus one of the party must stop. Since $i > 1$, it has to be party B and in this case, $Y_B \neq 1$.

For this reason, we have:

$$\Pr(Y_A = Y_B = 1) \tag{134}$$
$$= \Pr(K_A = K_B, Y_A = 1, Y_B = 1) + \Pr(K_A \neq K_B, Y_A = 1, Y_B = 1) \tag{135}$$
$$= \Pr(K_A = K_B, Y_A = 1) + \sum_{i \neq j} \Pr(K_A = i, K_B = j, Y_A = 1, Y_B = 1) \tag{136}$$
$$= \Pr(K_A = K_B, Y_A = 1) \tag{137}$$
$$= \Pr(K_A = K_B = 1, Y_A = 1) + \Pr(K_A = K_B > 1, Y_A = 1) \tag{138}$$
$$= Q_Y(Y = 1) \tag{139}$$
$$= \frac{1}{2k + 1} \tag{140}$$

and hence:

$$\Pr(Y_A = Y_B | Y_A = 1) = \frac{1}{k + 1} \tag{141}$$

which approaches 0.0 as $n \to \infty$. Overall, due to its greedy selection approach, GRS may yield lower matching probabilities compared to other methods such as PML which we provide the analysis below.

Matching Probability of PML. In PML, the matching probability is $\Pr(Y_A = Y_B \mid Y_A = 1) = 1$. This results from PML's more global selection process compared to GRS, as it evaluates all candidates comprehensively. In particular, let $W = (S_1, Y_1), ..., (S_n, Y_n)$ where $S_i \sim \mathrm{Exp}(1)$ and let K_A, K_B be the value within W that each party respectively select in this case. Note that the construction of W in the discrete case for PML does not require Q_Y. The selection process according to PML is as follows:

$$K_A = \arg\min_{1 \leq i \leq n} \frac{S_i}{P_Y^A(Y_i)} \quad K_B = \arg\min_{1 \leq i \leq n} \frac{S_i}{P_Y^B(Y_i)}, \tag{142}$$

and each party outputs $Y_A = Y_{K_A}, Y_B = Y_{K_B}$. We see that if $K_A = 1$, then we must have $K_B = 1$. This is because for any $i > 1$, we have $P_Y^A(Y = 1) = P_Y^B(Y = 1) > P_Y^B(Y = i)$ and $P_Y^A(Y = i) = P_Y^B(Y = i + 1 + k)$. Thus, this gives $\Pr(Y_A = Y_B \mid Y_A = 1) = 1$.

Algorithm 1: Ensemble Rejection Sampling - $\text{ERS}(W; P_Y, Q_Y, \omega = \max_y \frac{P_Y(y)}{Q_Y(y)}, \text{scale} = 1)$

Input: Target distribution P_Y, Proposal distribution Q_Y, and the source of randomness W (see Section 5.1). Default value $\omega = \max_y \frac{P_Y(y)}{Q_Y(y)}$ unless override by some value $> \omega$.
Default scaling factor $\text{scale} = 1$ unless override by some value within $(0, 1]$.

Output: Selected Index K and sample $Y_K \sim P_Y$

1. Observe batch $\{B_i, U_i\}$
2. Select candidate index K_i^{cand}:

$$K_i^{\text{cand}} = \underset{1 \le k \le N}{\text{argmin}} \frac{S_{ik}}{\lambda_{ik}}, \quad \text{where:} \quad \lambda_{ik} = \frac{P_Y(Y_{ik})}{Q_Y(Y_{ik})}$$

3. Compute:

$$\hat{Z}(Y_{i,1:N}) = \sum_{k=1}^{N} \lambda_{ik}\,, \quad \bar{Z}(Y_{i,1:N}, K_i^{\text{cand}}) = \hat{Z}(Y_{i,1:N}) + \omega - \lambda_{i,K_i^{\text{cand}}}$$

4. Set $K_1 = i$, $K_2 = K_i^{\text{cand}}$, $K = (N-1)1i + K_i^{\text{cand}}$ and return Y_K if:

$$U_i \le \frac{\hat{Z}(Y_{i,1:N})}{\bar{Z}(Y_{i,1:N}, K_i^{\text{cand}})} \cdot \text{scale},$$

else repeat Step 1 with B_{i+1}.

E ERS Coding Scheme

E.1 Prelimaries

We show the standard ERS algorithm in Algorithm 1, following the original version introduced by Deligiannidis et al. [8] with a slight generalization in terms of the scaling factor ($0 < \text{scale} \le 1$) that we will use for channel simulation purpose. This section begins by establishing some detailed quantities that will be used repeatedly. For simplicity, we use $P_x(.)$ for the target distribution $P_{Y|X=x}$ and $Q(.)$ for the proposal distribution. Let $\omega_x = \max_y P_x(y)/Q(y)$, we define the quantities:

$$\hat{Z}_x(y_{1:N}) = \sum_{j=1}^{N} \frac{P_x(y_j)}{Q(y_j)}, \quad \bar{Z}_x(y_{1:N}, k) = \hat{Z}_x(y_{1:N}) - \frac{P_x(y_k)}{Q(y_k)} + \omega_x \tag{143}$$

and denote the following constants:

$$\Delta_x = \mathbb{E}_{Y_{1:N} \sim Q}\left[\frac{N}{\bar{Z}_x(Y_{1:N}, 1)}\right], \quad \Delta = \frac{N}{N - 1 + \omega}, \tag{144}$$

where we recall that $\omega = \max_{x,y} P_x(y)/Q(y)$, by Jensen's inequality we have the following:

$$\Delta_x \ge \frac{N}{N - 1 + \omega_x} \ge \frac{N}{N - 1 + \omega} = \Delta \text{ for every } x. \tag{145}$$

From this, we can see that as $N \to \infty$, we achieve $\Delta \to 1.0$. This value Δ turns out to be the average batch acceptance probability when we set $\text{scale} = \frac{\Delta}{\Delta_x}$, which we elaborate on below.

Scaled Acceptance Probability. For the channel simulation setting in this section, we slightly modify the acceptance probability in Algorithm 1 (step 4) with a scaling factor $\text{scale} = \frac{\Delta}{\Delta_x} \le 1$ such that the average batch acceptance probability is the same, regardless of the target distribution P_x [2]. In particular, the encoder selects the index according to:

$$K = \text{ERS}(W; P_x, Q, \text{scale} = \frac{\Delta}{\Delta_x}), \tag{146}$$

[2]This is similar to the case of standard RS where we accept/reject based on the global ratio bound ω instead of ω_x.

which means for a batch i containing samples $Y_{i,1:N} = y_{1:N}$, we accept it within step 4 if:

$$\text{Accept if } U_i \leq \frac{\hat{Z}_x(y_{1:N})}{\bar{Z}_x(y_{1:N}, k)} \frac{\Delta}{\Delta_x}, \tag{147}$$

where we modify the scaling $\text{scale} = \frac{\Delta}{\Delta_x} \leq 1$ in Algorithm 1, which is a constant and does not affect the resulting output distribution. The value of k is determined via the Gumbel-Max selection procedure in Step 2. The intuition is, within every accepted batch without scaling, we randomly reject $(1 - \text{scale})$ of them. Formally, first consider the following **ERS proposal distribution**:

$$\bar{Q}_{Y_{1:N},K}(y_{1:N}, k; x) = \left(\frac{P_x(y_k)/Q(y_k)}{\sum_{j=1}^N P_x(y_j)/Q(y_j)} \right) \prod_{j=1}^N Q(y_j) \tag{148}$$

$$= \left(\frac{P_x(y_k)/Q(y_k)}{\hat{Z}_x(y_{1:N})} \right) \prod_{j=1}^N Q(y_j), \tag{149}$$

where the first product in the RHS is the likelihood we obtain the samples $y_{1:N}$ from the original proposal distribution $Q_Y(.)$ and the ratio is due to the IS process. Now, the **ERS target distribution** is $\bar{P}_{Y_{1:N},K}(y_{1:N}, k; x)$ where

$$\bar{P}_{Y_{1:N},K}(y_{1:N}, k; x) = \frac{1}{\alpha} \left(\frac{P_x(y_k)/Q(y_k)}{\hat{Z}_x(y_{1:N})} \frac{\hat{Z}_x(y_{1:N})}{\bar{Z}_x(y_{1:N}, k)} \frac{\Delta}{\Delta_x} \right) \prod_{j=1}^N Q(y_j) \tag{150}$$

$$= \left(\frac{P_x(y_k)/Q(y_k)}{\Delta_x \bar{Z}_x(y_{1:N}, k)} \right) \prod_{j=1}^N Q(y_j), \tag{151}$$

which is the batch target distribution that yields $Y \sim P_x$ when no scaling occur (see [8], Section 2.2), since the normalization factor α is:

$$\alpha = \sum_{k=1}^N \int_{-\infty}^{\infty} \left(\frac{P_x(y_k)/Q(y_k)}{\bar{Z}_x(y_{1:N}, k)} \right) \frac{\Delta}{\Delta_x} \left(\prod_{j=1}^N Q(y_j) \right) dy_{1:N} \tag{152}$$

$$= N \frac{\Delta}{\Delta_x} \int_{-\infty}^{\infty} \left(\frac{P_x(y_k)/Q(y_k)}{\bar{Z}_x(y_{1:N}, 1)} \right) \left(\prod_{j=1}^N Q(y_j) \right) dy_{1:N} \quad \text{(Due to symmetry)} \tag{153}$$

$$= N \frac{\Delta}{\Delta_x} \int_{-\infty}^{\infty} \frac{1}{\bar{Z}_x(y_{1:N}, 1)} \left(\prod_{j=2}^N Q(y_j) \right) dy_2^N \tag{154}$$

$$= \Delta \tag{155}$$

It turns out that Δ is also the **batch acceptance probability** since:

$$\Pr(\text{Accept batch } B) = \mathbb{E}_{(Y_{1:N},K) \sim \bar{Q}} \left[\frac{\Delta}{\Delta_x} \frac{\hat{Z}_x(y_{1:N})}{\bar{Z}_x(y_{1:N}, k)} \right] \tag{156}$$

$$= \frac{\Delta}{\Delta_x} \sum_{k=1}^N \int_{-\infty}^{\infty} \left(\frac{P_x(y_k)/Q(y_k)}{\bar{Z}_x(y_{1:N}, k)} \right) \tag{157}$$

$$= \frac{\Delta}{\Delta_x} N \int_{-\infty}^{\infty} \left(\frac{P_x(y_1)/Q(y_1)}{\bar{Z}_x(y_{1:N}, 1)} \right) \left(\prod_{j=1}^N Q(y_j) \right) dy_{1:N} \tag{158}$$

$$= \Delta, \tag{159}$$

and it can be observed that, without the scaling factor $\frac{\Delta}{\Delta_x}$, the batch acceptance probability is Δ_x. Finally, we can view the ERS as a standard RS procedure with proposal distribution $\bar{Q}_{Y_1^N,K}$ and target distribution $\bar{P}_{Y_1^N,K}$.

Harris-FKG/Chebyshev Inequality. We introduce the following inequality (Harris-FKG/Chebyshev), which will be used in the proof:

Proposition E.1. *For function f, g on $Y \sim P(.)$ where f is non-increasing and g is non-decreasing, we have:*

$$\mathbb{E}[f(Y)g(Y)] \leq \mathbb{E}[f(Y)]\mathbb{E}[g(Y)]$$

Proof. Let $Y_1, Y_2 \sim P(.)$ and they are independent. Then we have:

$$[f(Y_1) - f(Y_2)][g(Y_1) - g(Y_2)] \leq 0 \tag{160}$$

Hence:

$$\mathbb{E}\{[f(Y_1) - f(Y_2)][g(Y_1) - g(Y_2)]\} \leq 0 \tag{161}$$

This gives us:

$$\mathbb{E}[f(Y_1)g(Y_1)] + \mathbb{E}[f(Y_2)g(Y_2)] \leq \mathbb{E}[f(Y_1)]\mathbb{E}[g(Y_2)] + \mathbb{E}[f(Y_2)]\mathbb{E}[g(Y_1)], \tag{162}$$

which completes the proof. □

E.2 Encoding K_1.

We encode K_1 the same way as the scheme for standard RS. Similar to standard RS, we encode K_1 into two messages. Specifically:

- Step 1: the encoder sends the ceiling $L = \lceil \frac{K_1}{\lfloor \Delta^{-1} \rfloor} \rceil$ to the decoder. The decoder then knows $(L-1)\lfloor \Delta^{-1} \rfloor^{-1} + 1 \leq L \leq L\lfloor \Delta^{-1} \rfloor^{-1}$, i.e. K_1 is in chunk L that consists of $\lfloor \Delta^{-1} \rfloor^{-1}$ batches. We have $\mathbb{E}[\log(L)] \leq 1$ bit.
- Step 2: The encoder and decoder both sort the uniform random variables U_i within the selected chunk $(L-1)\lfloor \Delta^{-1} \rfloor^{-1} + 1 \leq i \leq L\lfloor \Delta^{-1} \rfloor^{-1}$. Let the sorted list be $U_{\pi(1)} \leq U_{\pi(2)} \leq ... \leq U_{\pi(\lfloor \Delta^{-1} \rfloor)}$ where $\pi(.)$ is the mapping between the sorted index and the original unsorted one. The encoder sends the rank of U_{K_1} within this list, i.e. sends the value T such that $K_1 = \pi(\hat{K}_1)$. The decoder receive $\hat{K}_1$ and retrieve B_{K_1} accordingly. Section E.2.2 shows the coding cost for this step.

We provide the detail analysis in Section E.2.1 and E.2.2. Notice that the role Δ plays here is similar to that of ω in standard RS.

E.2.1 Coding Cost of L

Similar to RS, since each batch is accepted with probability Δ (see (159)), this means:

$$\Pr(K_1 > \ell\Delta^{-1}) = (1-\Delta)^{\ell\lfloor \Delta^{-1} \rfloor} < 0.5^{-\ell},$$

which is equivalent to $\Pr(L > \ell) < 0.5^{-\ell}$. Note that we reuse the inequality in Appendix B.3. We have:

$$\mathbb{E}[L] = \sum_{\ell=0}^{\infty} \Pr(L > \ell) < 1 + 0.5^{-1} + 0.5^{-2} + ... = 2, \tag{163}$$

implying $\mathbb{E}[\log L] \leq 1$.

E.2.2 Coding Cost of $\hat{K}_1$

We will show that:

$$\mathbb{E}[\log \hat{K}_1] \leq \frac{N}{\Delta_x} \mathbb{E}_{Y_{1:N} \sim Q} \left[\frac{P_x(Y_1)/Q(Y_1)}{\bar{Z}_x(Y_{1:N}, 1)} \log \left(\frac{\hat{Z}_x(Y_{1:N})}{\Delta_x \bar{Z}_x(Y_{1:N}, 1)} \right) \right], \tag{164}$$

where we provide the result of (164) in Section E.4.2.

E.3 Encoding K_2.

Given an accepted batch $\{(Y_i, S_i)\}_{i=1}^N$, we have:

$$K_2 = \arg\min_{1\le i\le N} \frac{S_i}{\lambda_i}; \qquad \Theta_P = \min_{1\le i\le N} \frac{S_i}{\lambda_i}, \tag{165}$$

where we have the weights λ_i defined as:

$$\lambda_i = \frac{P(Y_i)}{Q(Y_i)} \tag{166}$$

After communicating the selected batch index K_1, the encoder and decoder sort the exponential random variables $\{S_{K_1,i}\}_{i=1}^N$, i.e.

$$S_{K_2,\pi(1)} \le S_{K_2,\pi(2)} \le \ldots \le S_{K_2,\pi(N)}, \tag{167}$$

and send the sorted index $\hat{K}_2$ of K_2, i.e. $\pi(\hat{K}_2) = K_2$. The decoder also performs the sorting operation and retrieve K_2 accordingly. Since K_2 are obtained from the batch selected by ERS, we analyze $\mathbb{E}[\log \hat{K}_2' | Y_{1:N} \text{ are selected}]$, where $\hat{K}_2'$ and K_2' are defined the same as $\hat{K}_2$ and K_2 (follows the same Gumbel-Max procedure) but for arbitrary N i.i.d. proposals $Y_{1:N} \sim Q(.)$. In this case:

$$\mathbb{E}[\log \hat{K}_2] = \mathbb{E}[\log \hat{K}_2' | Y_{1:N} \text{ are selected}]$$

Notice the following identity:

$$\bar{P}(y_{1:N}, k_2; x) = P(Y_{1:N} = y_{1:N} | Y_{1:N} \text{ are selected}, K_2' = k_2) \Pr(K_2' = k_2 | Y_{1:N} \text{ are selected})$$

where $\bar{P}(y_{1:N}, k_2; x)$ is the ERS target distribution described previously in Appendix E.1. Then, we obtain the following likelihood:

$$P(Y_{1:N} = y_{1:N} | Y_{1:N} \text{ are selected}, K_2' = 1) \tag{168}$$

$$= \frac{\bar{P}(y_{1:N}, j; x)}{\Pr(K_2' = 1 | Y_{1:N} \text{ are selected})} \tag{169}$$

$$= N \frac{P_x(y_1)/Q(y_1)}{\Delta_x \bar{Z}_x(y_{1:N}, 1)} \prod_{i=1}^N Q(y_i) \tag{170}$$

With this, we now bound the expectation term of interest $\mathbb{E}[\log \hat{K}_2]$ as follows:

$$\mathbb{E}[\log \hat{K}_2] \tag{171}$$

$$= \mathbb{E}[\log \hat{K}_2' | Y_{1:N} \text{ are selected}] \tag{172}$$

$$= \mathbb{E}[\log \hat{K}_2' | Y_{1:N} \text{ are selected}, K_2' = 1] \quad \text{(Due to Symmetry)} \tag{173}$$

$$= \mathbb{E}_{Y_{1:N}}[\mathbb{E}[\log \hat{K}_2' | Y_{1:N} \text{ are selected}, K_2' = 1, Y_{1:N} = y_{1:N}]] \tag{174}$$

$$= N \int_{-\infty}^{\infty} \frac{P_x(y_1)/Q(y_1)}{\Delta_x \bar{Z}_x(y_{1:N}, 1)} \mathbb{E}[\log \hat{K}_2' | Y_{1:N} \text{ are selected}, Y_{1:N} = y_{1:N}, K_2' = 1] \left(\prod_{i=1}^N Q(y_i) \right) dy_{1:N} \tag{175}$$

$$= N \int_{-\infty}^{\infty} \frac{P_x(y_1)/Q(y_1)}{\Delta_x \bar{Z}_x(y_{1:N}, 1)} \mathbb{E}[\log \hat{K}_2' | Y_{1:N} = y_{1:N}, K_2' = 1] \left(\prod_{i=1}^N Q(y_i) \right) dy_{1:N}, \tag{176}$$

where the last equality is because, given $\{Y_{1:N}{=}y_{1:N}, K_2'{=}1\}$, the event $\{Y_{1:N} \text{ are selected}\}$ and the random variable $\hat{K}_2'$ are independent. In particular, the decision whether to accept a batch or not does not depends on the rank of $S_{K_2'}$, that is:

$$\Pr(Y_{1:N} \text{ are selected} | Y_{1:N} = y_{1:N}, K_2' = 1, \hat{K}_2' = k_2) \tag{177}$$

$$= \Pr(Y_{1:N} \text{ are selected} | Y_{1:N} = y_{1:N}, K_2' = 1) \tag{178}$$

$$= \frac{\hat{Z}_x(y_{1:N})}{\bar{Z}_x(y_{1:N}, 1)} \frac{\Delta}{\Delta_x} \tag{179}$$

We then have:

$$\mathbb{E}[\log \hat{K}_2'|Y_{1:N} \text{ are selected}] \tag{180}$$

$$=N\int_{-\infty}^{\infty}\prod_{i=1}^{N}Q(y_i)\frac{P_x(y_1)/Q(y_1)}{\Delta_x\bar{Z}_x(y_{1:N},1)}\left(\int_0^{\infty}e^{-\theta}\mathbb{E}[\log \hat{K}_2'|Y_{1:N}{=}y_{1:N}, K_2'{=}1, \Theta_P{=}\theta]d\theta\right)dy_{1:N}, \tag{181}$$

since, given $Y_{1:N}$, Θ_P is independent of K_2' and $\Theta_P{\sim}\text{Exp}(1)$ (see [32], Appendix 18). We now provide an upperbound of $\mathbb{E}[\log \hat{K}_2|Y_{1:N}{=}y_{1:N}, K_2{=}1, \Theta_P{=}\theta]$, which follows the argument presented in [32], and is repeated here. Applying Jensen's inequality, we have:

$$\mathbb{E}[\log \hat{K}_2'|Y_{1:N}=y_{1:N}, K_2'=1, \Theta_P=\theta] \leq \log \mathbb{E}[\hat{K}_2'|Y_{1:N}=y_{1:N}, K_2'=1, \Theta_P=\theta], \tag{182}$$

We then rewrite $\hat{K}_2'$ as the following:

$$\hat{K}_2' = |\{S_i < S_{K_2'}\}| + 1, \tag{183}$$

which gives us:

$$\mathbb{E}[\hat{K}_2'|Y_{1:N}=y_{1:N}, K_2'=1, \Theta_P=\theta] \tag{184}$$

$$= 1 + \mathbb{E}[|\{S_i < S_{K_2'}\}||Y_{1:N}=y_{1:N}, K_2'=1, \Theta_P=\theta] \tag{185}$$

$$= 1 + \mathbb{E}\left[\left|\left\{S_i < \theta\frac{P_x(Y_{K_2'})/Q(Y_{K_2'})}{\hat{Z}_x(Y_{1:N})}\right\}\right|\middle| Y_{1:N}=y_{1:N}, K_2'=1, \Theta_P=\theta\right] \tag{186}$$

$$= 1 + \sum_{i=2}^{N}\Pr\left(S_i < \theta\frac{P_x(Y_{K_2'})/Q(Y_{K_2'})}{\hat{Z}_x(Y_{1:N})}\middle| Y_{1:N}=y_{1:N}, K_2'=1, \Theta_P=\theta\right) \tag{187}$$

$$= 1 + \sum_{i=2}^{N}\Pr\left(S_i < \theta\frac{P_x(Y_1)/Q(Y_1)}{\hat{Z}_x(Y_{1:N})}\middle| Y_{1:N}=y_{1:N}, \frac{S_j}{\frac{P_x(y_j)/Q(y_j)}{\hat{Z}_x(y_{1:N})}}{\geq}\theta \text{ for } j{\neq}1, \frac{S_1}{\frac{P_x(y_1)/Q(y_1)}{\hat{Z}_x(y_{1:N})}}{=}\theta\right) \tag{188}$$

$$= 1 + \sum_{i=2}^{N}\Pr\left(S_i < \theta\frac{P_x(Y_1)/Q(Y_1)}{\hat{Z}_x(Y_{1:N})}\middle| Y_{1:N}{=}y_{1:N}, \frac{S_j}{\frac{P_x(y_j)/Q(y_j)}{\hat{Z}_x(y_{1:N})}}{\geq}\theta \text{ for } j{\neq}1, \frac{S_1}{\frac{P_x(y_1)/Q(y_1)}{\hat{Z}_x(y_{1:N})}}{=}\theta\right) \tag{189}$$

$$= 1 + \sum_{i=2}^{N}\Pr\left(S_i < \theta\frac{P_x(Y_1)/Q(Y_1)}{\hat{Z}_x(Y_{1:N})}\middle| Y_{1:N}=y_{1:N}, \frac{S_i}{\frac{P_x(y_i)/Q(y_i)}{\hat{Z}_x(y_{1:N})}} \geq \theta\right) \tag{190}$$

Note that:

$$\Pr\left(S_i < \theta\frac{P_x(Y_1)/Q(Y_1)}{\hat{Z}_x(y_{1:N})}\middle| Y_{1:N}=y_{1:N}, \frac{S_i}{\frac{P_x(y_i)/Q(y_i)}{\hat{Z}_x(y_{1:N})}} \geq \theta\right) \tag{191}$$

$$= \mathbf{1}\left\{\theta\frac{P_x(Y_1)/Q(Y_1)}{\hat{Z}_x(y_{1:N})} \geq \theta\frac{P_x(Y_i)/Q(Y_i)}{\hat{Z}_x(y_{1:N})}\right\}\left[1 - \exp\left(-\theta\frac{P_x(y_1)/Q(y_1) - P_x(y_i)/Q(y_i)}{\hat{Z}_x(y_{1:N})}\right)\right] \tag{192}$$

$$\leq 1 - \exp\left(-\theta\frac{P_x(y_1)/Q(y_1) - P_x(y_i)/Q(y_i)}{\hat{Z}_x(y_{1:N})}\right) \tag{193}$$

$$\leq \frac{\theta[P_x(y_1)/Q(y_1) - P_x(y_i)/Q(y_i)]}{\hat{Z}_x(y_{1:N})} \tag{194}$$

$$\leq \frac{\theta P_x(y_1)/Q(y_1)}{\hat{Z}_x(y_{1:N})} \tag{195}$$

As such:

$$\mathbb{E}[\hat{K}_2'|Y_{1:N}=y_{1:N}, K_2'=1, \Theta_P=\theta] \leq 1+\sum_{i=2}^{N}\frac{\theta P_x(y_1)/Q(y_1)}{\hat{Z}_x(y_{1:N})} \tag{196}$$

$$\leq 1+\frac{N\theta P_x(y_1)/Q(y_1)}{\hat{Z}_x(y_{1:N})} \tag{197}$$

and thus:

$$\int_0^\infty e^{-\theta}\mathbb{E}[\log K|Y_{1:N}=y_{1:N}, K_2=1, \Theta_P=\theta]d\theta \tag{198}$$

$$\leq \int_0^\infty e^{-\theta}\log\left(1+\frac{N\theta P_x(y_1)/Q(y_1)}{\hat{Z}_x(y_{1:N})}\right)d\theta \tag{199}$$

$$\leq \log\left(\frac{NP_x(y_1)/Q(y_1)}{\hat{Z}_x(y_{1:N})}+1\right), \tag{200}$$

which is due to Jensen's inequality for concave function $\log(.)$. Finally, we have:

$$\mathbb{E}[\log\hat{K}_2] \tag{201}$$

$$\leq \mathbb{E}_{Y_{1:N}\sim Q(.)}\left[\frac{NP_x(Y_1)/Q(Y_1)}{\Delta_x\bar{Z}_x(Y_{1:N},1)}\log\left(\frac{NP_x(Y_1)/Q(Y_1)}{\hat{Z}_x(Y_{1:N})}+1\right)\right] \tag{202}$$

$$\leq \mathbb{E}_{Y_{1:N}\sim Q(.)}\left[\frac{NP_x(Y_1)/Q(Y_1)}{\Delta_x\bar{Z}_x(Y_{1:N},1)}\log\left(\frac{NP_x(Y_1)/Q(Y_1)}{\hat{Z}_x(Y_{1:N})}\right)+\frac{\log(e)\hat{Z}_x(Y_{1:N})}{NP_x(Y_1)/Q(Y_1)}\frac{NP_x(Y_1)/Q(Y_1)}{\Delta_x\bar{Z}_x(Y_{1:N},1)}\right] \tag{203}$$

$$= \mathbb{E}_{Y_{1:N}\sim Q(.)}\left[\frac{NP_x(Y_1)/Q(Y_1)}{\Delta_x\bar{Z}_x(Y_{1:N},1)}\log\left(\frac{NP_x(Y_1)/Q(Y_1)}{\hat{Z}_x(Y_{1:N})}\right)\right]+\log(e) \tag{204}$$

The last inequality is due to the FKG inequality:

$$\mathbb{E}_{Y_{1:N}\sim Q(.)}\left[\frac{\log(e)\hat{Z}_x(Y_{1:N})}{\Delta_x\bar{Z}_x(Y_{1:N},1)}\right] \tag{205}$$

$$= \log(e)\mathbb{E}_{Y_2^N\sim Q(.)}\left[\left(1+\sum_{i=2}^{N}\frac{P_x(Y_i)}{Q(Y_i)}\right)\left(\frac{1}{\Delta_x\bar{Z}_x(Y_{1:N},1)}\right)\right] \tag{206}$$

$$\leq \log(e)\mathbb{E}_{Y_2^N\sim Q(.)}\left[1+\sum_{i=2}^{N}\frac{P_x(Y_i)}{Q(Y_i)}\right]\mathbb{E}_{Y_2^N\sim Q(.)}\left[\frac{1}{\Delta_x\bar{Z}_x(Y_{1:N},1)}\right] \tag{207}$$

$$= \log(e) \tag{208}$$

So we have the bound on $\mathbb{E}[\log(\hat{K}_2)]$ as:

$$\mathbb{E}[\log(\hat{K}_2)] \leq \mathbb{E}_{Y_{1:N}\sim Q(.)}\left[\frac{NP_x(Y_1)/Q(Y_1)}{\Delta_x\bar{Z}_x(Y_{1:N},1)}\log\left(\frac{NP_x(Y_1)/Q(Y_1)}{\hat{Z}_x(Y_{1:N})}\right)\right]+\log(e) \tag{209}$$

E.4 Total Coding Cost of K

We now provide an upperbound on the total coding cost of K. We have:

$$H(K|W) = H(L,\hat{K}_1,\hat{K}_2|W) \tag{210}$$

$$\leq H(L|W)+H(\hat{K}_1|W)+H(\hat{K}_2|W) \tag{211}$$

$$\leq H(L)+H(\hat{K}_1)+H(\hat{K}_2) \tag{212}$$

For each of the message, we encode using Zipf distribution. Since $\mathbb{E}[\log(L)] \leq 1$, then:

$$H(L) \leq 3$$

For $H(\hat{K}_1)$, we have:

$$H(\hat{K}_1) \leq \mathbb{E}_X[\mathbb{E}[\log(\hat{K}_1)]] + \log(\mathbb{E}_X[\mathbb{E}[\log(\hat{K}_1)]] + 1) + 1 \tag{213}$$

and $H(\hat{K}_2)$, we have:

$$H(\hat{K}_2) \leq \mathbb{E}_X[\mathbb{E}[\log(\hat{K}_2)]] + \log(\mathbb{E}_X[\mathbb{E}[\log(\hat{K}_2)]] + 1) + 1 \tag{214}$$

and thus we have:

$$H(K|W) \tag{215}$$

$$\leq (\mathbb{E}_X[\mathbb{E}[\log(\hat{K}_1)] + \mathbb{E}[\log(\hat{K}_2)]]) + \log((\mathbb{E}_X[\mathbb{E}[\log(\hat{K}_1)]] + 1)(\mathbb{E}_X[\mathbb{E}[\log(\hat{K}_2)]] + 1)) + 5 \tag{216}$$

By AM-GM inequality, we have:

$$\log((\mathbb{E}_X[\mathbb{E}[\log(\hat{K}_1)]] + 1)(\mathbb{E}_X[\mathbb{E}[\log(\hat{K}_2)]] + 1)) \tag{217}$$

$$\leq \log(\frac{1}{4}(\mathbb{E}_X[\mathbb{E}[\log(\hat{K}_1)]] + 1 + \mathbb{E}_X[\mathbb{E}[\log(\hat{K}_2)]] + 1)^2) \tag{218}$$

$$= 2\log(\mathbb{E}_X[\mathbb{E}[\log(\hat{K}_1)]] + \mathbb{E}_X[\mathbb{E}[\log(\hat{K}_2)]] + 2) - 2 \tag{219}$$

We will show $\mathbb{E}[\log(\hat{K}_1)] + \mathbb{E}[\log(\hat{K}_2)] \leq D_{KL}(P_x||Q) + 3 + 2\log(e)$ at the end of this section. Given this, we have:

$$H(K|W) \leq I(X;Y) + 3 + 2\log(e) + 2\log(I(X;Y) + 5 + 2\log(e)) - 2 + 5 \tag{220}$$

$$\leq I(X;Y) + 2\log(I(X;Y) + 8) + 9. \tag{221}$$

Since we are encoding 3 messages separately, we add 1 bit overhead for each message and thus arrive to the constant 12 as in the original result.

The rest is to bound $\mathbb{E}[\log(\hat{K}_1)] + \mathbb{E}[\log(\hat{K}_2)]$, note that:

$$\mathbb{E}[\log(\hat{K}_1)] + \mathbb{E}[\log(\hat{K}_2)] \tag{222}$$

$$\leq 2\log(e) + \mathbb{E}_{Y_{1:N}\sim Q(.)}\left[\frac{NP_x(Y_1)/Q(Y_1)}{\Delta_x\bar{Z}_x(Y_{1:N},1)}\left(\log\left(\frac{\hat{Z}_x(Y_{1:N})}{\Delta_x\bar{Z}_x(Y_{1:N},1)}\right) + \log\left(\frac{NP_x(Y_1)/Q(Y_1)}{\hat{Z}_x(Y_{1:N})}\right)\right)\right] \tag{223}$$

$$= 2\log(e) + \mathbb{E}_{Y_{1:N}\sim Q(.)}\left[\frac{NP_x(Y_1)/Q(Y_1)}{\Delta_x\bar{Z}_x(Y_{1:N},1)}\log\left(\frac{NP_x(Y_1)/Q(Y_1)}{\Delta_x\bar{Z}_x(Y_{1:N},1)}\right)\right] \tag{224}$$

$$= 2\log(e) + \mathbb{E}_{Y_{1:N}\sim Q(.)}\left[\frac{NP_x(Y_1)/Q(Y_1)}{\Delta_x\bar{Z}_x(Y_{1:N},1)}\left(\log\frac{P_x(Y_1)}{Q(Y_1)} + \log\left(\frac{N}{\Delta_x\bar{Z}_x(Y_{1:N},1)}\right)\right)\right] \tag{225}$$

$$= 2\log(e) + D_{KL}(P_x||Q) + \mathbb{E}_{Y_{1:N}\sim Q(.)}\left[\frac{N}{\Delta_x\bar{Z}_x(Y_{1:N},1)}\log\left(\frac{N}{\Delta_x\bar{Z}_x(Y_{1:N},1)}\right)\right] \tag{226}$$

$$= 2\log(e) + D_{KL}(P_x||Q) + E_1 \tag{227}$$

where:

$$E_1 = \mathbb{E}_{Y_{1:N}\sim Q(.)}\left[\frac{N}{\Delta_x\bar{Z}_x(Y_{1:N},1)}\log\left(\frac{N}{\Delta_x\bar{Z}_x(Y_{1:N},1)}\right)\right] \tag{228}$$

We will show in Appendix E.4.1 that:

$$E_1 \leq 3 \tag{229}$$

and thus:

$$\mathbb{E}[\log(\hat{K}_1)] + \mathbb{E}[\log(\hat{K}_2)] \leq 2\log(e) + 3 + D_{KL}(P_x||Q) \tag{230}$$

E.4.1 Bound on E_1

We consider two cases, when the batch size $N \le 7\omega_x$ and when $N > 7\omega_x$.

<u>*Case 1*</u>: $N \le 7\omega_x$

Recall that $\bar{Z}_x(Y_{1:N}, 1) > \omega_x$ and $\Delta_x \ge \frac{N}{N-1+\omega_x}$, we have:

$$\begin{aligned}
\frac{N}{\Delta_x \bar{Z}_x(Y_{1:N},1)} &\le \frac{N-1+\omega_x}{\omega_x} && (231)\\
&< \frac{8\omega_x - 1}{\omega_x} \quad (\text{ Since } N \le 7\omega) && (232)\\
&< 8 && (233)
\end{aligned}$$

Thus, we have:

$$\begin{aligned}
E_1 &= \mathbb{E}_{Y_{1:N}\sim Q(.)}\left[\frac{N}{\Delta_x \bar{Z}_x(Y_{1:N},1)} \log\left(\frac{N}{\Delta_x \bar{Z}_x(Y_{1:N},1)}\right)\right] && (234)\\
&\le \mathbb{E}_{Y_{1:N}\sim Q(.)}\left[\frac{N}{\Delta_x \bar{Z}_x(Y_{1:N},1)} \log(8)\right] && (235)\\
&= 3 \quad (\text{Since } \Delta_x = \mathbb{E}_{Y_{1:N}\sim Q(.)}\left[\frac{N}{\bar{Z}_x(Y_{1:N},1)}\right]), && (236)
\end{aligned}$$

and hence $E_1 \le 3$ bit.

<u>*Case 2*</u>: $N > 7\omega$

To upper-bound E_2 in this regime, we first note that:

$$\Delta_x = \mathbb{E}_{Y_{1:N}\sim Q(.)}\left[\frac{N}{\bar{Z}_x(Y_{1:N},1)}\right] = \Pr(\text{Accept batch } B) \le 1 \tag{237}$$

Another way to see this is through the following arguments:

$$\begin{aligned}
&\mathbb{E}_{Y_{1:N}\sim Q(.)}\left[\frac{N}{\bar{Z}_x(Y_{1:N},1)}\right] && (238)\\
&= \mathbb{E}_{Y_{1:N}\sim Q(.)}\left[\frac{NP_x(Y_1)/Q(Y_1)}{\bar{Z}_x(Y_{1:N},1)}\right] && (239)\\
&= \mathbb{E}_{Y_{1:N}\sim Q(.)}\left[\frac{NP_x(Y_1)/Q(Y_1)}{\hat{Z}_x(Y_{1:N})}\frac{\hat{Z}_x(Y_{1:N})}{\bar{Z}_x(Y_{1:N},1)}\right] && (240)\\
&\le \mathbb{E}_{Y_{1:N}\sim Q(.)}\left[\frac{NP_x(Y_1)/Q(Y_1)}{\hat{Z}_x(Y_{1:N})}\right] \quad \left(\text{Since } \frac{\hat{Z}_x(Y_{1:N})}{\bar{Z}_x(Y_{1:N},1)} \le 1\right) && (241)\\
&= \sum_{i=1}^{N} \mathbb{E}_{Y_{1:N}\sim Q(.)}\left[\frac{P_x(Y_i)/Q(Y_i)}{\hat{Z}_x(Y_{1:N})}\right] \quad (\text{Due to symmetry}) && (242)\\
&= 1, && (243)
\end{aligned}$$

and as a consequence (which we will be using later), we have:

$$\begin{aligned}
&\mathbb{E}_{Y_{1:N}\sim Q(.)}\left[\frac{N+1}{\omega_x + \hat{Z}_x(Y_{1:N})}\right] && (244)\\
&= \mathbb{E}_{Y_1^{N+1}\sim Q(.)}\left[\frac{N+1}{\omega_x + \hat{Z}_x(Y_{1:N})}\right] && (245)\\
&\le 1. && (246)
\end{aligned}$$

Then, observe that:

$$
\begin{aligned}
E_1 &= \mathbb{E}_{Y_{1:N} \sim Q(.)} \left[\frac{N}{\Delta_x \bar{Z}_x(Y_{1:N}, 1)} \log \left(\frac{N}{\Delta_x \bar{Z}_x(Y_{1:N}, 1)} \right) \right] && (247) \\
&= \frac{1}{\Delta_x} \mathbb{E}_{Y_{1:N} \sim Q(.)} \left[\frac{N}{\bar{Z}_x(Y_{1:N}, 1)} \log \left(\frac{N}{\bar{Z}_x(Y_{1:N}, 1)} \right) \right] + \log \frac{1}{\Delta_x} && (248) \\
&\leq 3 \text{ bits} && (249)
\end{aligned}
$$

where, to show the inequality at the end, we will prove the following two inequalities:

$$
\begin{aligned}
&\log \frac{1}{\Delta_x} \leq \log \left(\frac{8}{7} \right) && (250) \\
&\frac{1}{\Delta_x} \mathbb{E}_{Y_{1:N} \sim Q(.)} \left[\frac{N}{\bar{Z}_x(Y_{1:N}, 1)} \log \left(\frac{N}{\bar{Z}_x(Y_{1:N}, 1)} \right) \right] \leq \frac{16}{7}, && (251)
\end{aligned}
$$

and hence $E_2 \leq 3$ (bits). For the first inequality, we have:

$$
\begin{aligned}
\Delta_x &= \mathbb{E}_{Y_{1:N} \sim Q(.)} \left[\frac{N}{\bar{Z}_x(Y_{1:N}, 1)} \right] && (252) \\
&\geq \frac{N}{\mathbb{E}_{Y_{1:N} \sim Q(.)}[\bar{Z}_x(Y_{1:N}, 1)]} \quad \text{(Jensen's Inequality)} && (253) \\
&= \frac{N}{N - 1 + \omega_x} && (254) \\
&\geq \frac{N}{N - 1 + N/7} \quad \text{(Since } N > 7\omega_x \text{)} && (255) \\
&\geq \frac{7}{8}, && (256)
\end{aligned}
$$

hence, we have:

$$
\frac{1}{\Delta_x} \leq 8/7, \tag{257}
$$

which yields the first inequality after taking the $\log(.)$ in both sides.

For the second inequality, we begin by establishing the following key inequality:

$$
\frac{N}{\bar{Z}_x(Y_{1:N}, 1)} \leq \frac{2N}{\hat{Z}_x(Y_{1:N}) + \omega_x}, \tag{258}
$$

which is due to:

$$
\begin{aligned}
\frac{N}{\bar{Z}_x(Y_{1:N}, 1)} &= \frac{N}{\omega_x + \sum_{i=2}^{N} \frac{P_x(Y_i)}{Q(Y_i)}} && (259) \\
&\leq \frac{N}{\omega_x + \frac{1}{2} \sum_{i=2}^{N} \frac{P_x(Y_i)}{Q(Y_i)}} \quad \text{(Since } \frac{P_x(Y_i)}{Q(Y_i)} \geq 0 \text{ for all } i \text{)} && (260) \\
&= \frac{2N}{2\omega_x + \sum_{i=2}^{N} \frac{P_x(Y_i)}{Q(Y_i)}} && (261) \\
&\leq \frac{2N}{\omega_x + \sum_{i=1}^{N} \frac{P_x(Y_i)}{Q(Y_i)}} \quad \text{(Since } \frac{P_x(Y_i)}{Q(Y_i)} \leq \omega \text{ for all } i \text{)} && (262) \\
&= \frac{2N}{\hat{Z}_x(Y_{1:N}) + \omega_x}, && (263)
\end{aligned}
$$

Then, we have:

$$
\begin{aligned}
&\frac{1}{\Delta_x}\mathbb{E}_{Y_{1:N}\sim Q(.)}\left[\frac{N}{\bar{Z}_x(Y_{1:N},1)}\log\left(\frac{N}{\bar{Z}_x(Y_{1:N},1)}\right)\right] && (264)\\
\leq&\frac{8}{7}\mathbb{E}_{Y_{1:N}\sim Q(.)}\left[\frac{N}{\bar{Z}_x(Y_{1:N},1)}\log\left(\frac{N}{\bar{Z}_x(Y_{1:N},1)}\right)\right] \quad \text{(Since } \Delta_x \geq \frac{7}{8} \text{ from (256))} && (265)\\
=&\frac{8}{7}\mathbb{E}_{Y_{1:N}\sim Q(.)}\left[\frac{NP_x(Y_1)/Q(Y_1)}{\bar{Z}_x(Y_{1:N},1)}\log\left(\frac{N}{\bar{Z}_x(Y_{1:N},1)}\right)\right] && (266)\\
\leq&\frac{8}{7}\mathbb{E}_{Y_{1:N}\sim Q(.)}\left[\frac{NP_x(Y_1)/Q(Y_1)}{\bar{Z}_x(Y_{1:N},1)}\log\left(\frac{N}{\bar{Z}_x(Y_{1:N},1)}+1\right)\right] && (267)\\
\leq&\frac{8}{7}\mathbb{E}_{Y_{1:N}\sim Q(.)}\left[\frac{NP_x(Y_1)/Q(Y_1)}{\bar{Z}_x(Y_{1:N},1)}\log\left(\frac{2N}{\hat{Z}_x(Y_{1:N})+\omega_x}+1\right)\right] \text{ (Due to Inequality (258))} && (268)\\
\leq&\frac{8}{7}\mathbb{E}_{Y_{1:N}\sim Q(.)}\left[\frac{NP_x(Y_1)/Q(Y_1)}{\hat{Z}_x(Y_{1:N})}\log\left(\frac{2N}{\hat{Z}_x(Y_{1:N})+\omega_x}+1\right)\right] \text{ (Since } \hat{Z}_x(Y_{1:N})\leq\bar{Z}_x(Y_{1:N},1)) && (269)\\
=&\frac{8}{7}\sum_{i=1}^{N}\mathbb{E}_{Y_{1:N}\sim Q(.)}\left[\frac{P_x(Y_i)/Q(Y_i)}{\hat{Z}_x(Y_{1:N})}\log\left(\frac{2N}{\hat{Z}_x(Y_{1:N})+\omega_x}+1\right)\right] \quad \text{(Due to symmetry)} && (270)\\
=&\frac{8}{7}\mathbb{E}_{Y_{1:N}\sim Q(.)}\left[\frac{\sum_{i=1}^{N}P_x(Y_i)/Q(Y_i)}{\hat{Z}_x(Y_{1:N})}\log\left(\frac{2N}{\hat{Z}_x(Y_{1:N})+\omega_x}+1\right)\right] && (271)\\
=&\frac{8}{7}\mathbb{E}_{Y_{1:N}\sim Q(.)}\left[\log\left(\frac{2N}{\hat{Z}_x(Y_{1:N})+\omega_x}+1\right)\right] && (272)\\
\leq&\frac{8}{7}\log\left(\mathbb{E}_{Y_{1:N}\sim Q(.)}\left[\frac{2N}{\hat{Z}_x(Y_{1:N})+\omega_x}+1\right]\right) \quad \text{(Jensen's Inequality)} && (273)\\
=&\frac{8}{7}\log\left(1+\frac{2N}{N+1}\mathbb{E}_{Y_{1:N}\sim Q(.)}\left[\frac{N+1}{\hat{Z}_x(Y_{1:N})+\omega_x}\right]\right) && (274)\\
\leq&\frac{8}{7}\log\left(1+\frac{2N}{N+1}\right) \text{ (Since } \mathbb{E}_{Y_{1:N}\sim Q(.)}\left[\frac{N+1}{\hat{Z}_x(Y_{1:N})+\omega_x}\right] < 1 \text{ due to Inequality (246))} && (275)\\
\leq&\frac{8}{7}\log(4) && (276)\\
=&\frac{16}{7}\text{(bits)} && (277)
\end{aligned}
$$

which completes the proof for this part.

E.4.2 Proof of Inequality (164)

We first express the quantity $\mathbb{E}[\log \hat{K}_1]$ with conditional expectation. The accepted batch and selected local index K_2 are distributed according to $Y_{K_1,1:N}, K_2 \sim \bar{P}_{Y_{1:N},K;x}$, then:

$$
\begin{aligned}
&\mathbb{E}[\log \hat{K}_1] && (278)\\
&= \mathbb{E}[\mathbb{E}[\log \hat{K}_1 | Y_{K_1,1:N} = y_{1:N}, K_2 = k_2]] && (279)\\
&= \sum_{k_2=1}^{N}\int_{-\infty}^{\infty}\left(\prod_{j=1,j\neq k_2}^{N} Q(y_j)\right)\frac{P_x(y_k)}{\bar{Z}_x(y_{1:N},k)\Delta_x}\mathbb{E}[\log \hat{K}_1 | Y_{K_1,1:N} = y_{1:N}, K_2 = k_2] dy_{1:N} && (280)\\
&= N\int_{-\infty}^{\infty}\left(\prod_{j=2}^{N} Q(y_j)\right)\frac{P_x(y_1)}{\bar{Z}_x(y_{1:N},1)\Delta_x}\mathbb{E}[\log \hat{K}_1 | Y_{K_1,1:N} = y_{1:N}, K_2 = 1] dy_{1:N} && (281)
\end{aligned}
$$

Notice that, since we accept a batch i when $U_i \leq \frac{\hat{Z}_x(y_{1:N})}{\bar{Z}_x(y_{1:N},1)}\frac{\Delta}{\Delta_x}$, we have that:

$$P(U_{K_1} = u|Y_{K,1:N} = y_{1:N}, K_2 = 1) = \frac{\bar{Z}_x(y_{1:N},1)}{\hat{Z}_x(y_{1:N})}\frac{\Delta_x}{\Delta},$$

then conditioning on U_{K_1} for the last expectation term above:

$$\mathbb{E}[\log \hat{K}_1|Y_{K_1,1:N} = y_{1:N}, K_2 = k] \tag{282}$$

$$= \int_{-\infty}^{\infty} \mathbb{E}[\log \hat{K}_1|Y_{K_1,1:N} = y_{1:N}, K_2 = 1, U_{K_1} = u]P(U_{K_1} = u|Y_{K_1,1:N} = y_{1:N}, K_2 = 1)du \tag{283}$$

$$= \int_0^{\frac{\hat{Z}_x(y_{1:N})}{\bar{Z}_x(y_{1:N},1)}\frac{\Delta}{\Delta_x}} \mathbb{E}[\log \hat{K}_1|Y_{K_1,1:N} = y_{1:N}, K_2 = 1, U_{K_1} = u]\frac{\bar{Z}_x(y_{1:N},1)}{\hat{Z}_x(y_{1:N})}\frac{\Delta_x}{\Delta}du \tag{284}$$

$$\leq \frac{\Delta_x}{\Delta}\int_0^{\frac{\hat{Z}_x(y_{1:N})}{\bar{Z}_x(y_{1:N},1)}\frac{\Delta}{\Delta_x}} \frac{\bar{Z}_x(y_{1:N},1)}{\hat{Z}_x(y_{1:N})}\log\left[1 + \frac{u}{\Delta}\right]du \quad \text{(See the Sort Coding bound below.)} \tag{285}$$

$$\leq \frac{\Delta_x}{\Delta}\int_0^{\frac{\hat{Z}_x(y_{1:N})}{\bar{Z}_x(y_{1:N},1)}\frac{\Delta}{\Delta_x}} \frac{\bar{Z}_x(y_{1:N},1)}{\hat{Z}_x(y_{1:N})}\log\left[1 + \frac{\hat{Z}_x(y_{1:N})}{\Delta_x\bar{Z}_x(y_{1:N},1)}\right]du \tag{286}$$

$$= \log\left[1 + \frac{\hat{Z}_x(y_{1:N})}{\Delta_x\bar{Z}_x(y_{1:N},1)}\right], \tag{287}$$

Finally, we have:

$$\mathbb{E}[\log \hat{K}_1] \tag{288}$$

$$= N\int_{-\infty}^{\infty}\left(\prod_{j=2}^{N} Q(y_j)\right)\frac{P_x(y_1)}{\bar{Z}_x(y_{1:N},1)\Delta_x}\log\left[1 + \frac{\hat{Z}_x(y_{1:N})}{\Delta_x\bar{Z}_x(y_{1:N},1)}\right]dy_{1:N} \tag{289}$$

$$\leq N\int_{-\infty}^{\infty}\left(\prod_{j=2}^{N} Q(y_j)\right)\frac{P_x(y_1)}{\bar{Z}_x(y_{1:N},1)\Delta_x}\left(\log\left[\frac{\hat{Z}_x(y_{1:N})}{\Delta_x\bar{Z}_x(y_{1:N},1)}\right] + \log e\frac{\Delta_x\bar{Z}_x(y_{1:N},1)}{\hat{Z}_x(y_{1:N})}\right)dy_{1:N} \tag{290}$$

$$= N\int_{-\infty}^{\infty}\left(\prod_{j=2}^{N} Q(y_j)\right)\frac{P_x(y_1)}{\bar{Z}_x(y_{1:N},1)\Delta_x}\log\left[\frac{\hat{Z}_x(y_{1:N})}{\Delta_x\bar{Z}_x(y_{1:N},1)}\right]dy_{1:N} + \log(e) \tag{291}$$

$$= \frac{N}{\Delta_x}\mathbb{E}_{Y_{1:N}\sim Q}\left[\frac{P_x(Y_1)/Q(Y_1)}{\bar{Z}_x(Y_{1:N},1)}\log\left(\frac{\hat{Z}_x(Y_{1:N})}{\Delta_x\bar{Z}_x(Y_{1:N},1)}\right)\right] \tag{292}$$

We show the proof for (285) below.

Sort Coding Bound. To bound the expectation term, we first apply Jensen's inequality and conditioning on the accepted chunk of batches $L = \ell$:

$$\mathbb{E}[\log \hat{K}_1|Y_{K_1,1:N} = y_{1:N}, K_2 = 1, U_{K_1} = u] \tag{293}$$

$$\leq \log(\mathbb{E}[\hat{K}_1|Y_{K_1,1:N} = y_{1:N}, K_2 = 1, U_{K_1} = u]) \tag{294}$$

$$= \log(\mathbb{E}[\hat{K}_1|Y_{K_1,1:N} = y_{1:N}, K_2 = 1, U_{K_1} = u]) \tag{295}$$

$$= \log(\mathbb{E}_L[\mathbb{E}[\hat{K}_1|Y_{K_1,1:N} = y_{1:N}, K_2 = 1, U_{K_1} = u, L = \ell]]) \tag{296}$$

We now repeat the previous argument in standard RS. Specifically, given $\hat{K}_1$ is within the range $L = \ell$ and $U_{K_1} = u$, we can express $\hat{K}_1$ as follows:

$$\hat{K}_1 = |\{U_i < u, (\ell - 1)\lfloor\Delta^{-1}\rfloor + 1 \leq i \leq \ell\lfloor\Delta^{-1}\rfloor\}| + 1, \tag{297}$$

$$= \Omega(u, \ell) + 1 \tag{298}$$

i.e. the number of U_i (plus 1 for the ranking) within the range L that has value lesser than u.

We can see that the the index i within the range L satisfying $U_i < u$ are from the indices that are either (1) rejected, i.e. index $i < \hat{K}_1$ or (2) not examined by the algorithm, i.e. index $i > \hat{K}_1$. The rest of this proof will show the following bound:

$$\mathbb{E}[\Omega(u,\ell)|Y_{K_1,1:N} = y_{1:N}, K_2 = 1, U_{K_1} = u, L = \ell] \leq \Delta^{-1}u, \text{ for any } \ell \tag{299}$$

For readability, we split the proof into different proof steps.

Proof Step 1: We condition on the mapped index of $\pi(\hat{K})$ on the original array:

$$\begin{aligned}
&\mathbb{E}[\hat{K}_1|Y_{K_1,1:N} = y_{1:N}, K_2 = 1, U_{K_1} = u, L = \ell] && (300)\\
&= \mathbb{E}_{\pi(\hat{K}_1)}\left[\mathbb{E}[\hat{K}_1 \mid Y_{K_1,1:N} = y_{1:N}, K_2 = 1, U_{K_1} = u, L = \ell, \pi(\hat{K}_1) = k_1]\right] && (301)\\
&= \mathbb{E}_{\pi(\hat{K}_1)}\left[\mathbb{E}[\Omega(u,\ell) + 1 \mid Y_{K_1,1:N} = y_{1:N}, K_2 = 1, U_{K_1} = u, L = \ell, \pi(\hat{K}_1) = k_1]\right] && (302)\\
&= \mathbb{E}_{\pi(\hat{K}_1)}\left[\mathbb{E}[\Omega(u,\ell) \mid Y_{K_1,1:N} = y_{1:N}, K_2 = 1, U_{K_1} = u, L = \ell, \pi(\hat{K}_1) = k_1]\right] + 1 && (303)\\
&= \mathbb{E}_{\pi(\hat{K}_1)}\left[\mathbb{E}[\Omega_1(u,\ell,k_1) + \Omega_2(u,\ell,k_1) \mid Y_{K_1,1:N} = y_{1:N}, K_2 = 1, U_{K_1} = u, L = \ell, \pi(\hat{K}_1) = k_1]\right] + 1, && (304)
\end{aligned}$$

where $\Omega_1(u,\ell,k_1), \Omega_2(u,\ell,k_1)$ are the number of $U_i < u$ within the range $L = \ell$ that occurs before and after the selected index k_1 respectively. Specifically:

$$\begin{aligned}
\Omega_1(u,\ell,k_1) &= |\{U_i < u, (\ell-1)\lfloor\Delta^{-1}\rfloor + 1 \leq i < (\ell-1)\lfloor\Delta^{-1}\rfloor + k_1\}| && (305)\\
\Omega_2(u,\ell,k_1) &= |\{U_i < u, (\ell-1)\lfloor\Delta^{-1}\rfloor + k_1 + 1 \leq i \leq \ell\lfloor\Delta^{-1}\rfloor\}|, && (306)
\end{aligned}$$

which also naturally gives $\Omega(u,\ell) = \Omega_1(u,\ell,k_1) + \Omega_2(u,\ell,k_1)$.

Proof Step 2: Consider $\Omega_2(u,\ell,k_1)$, since each proposal $(Y_{i,1:N}, U_i)$ is i.i.d distributed and the fact that k_1 is the index of the *first accepted batch*, for every $i > k_1$, we have:

$$\Pr(U_i < u \mid \bar{Y}_{1:N} = y_{1:N}, K_2 = 1, U_{K_1} = u, L = \ell, \pi(\hat{K}_1) = k_1) = \Pr(U_i < u)$$

This gives us:

$$\begin{aligned}
&\mathbb{E}[\Omega_2(u,\ell,k_1) \mid Y_{K_1,1:N} = y_{1:N}, K_2 = 1, U_{K_1} = u, L = \ell, \pi(\hat{K}_1) = k_1] && (307)\\
&= (\lfloor\Delta^{-1}\rfloor - k_1)\Pr(U < u) && (308)\\
&= (\lfloor\Delta^{-1}\rfloor - k_1)u && (309)\\
&\leq \frac{(\lfloor\Delta^{-1}\rfloor - k_1)u}{\Pr(\text{Batch is rejected})} && (310)\\
&\leq \frac{(\lfloor\Delta^{-1}\rfloor - k_1)u}{1-\Delta} && (311)
\end{aligned}$$

Proof Step 3: For $\Omega_1(u,\ell,\hat{k}_1)$, we do not have such independent property since for every batch with index $i < k_1$, we know that they are rejected batches, and hence for $i < k_1$:

$$\begin{aligned}
&\Pr(U_i < u \mid Y_{K_1,1:N} = y_{1:N}, K_2 = k, U_{K_1} = u, L = \ell, \pi(\hat{K}_1) = k_1) && (312)\\
&= \Pr(U_i < u | Y_{i,1:N} \text{ is rejected}) && (313)\\
&= \frac{\Pr(U_i < u, Y_{i,1:N} \text{ is rejected})}{\Pr(Y_{i,1:N} \text{ is rejected})} && (314)\\
&\leq \frac{\Pr(U_i < u)}{\Pr(Y_{i,1:N} \text{ is rejected})} && (315)\\
&= \frac{u}{1-\Delta}, && (316)
\end{aligned}$$

which gives us:

$$\mathbb{E}[\Omega_2(u,\ell,k_1) \mid Y_{K_1,1:N} = y_{1:N}, K_2 = 1, U_{K_1} = u, L = \ell, \pi(\hat{K}_1) = k_1] \leq \frac{(k_1-1)u}{1-\Delta} \tag{317}$$

To prove Equation (313), note that the following events are equivalent:

$$\{Y_{K_1,1:N} = y_{1:N}, K_2 = 1, U_{K_1} = u, L = \ell, \pi(\hat{K}_1) = k_1\} \tag{318}$$
$$= \{Y_{k_1,1:N} = y_{1:N}, K_2 = 1, U_k = u, B_{1,..,k-1} \text{ are rejected}\} \tag{319}$$
$$\triangleq \Lambda(u, y, k_1) \tag{320}$$

Here, we note that Y_{k_1}, U_{k_1} denote the value at batch index k within W, which is different from Y_{K_1}, U_{K_1}, the value selected by the rejection sampler. Hence:

$$\Pr(U_i < u | \Lambda(u, y, k_1)) \tag{321}$$
$$= \frac{\Pr(U_i < u, B_{1\ldots k_1-1} \text{ are rejected} | Y_{k,1:N} = y_{1:N}, U_k = u, K_2 = 1)}{\Pr(B_{1\ldots k_1-1} \text{ are rejected} | Y_{k,1:N} = y_{1:N}, U_k = u, K_2 = 1)} \tag{322}$$
$$= \frac{\Pr(U_i < u,\ B_{1\ldots k_1-1} \text{ are rejected})}{\Pr(B_{1\ldots k_1-1} \text{ are rejected})} \quad \text{(Since } (Y_i, U_i) \text{ are i.i.d)} \tag{323}$$
$$= \Pr(U_i < u | B_i \text{ is rejected}), \tag{324}$$

Proof Step 4: From the above result from Step 2 and 3, we have $\Omega(u, \ell) = \Omega_1(u, \ell, k) + \Omega_2(u, \ell, k) \leq \frac{(\lfloor \Delta^{-1} \rfloor - 1)u}{1-\Delta}$ and as a result:

$$\mathbb{E}[\log \hat{K}_1 | Y_{K_1,1:N} = y_{1:N}, K_2 = 1, U_{K_1} = u] \leq \frac{(\lfloor \Delta^{-1} \rfloor - 1)u}{1 - \Delta} + 1 \tag{325}$$
$$\leq \frac{(\Delta^{-1} - 1)u}{1 - \Delta} + 1 \quad \text{(Since} \lfloor \Delta^{-1} \rfloor \leq \Delta^{-1}) \tag{326}$$
$$= \Delta^{-1} u + 1 \tag{327}$$

which completes the proof.

F ERS Matching Lemmas

F.1 Preliminaries

We begin by providing the following bounds on inverse moments of averages.

Proposition F.1. *Let $Y_1, Y_2, ..., Y_N \sim Q_Y(.)$ and suppose the target distribution P_Y satifies:*

$$d_2(Q_Y||P_Y) \triangleq \mathbb{E}_{Y\sim Q_Y(.)}\left[\frac{Q_Y(Y)}{P_Y(Y)}\right] < \infty, \tag{328}$$

then we have:

$$\mathbb{E}_{Y_{1:N}\sim Q_Y(.)}\left[\frac{N}{\sum_{i=1}^{N}\frac{P_Y(Y_i)}{Q_Y(Y_i)}}\right] \leq d_2(Q_Y||P_Y). \tag{329}$$

Proof. Applying the Cauchy-Schwarz inequality, we have:

$$\frac{N}{\sum_{i=1}^{N}\frac{P_Y(Y_i)}{Q_Y(Y_i)}} \leq \frac{1}{N}\sum_{j=1}^{N}\frac{Q_Y(Y_i)}{P_Y(Y_i)} \tag{330}$$

Taking the expectation of both sides yield the desired inequality. □

Remark F.2. *In general, stronger results on the inverse moments of averages exist under weaker moment assumptions, specifically:*

$$\mathbb{E}_{Y\sim Q_Y}\left[\left(\frac{Q_Y(Y)}{P_Y(Y)}\right)^{\eta}\right]$$

is finite for some $\eta < 0$. The resulting bound has a similar form (some power terms involved) to that of Proposition F.1 but requires a mild threshold on N. For further details, see Proposition A.1 in [9].

We show an application of Proposition F.1, which we will use repeatly:

Corollary F.3. *Let $Y_1, Y_2, ..., Y_N \sim Q_Y(.)$ and suppose the target distributions P_Y^A, P_Y^B satisfy:*

$$d_2(Q_Y||P_Y^A) \triangleq \mathbb{E}_{Y\sim Q_Y(.)}\left[\frac{Q_Y(Y)}{P_Y^A(Y)}\right] < \infty, \quad \textit{and} \quad \frac{P_Y^A(y)}{Q_Y(y)}, \frac{P_Y^B(y)}{Q_Y(y)} \leq \omega \textit{ for all } y. \tag{331}$$

Then, for any $N \geq 1$,

$$\mathbb{E}_{Y_{1:N}\sim Q_Y(.)}\left[\frac{\sum_{j=1}^{N}\frac{P_Y^B(Y_j)}{Q_Y(Y_j)}}{\sum_{i=1}^{N}\frac{P_Y^A(Y_i)}{Q_Y(Y_i)}}\right] \leq \mathbb{I}_N(\omega, 1)\cdot d_2(Q_Y||P_Y), \tag{332}$$

where we define $\mathbb{I}_N(\omega, i) \triangleq (2\mathbb{1}_{N>i} + \omega\mathbb{1}_{N=i})$.

Proof. For $N = 1$, applying the conditions for P_Y^A and P_Y^B gives us an upper-bound of $\omega d_2(Q_Y||P_Y^A)$. For $N > 1$, we have:

$$\begin{aligned}
\mathbb{E}_{Y_{1:N}\sim Q_Y(.)}\left[\frac{\sum_{j=1}^{N}\frac{P_Y^B(Y_i)}{Q_Y(Y_i)}}{\sum_{i=1}^{N}\frac{P_Y^A(Y_i)}{Q_Y(Y_i)}}\right] &= N\mathbb{E}_{Y_{1:N}\sim Q_Y(.)}\left[\frac{\frac{P_Y^B(Y_1)}{Q_Y(Y_1)}}{\sum_{i=1}^{N}\frac{P_Y^A(Y_i)}{Q_Y(Y_i)}}\right] \text{ (Due to symmetry)} && (333)\\
&\leq N\mathbb{E}_{Y_{1:N}\sim Q_Y(.)}\left[\frac{\frac{P_Y^B(Y_1)}{Q_Y(Y_1)}}{\sum_{i=2}^{N}\frac{P_Y^A(Y_i)}{Q_Y(Y_i)}}\right] \text{ (since } \frac{P_Y^A(Y_i)}{Q_Y(Y_i)} \geq 0) && (334)\\
&= \frac{N}{N-1}\mathbb{E}_{Y_{1:N}\sim Q_Y(.)}\left[\frac{N-1}{\sum_{i=2}^{N}\frac{P_Y^A(Y_i)}{Q_Y(Y_i)}}\right] && (335)\\
&\leq 2d_2(Q_Y||P_Y^A) \quad \text{(Proposition F.1 and } N > 1), && (336)
\end{aligned}$$

which completes the proof. □

F.2 Distributed Matching Without Batch Communication

Before the proof, we outline the details of each case in Section 3.2, covering scenarios without and with communication between the encoder and decoder.

Without-Communication. In this scenario, let $P_Y^A(.)$ and $P_Y^B(.)$ be the target distributions at the encoder and decoder respectively, where we use the same shared randomness W as in Section 5.1 where we use the proposal distribution $Q_Y(.)$. Furthermore, we assume that:

$$\max_y \left(\frac{P_Y^A(y)}{Q_Y(y)}\right) = \omega_A, \quad \max_y \left(\frac{P_Y^B(y)}{Q_Y(y)}\right) = \omega_B, \quad \max_y \left(\frac{P_Y^A(y)}{Q_Y(y)}, \frac{P_Y^B(y)}{Q_Y(y)}\right) \leq \omega, \tag{337}$$

Using the ERS procedure, the encoder and decoder select the indices K_A and K_B respectively.

$$K_A = \mathrm{ERS}(W; P_Y^A, Q_Y), \quad K_B = \mathrm{ERS}(W; P_Y^B, Q_Y), \tag{338}$$

The $\mathrm{ERS}(\cdot)$ function follows Algorithm 1, without requiring any specific scaling factor. During the selection process, the calculation of $\bar{Z}$ in Step 3 of this algorithm, which determines the acceptance probability, uses the ratio upper bounds ω_A and ω_B for parties A and B, respectively. Proposition F.4 establishes the bound on the probability that both parties produce the same output, conditioned on $Y_{K_A} = y$.

Proposition F.4. *Let K_A, K_B and P_Y^A, P_Y^B defined as above and $N \geq 2$, we have:*

$$\Pr(Y_{K_A} = Y_{K_B} | Y_{K_A} = y) \geq \left(1 + \mu_1(N) + \frac{P_Y^A(y)}{P_Y^B(y)}\left(1 + \mu_2(N)\right)\right)^{-1}, \tag{339}$$

where $\mu_1(N)$ and $\mu_2(N)$ are defined as in Appendix F.3 and we note that $\mu_1(N), \mu_2(N) \to 0$ as $N \to \infty$ under mild assumptions on the distributions P_Y^A, P_Y^B and Q_Y.

Proof. See Appendix F.4. □

With Communication. Following the setup described in Section 5.3, we define the ratio upperbounds in the communication case as below:

$$\max_z \left(\frac{P_{Y|Z}(y|x)}{Q_Y(y)}\right) = \omega_x, \quad \max_z \left(\frac{\tilde{P}_{Y|Z}(y|z)}{Q_Y(y)}\right) = \omega_z, \quad \max_{y,z} \left(\frac{P_{Y|X}(y|x)}{Q_Y(y)}, \frac{\tilde{P}_{Y|Z}(y|z)}{Q_Y(y)}\right) \leq \omega,$$

and similar to the case without communication, the $\mathrm{ERS}(.)$ selection process at the encoder and decoder also follows Algorithm 1, with the calculation of $\bar{Z}$ in Step 3 uses the upperbound ω_x and ω_z respectively for the encoder and decoder. The bound for this case is shown below.

Proposition F.5. *For $N \geq 2$ and X, Y, Z defined as above, we have:*

$$\Pr(Y_{K_A}{=}Y_{K_B} | Y_{K_A}{=}y, X{=}x, Z{=}z) \geq \left(1{+}\mu_1^{\mathrm{cond}}(N){+}\frac{P_{Y|X}(y|x)}{\tilde{P}_{Y|Z}(y|z)}\left(1{+}\mu_2^{\mathrm{cond}}(N)\right)\right)^{-1}, \tag{340}$$

where $\mu_1^{\mathrm{cond}}(N)$ and $\mu_2^{\mathrm{cond}}(N)$ are defined as in Appendix F.5 and we note that $\mu_1^{\mathrm{cond}}(N), \mu_2^{\mathrm{cond}}(N) \to 0$ as $N \to \infty$ under mild assumptions on the distributions $P_{Y|X}(.|x), \tilde{P}_{Y|Z}(.|z)$ and $Q_Y(.)$.

Proof. See Appendix F.6. □

F.3 Coefficients in Proposition F.4

We first define the coefficient $\mu_1(N)$ and $\mu_2(N)$ in Proposition F.4.

$$\mu_1(N) = \frac{1}{N}\left[\omega + \omega \mathbb{I}_N(\omega, 2) d_2(Q_Y || P_Y^B) + \frac{\omega^2}{N-1} d_2(Q_Y || P_Y^B)\right] \tag{341}$$

$$\mu_2(N) = \frac{1}{N}\left[\omega + \omega \mathbb{I}_N(\omega, 2) d_2(Q_Y || P_Y^A) + \frac{\omega^2}{N-1} d_2(Q_Y || P_Y^A)\right] \tag{342}$$

where we define $\mathbb{I}_N(\omega, i) \triangleq (2\mathbb{1}_{N>i} + \omega \mathbb{1}_{N=i})$ as in Proposition F.3.

F.4 Proof of Proposition F.4

We prove the matching probability for the case of ERS. We note that in this proof, we will use the global index for the proposals $Y_1, ...Y_N \sim Q(.)$ instead of $Y_{1,1}, ...Y_{1,N}$ unless otherwise stated. First, consider:

$$
\begin{aligned}
&\Pr(Y_{K_A} = Y_{K_B} | Y_{K_A} = y_1) && (343)\\
&\geq \Pr(K_A{=}K_B | Y_{K_A}{=}y_1) && (344)\\
&= \sum_{k=1}^{\infty} \Pr(K_A = K_B = k | Y_{K_A} = y_1) && (345)\\
&\geq \sum_{k=1}^{N} \Pr(K_A = K_B = k | Y_{K_A} = y_1) && (346)\\
&= N \Pr(K_A = K_B = 1 | Y_{K_A} = y_1) && (347)\\
&= \frac{N Q_Y(y_1)}{P_Y^A(y_1)} \Pr(K_{2,A} = K_{2,B} = 1, K_{1,A} = K_{1,B} = 1 | Y_1 = y_1) && (348)\\
&= \frac{N Q_Y(y_1)}{P_Y^A(y_1)} \int \Pr(K_{2,A} = K_{2,B} = 1, K_{1,A} = K_{1,B} = 1, Y_{2:N} = y_{2:N} | Y_1 = y_1) dy_{2:N} && (349)\\
&= \frac{N Q_Y(y_1)}{P_Y^A(y_1)} \int \Pr(K_{2,A} = K_{2,B} = 1, K_{1,A} = K_{1,B} = 1 | Y_{1:N}{=}y_{1:N}) Q_Y(y_{2:N}) dy_{2:N} && (350)\\
&= \frac{N Q_Y(y_1)}{P_Y^A(y_1)} \int \Pr(K_{2,A}{=}K_{2,B}{=}1 | Y_{1:N}{=}y_{1:N}) \\
&\qquad \times \Pr(K_{1,A}{=}K_{1,B}{=}1 | K_{2,A}{=}K_{2,B}{=}1, Y_{1:N}{=}y_{1:N}) Q_Y(y_{2:N}) dy_{2:N} && (351)\\
&= \frac{N Q_Y(y_1)}{P_Y^A(y_1)} \mathbb{E}_{Y_{2:N} \sim Q_Y(.)} [\Pr(K_{2,A}{=}K_{2,B}{=}1 | Y_{1:N}{=}y_{1:N}) \\
&\qquad \times \Pr(K_{1,A}{=}K_{1,B}{=}1 | K_{2,A}{=}K_{2,B}{=}1, Y_{1:N}{=}y_{1:N})] && (352)
\end{aligned}
$$

where (348) is due to the following fact that:

$$\{K_A = K_B = 1, Y_{K_A} = y_1\} = \{K_A = K_B = 1, Y_1 = y_1\}, \tag{353}$$

and thus:

$$
\begin{aligned}
\Pr(K_A = K_B = 1 | Y_{K_A} = y_1) &= \frac{\Pr(K_A = K_B = 1 | Y_1 = y_1) Q_Y(y_1)}{P(Y_{K_A} = y_1)} && (354)\\
&= \frac{\Pr(K_A = K_B = 1 | Y_1 = y_1) Q_Y(y_1)}{P_Y^A(y_1)} && (355)\\
& && (356)
\end{aligned}
$$

Define:

$$\hat{Z}(P_Y^A, y_{1:N}) = \sum_{i=1}^{N} \frac{P_Y^A(y_i)}{Q_Y(y_i)}, \quad \hat{Z}(P_Y^B, y_{1:N}) = \sum_{i=1}^{N} \frac{P_Y^B(y_i)}{Q_Y(y_i)} \tag{357}$$

Now, we note that:

$$
\begin{aligned}
&\Pr(K_{2,A}{=}K_{2,B}{=}1 | Y_{1:N}{=}y_{1:N}) && (358)\\
&= \Pr(K_{2,A}{=}1 | Y_{1:N}{=}y_{1:N}) \Pr(K_{2,B}{=}1 | Y_{1:N}{=}y_{1:N}, K_{2,A} = 1) && (359)\\
&= \frac{P_Y^A(y_1)/Q_Y(y_1)}{\sum_{i=1}^N P_Y^A(y_i)/Q_Y(y_i)} \Pr(K_{2,B}{=}1 | Y_{1:N}{=}y_{1:N}, K_{2,A} = 1) && (360)\\
&\geq \frac{P_Y^A(y_1)/Q_Y(y_1)}{\hat{Z}(P_Y^A, y_{1:N})} \left(1 + \frac{P_Y^A(y_1)}{P_Y^B(y_1)} \cdot \frac{\hat{Z}(P_Y^B, y_{1:N})}{\hat{Z}(P_Y^A, y_{1:N})}\right)^{-1}, && (361)
\end{aligned}
$$

where we denote $\hat{Z}(P_Y^A, y_{1:N}) = \sum_{i=1}^N P_Y^A(y_i)/Q_Y(y_i)$ and the last inequality is due to Proposition 1 in [32]. Also:

$$\Pr(K_{1,A}(1){=}K_{1,B}(1){=}1|K_{2,A}{=}K_{2,B}{=}1, Y_{1:N}{=}y_{1:N}) \tag{362}$$

$$\geq \min\left(\frac{\hat{Z}(P_Y^A, y_{1:N})}{\hat{Z}(P_Y^A, y_{2:N})+\omega}, \frac{\hat{Z}(P_Y^B, y_{1:N})}{\hat{Z}(P_Y^B, y_{2:N})+\omega}\right) (\text{ Since } \omega \geq \max_y \left(\frac{P_Y^A(y)}{Q_Y(y)}, \frac{P_Y^B(y)}{Q_Y(y)}\right)) \tag{363}$$

$$\geq \left(\frac{\hat{Z}(P_Y^A, y_{1:N})}{\hat{Z}(P_Y^A, y_{2:N})+\omega}\right)\left(\frac{\hat{Z}(P_Y^B, y_{1:N})}{\hat{Z}(P_Y^B, y_{2:N})+\omega}\right), \tag{364}$$

where we use the inequality $\min(a, b) \geq ab$ for $0 \leq a, b \leq 1$. Plug both in (352), we have:

$$\Pr(K_A{=}K_B|Y_{K_A}{=}y_1) \tag{365}$$

$$\geq \mathbb{E}_{Y_{2:N}\sim Q_Y(.)}\left[\frac{1}{\left(1+\frac{P_Y^A(y_1)}{P_Y^B(y_1)}\cdot\frac{\hat{Z}(P_Y^B,y_{1:N})}{\hat{Z}(P_Y^A,y_{1:N})}\right)}\left(\frac{N}{\hat{Z}(P_Y^A, y_{2:N}){+}\omega}\right)\left(\frac{\hat{Z}(P_Y^B, y_{1:N})}{\sum_{i=2}^N \hat{Z}(P_Y^B, y_{2:N}){+}\omega}\right)\right]$$

$$= \mathbb{E}_{Y_{2:N}\sim Q_Y(.)}\left[\frac{1}{\left(1{+}\frac{P_Y^A(y_1)}{P_Y^B(y_1)}\cdot\frac{\hat{Z}(P_Y^B,y_{1:N})}{\hat{Z}(P_Y^A,y_{1:N})}\right)\left(\frac{\hat{Z}(P_Y^A,y_{2:N}){+}\omega}{N}\right)\left(\frac{\hat{Z}(P_Y^B,y_{2:N}){+}\omega}{\hat{Z}(P_Y^B,y_{1:N})}\right)}\right] \tag{366}$$

$$\geq \left(\mathbb{E}_{Y_{2:N}\sim Q_Y(.)}\left[\left(1{+}\frac{P_Y^A(y_1)}{P_Y^B(y_1)}\cdot\frac{\hat{Z}(P_Y^B, y_{1:N})}{\hat{Z}(P_Y^A, y_{1:N})}\right)\left(\frac{\hat{Z}(P_Y^A, y_{2:N}){+}\omega}{N}\right)\left(\frac{\hat{Z}(P_Y^B, y_{2:N}){+}\omega}{\hat{Z}(P_Y^B, y_{1:N})}\right)\right]\right)^{-1} \tag{367}$$

$$= \left(\mathbb{E}_{Y_{2:N}\sim Q_Y(.)}\left[\zeta_1 + \frac{P_Y^A(y_1)}{P_Y^B(y_1)}\zeta_2\right]\right)^{-1}, \tag{368}$$

where we use Jensen's inequality for the convex function $1/x$ in line (367) and set:

$$\zeta_1 = \left(\frac{\hat{Z}(P_Y^A, y_{2:N}){+}\omega}{N}\right)\left(\frac{\hat{Z}(P_Y^B, y_{2:N}){+}\omega}{\hat{Z}(P_Y^B, y_{1:N})}\right) \tag{369}$$

$$= \frac{\hat{Z}(P_Y^A, y_{2:N})\cdot\hat{Z}(P_Y^B, y_{2:N})}{N\hat{Z}(P_Y^B, y_{1:N})} + \frac{\omega}{N}\cdot\frac{\hat{Z}(P_Y^A, y_{2:N})}{\hat{Z}(P_Y^B, y_{1:N})} + \frac{\omega}{N}\cdot\frac{\hat{Z}(P_Y^B, y_{2:N})}{\hat{Z}(P_Y^B, y_{1:N})} + \frac{\omega^2}{N\cdot\hat{Z}(P_Y^B, y_{1:N})}$$

$$\leq \frac{1}{N}\hat{Z}(P_Y^A, y_{2:N}) + \frac{\omega}{N}\cdot\frac{\hat{Z}(P_Y^A, y_{2:N})}{\hat{Z}(P_Y^B, y_{2:N})} + \frac{\omega}{N} + \frac{\omega^2}{N\hat{Z}(P_Y^B, y_{2:N})}, \tag{370}$$

with the last inequality due to $\sum_{i=1}^N z_i \geq \sum_{i=2}^N z_i$ for any positive z. We then have:

$$\mathbb{E}_{y_{2:N}\sim Q_Y(.)}[\zeta_1] \tag{371}$$

$$\leq \mathbb{E}_{y_{2:N}\sim Q_Y(.)}\left[\frac{\hat{Z}(P_Y^A, y_{2:N})}{N} + \frac{\omega}{N}\cdot\frac{\hat{Z}(P_Y^A, y_{2:N})}{\hat{Z}(P_Y^B, y_{2:N})} + \frac{\omega}{N} + \frac{\omega^2}{N\hat{Z}(P_Y^B, y_{2:N})}\right] \tag{372}$$

$$= \frac{N-1}{N} + \frac{\omega}{N} + \frac{\omega}{N}\mathbb{E}_{y_{2:N}\sim Q_Y(.)}\left[\frac{\hat{Z}(P_Y^A, y_{2:N})}{\hat{Z}(P_Y^B, y_{2:N})}\right] + \frac{\omega^2}{N}\mathbb{E}_{y_{2:N}\sim Q_Y(.)}\left[\frac{1}{\hat{Z}(P_Y^B, y_{2:N})}\right] \tag{373}$$

$$\leq 1 + \frac{1}{N}\left(\omega + \omega\mathbb{E}_{y_{2:N}\sim Q_Y(.)}\left[\frac{\hat{Z}(P_Y^A, y_{2:N})}{\hat{Z}(P_Y^B, y_{2:N})}\right] + \omega^2\mathbb{E}_{y_{2:N}\sim Q_Y(.)}\left[\frac{1}{\hat{Z}(P_Y^B, y_{2:N})}\right]\right) \tag{374}$$

$$\leq 1 + \frac{1}{N}\left[\omega + \omega\mathbb{I}_N(\omega, 2)d_2(Q_Y||P_Y^B) + \frac{\omega^2}{N-1}d_2(Q_Y||P_Y^B)\right] \tag{375}$$

$$= 1 + \mu_1(N), \tag{376}$$

where the last inequality is due to Proposition F.1 and Corollary F.3.. For the other term, we have:

$$\zeta_2 = \left(\frac{\hat{Z}(P_Y^B, y_{1:N})}{\hat{Z}(P_Y^A, y_{1:N})}\right)\left(\frac{\hat{Z}(P_Y^A, y_{2:N})+\omega}{N}\right)\left(\frac{\hat{Z}(P_Y^B, y_{2:N})+\omega}{\hat{Z}(P_Y^B, y_{1:N})}\right) \tag{377}$$

$$= \frac{1}{N}\left(\frac{\hat{Z}(P_Y^A, y_{2:N})}{\hat{Z}(P_Y^A, y_{1:N})} + \frac{\omega}{\hat{Z}(P_Y^A, y_{1:N})}\right)\left(\hat{Z}(P_Y^B, y_{2:N}) + \omega\right) \tag{378}$$

$$\leq \frac{1}{N}\left(1 + \frac{\omega}{\hat{Z}(P_Y^A, y_{2:N})}\right)\left(\hat{Z}(P_Y^B, y_{2:N}) + \omega\right) \tag{379}$$

$$= \frac{\hat{Z}(P_Y^B, y_{2:N})}{N} + \frac{1}{N}\left(\omega + \frac{\omega\hat{Z}(P_Y^B, y_{2:N})}{\hat{Z}(P_Y^A, y_{2:N})} + \frac{\omega^2}{\hat{Z}(P_Y^A, y_{2:N})}\right). \tag{380}$$

1177 where we again repeatly use the inequality $\sum_{i=1}^N z_i \geq \sum_{i=2}^N z_i$ for any positive z. This gives us:

$$\mathbb{E}_{y_{2:N}\sim Q_Y(.)}[\zeta_2] \tag{381}$$

$$\leq \frac{1}{N}\left(\omega + \omega\mathbb{E}_{y_{2:N}\sim Q_Y()}\left[\frac{\hat{Z}(P_Y^B, y_{2:N})}{\hat{Z}(P_Y^A, y_{2:N})}\right] + \omega^2\mathbb{E}_{y_{2:N}\sim Q_Y()}\left[\frac{1}{\hat{Z}(P_Y^A, y_{2:N})}\right]\right) \tag{382}$$

$$\leq \frac{1}{N}\left[\omega + \omega\mathbb{I}_N(\omega,2)d_2(Q_Y||P_Y^A) + \frac{\omega^2}{N-1}d_2(Q_Y||P_Y^A)\right] \tag{383}$$

$$= \mu_2(N), \tag{384}$$

1178 where the last inequality is due to Proposition F.1 and Corollary F.3. This completes the proof.

1179 F.5 Coefficients in Proposition F.5

1180 We define the coefficient $\mu_1^{\mathrm{cond}}(N)$ and $\mu_2^{\mathrm{cond}}(N)$ in Proposition F.5.

$$\mu_1^{\mathrm{cond}}(N) = \frac{1}{N}\left[\omega + \omega\mathbb{I}_N(\omega,2)d_2(Q_Y||\tilde{P}_{Y|Z}(.|z)) + \frac{\omega^2}{N-1}d_2(Q_Y||\tilde{P}_{Y|Z}(.|z))\right] \tag{385}$$

$$\mu_2^{\mathrm{cond}}(N) = \frac{1}{N}\left[\omega + \omega\mathbb{I}_N(\omega,2)d_2(Q_Y||P_{Y|X}(.|x)) + \frac{\omega^2}{N-1}d_2(Q_Y||P_{Y|X}(.|x))\right] \tag{386}$$

1181 where we define $\mathbb{I}_N(\omega, i) \triangleq (2\mathbb{1}_{N>i} + \omega\mathbb{1}_{N=i})$ as in Proposition F.3.

1182 F.6 Proof of Proposition F.5

1183 We will use the global index for the proposals $Y_1, ...Y_N \sim Q(.)$ instead of $Y_{1,1}, ...Y_{1,N}$ unless
1184 otherwise stated. For the communication version, we have:

$$\Pr(Y_{K_A} = Y_{K_B}|Y_{K_A} = y_1, X = x, Z = z) \tag{387}$$

$$\geq \Pr(K_A{=}K_B|Y_{K_A}{=}y_1, X = x, Z = z) \tag{388}$$

$$= \sum_{k=1}^{\infty} \Pr(K_A = K_B = k|Y_{K_A} = y_1, X = x, Z = z) \tag{389}$$

$$\geq \sum_{k=1}^{N} \Pr(K_A = K_B = k|Y_{K_A} = y_1, X = x, Z = z) \tag{390}$$

$$= N\Pr(K_A = K_B = 1|Y_{K_A} = y_1, X = x, Z = z) \tag{391}$$

$$= N\Pr(K_{1,A} = K_{1,B} = 1, K_{2,A} = K_{2_B} = 1|Y_{K_A} = y_1, X = x, Z = z) \tag{392}$$

1185 Define:

$$\hat{Z}(P_{Y|X=x}, y_{1:N}) = \sum_{i=1}^N \frac{P_{Y|X}(y_i|x)}{Q_Y(y_i)}, \quad \hat{Z}(\tilde{P}_{Y|Z=z}, y_{1:N}) = \sum_{i=1}^N \frac{\tilde{P}_{Y|Z}(y_i|z)}{Q_Y(y_i)} \tag{393}$$

Now consider the following terms:

$$
\begin{aligned}
E_1 &= \Pr(K_{1,A}=1, K_{2,A}=1 | Y_{K_A}=y_1, Y_{2:N}=y_{2:N}, X=x, Z=z) \\
&\quad \times P(Y_{2:N}=y_{2:N}|Y_{K_A}=y_1, X=x, Z=z) \qquad (394)\\
&= \frac{1}{P_{X,Y,Z}(x,y_1,z)} Q_Y(y_{1:N}) P_X(x) \Pr(K_{2,A}=1|Y_{1:N}=y_{1:N}, X=x) \\
&\quad \times \Pr(K_{1,A}=1|Y_{1:N}=y_{1:N}, X=x, K_{2,A}=1) P_Z(z|Y_{1:N}=y_{1:N}, X=x, K_A=1) \qquad (395)\\
&= \frac{1}{P_{X,Y,Z}(x,y_1,z)} Q_Y(y_{1:N}) P_X(x) \Pr(K_{2,A}=1|Y_{1:N}=y_{1:N}, X=x) \\
&\quad \times \Pr(K_{1,A}=1|Y_{1:N}=y_{1:N}, X=x, K_{2,A}=1) P_{Z|X,Y}(z|X=x, Y=y_1) \qquad (396)\\
&= \frac{Q_Y(y_{1:N})}{P_{Y|X}(y_1|x)} \Pr(K_{2,A}=1|Y_{1:N}=y_{1:N}, X=x) \\
&\quad \times \Pr(K_{1,A}=1|Y_{1:N}=y_{1:N}, X=x, K_{2,A}=1) \qquad (397)\\
&= \frac{Q_Y(y_{1:N})}{P_{Y|X}(y_1|x)} \frac{P_{Y|X}(y_1|x)/Q_Y(y_1)}{\hat{Z}(P_{Y|X=x}, y_{2:N}) + \omega_x} \qquad (398)\\
&= \frac{Q_Y(y_{2:N})}{\hat{Z}(P_{Y|X=x}, y_{2:N}) + \omega_x} \qquad (399)
\end{aligned}
$$

and:

$$
\begin{aligned}
&E_2 \qquad (400)\\
&= \Pr(K_{2,B}=1|K_A=1, Y_{1:N}=y_{1:N}, X=x, Z=z) \qquad (401)\\
&= 1 - \Pr(K_{2,B}\neq 1|K_A=1, Y_{1:N}=y_{1:N}, X=x, Z=z) \qquad (402)\\
&= 1 - \Pr\left(\min_{j\neq 1} \frac{S_j}{\frac{\tilde{P}_{Y|Z}(y_j|z)}{\hat{Z}(\tilde{P}_{Y|Z=z}, y_{1:N})}} \leq \frac{S_1}{\frac{\tilde{P}_{Y|Z}(y_1|z)}{\hat{Z}(\tilde{P}_{Y|Z=z}, y_{1:N})}} \middle| K_A=1, Y_{1:N}=y_{1:N}, X=x, Z=z \right) \qquad (403)\\
&= 1 - \Pr\left(\min_{j\neq 1} \frac{S_j}{\frac{\tilde{P}_{Y|Z}(y_j|z)}{\hat{Z}(\tilde{P}_{Y|Z=z}, y_{1:N})}} \leq \frac{S_1}{\frac{\tilde{P}_{Y|Z}(y_1|z)}{\hat{Z}(\tilde{P}_{Y|Z=z}, y_{1:N})}} \middle| K_A=1, Y_{1:N}=y_{1:N}, X=x \right) \qquad (404)\\
&= 1 - \Pr\left(\min_{j\neq 1} \frac{S_j}{\frac{\tilde{P}_{Y|Z}(y_j|z)}{\hat{Z}(\tilde{P}_{Y|Z=z}, y_{1:N})}} \leq \frac{S_1}{\frac{\tilde{P}_{Y|Z}(y_1|z)}{\hat{Z}(\tilde{P}_{Y|Z=z}, y_{1:N})}} \middle| K_{2,A}=1, Y_{1:N}=y_{1:N}, X=x \right) \qquad (405)\\
&\geq \left(1 + \frac{P_{Y|X}(y_1|x)}{\tilde{P}_{Y|Z}(y_1|z)} \frac{\hat{Z}(\tilde{P}_{Y|Z=z}, y_{1:N})}{\hat{Z}(P_{Y|X=x}, y_{1:N})} \right)^{-1}, \qquad (406)
\end{aligned}
$$

where (404) is due to the Markov condtion $Z-(X,Y)-W$, (405) is due to the fact that the uniform random variable U is independent of S_1^N and (406) is due to the conditional importance matching lemma [32]. We note the following events are equivalent:

$$
\begin{aligned}
&\{K_A=1, Y_{1:N}=y_{1:N}, X=x, Z=z, K_{2,B}=1\} \qquad (407)\\
&\triangleq \left\{ U \leq \frac{\hat{Z}(P_{Y|X=x}, y_{1:N})}{\hat{Z}(P_{Y|X=x}, y_{2:N}) + \omega_x}, Y_{1:N}=y_{1:N}, X=x, Y_{K_A}=y_1, Z=z \right\} \qquad (408)\\
&\triangleq \mathcal{E} \cap \{Z=z\} \qquad (409)
\end{aligned}
$$

where $\mathcal{E} = \left\{ U \leq \frac{\hat{Z}(P_{Y|X=x}, y_{1:N})}{\hat{Z}(P_{Y|X=x}, y_{2:N}) + \omega_x}, Y_{1:N} = y_{1:N}, X = x, Y_{K_A} = y_1 \right\}$. Then, we have:

$$E_3 = \Pr(K_{1,B} = 1 | K_A = 1, Y_{1:N} = y_{1:N}, X = x, Z = z, K_{2,B} = 1) \tag{410}$$

$$= \Pr\left(U \leq \frac{\hat{Z}(\tilde{P}_{Y|Z=z}, Y_{1:N})}{\hat{Z}(\tilde{P}_{Y|Z=z}, Y_{2:N}) + \omega_z} \middle| \mathcal{E}, Z = z \right) \tag{411}$$

$$= \Pr\left(U \leq \frac{\hat{Z}(\tilde{P}_{Y|Z=z}, Y_{1:N})}{\hat{Z}(\tilde{P}_{Y|Z=z}, Y_{2:N}) + \omega_z} \middle| \mathcal{E} \right) \tag{412}$$

$$= \min\left(1, \frac{\hat{Z}(\tilde{P}_{Y|Z=z}, Y_{1:N})}{\hat{Z}(\tilde{P}_{Y|Z=z}, Y_{2:N}) + \omega_z} \cdot \frac{\hat{Z}(P_{Y|X=x}, y_{2:N}) + \omega_x}{\hat{Z}(P_{Y|X=x}, y_{1:N})} \right), \tag{413}$$

where the second to last equality is due to the Markov condition $Z - (X, Y) - W$.

Combining all three terms E_1, E_2, E_3 and continue from step (392), we have:

$$\Pr(Y_{K_A} = Y_{K_B} | Y_{K_A} = y_1, X = x, Z = z) \tag{414}$$

$$\geq N \int \frac{Q_Y(y_{2:N})}{\hat{Z}(P_{Y|X=x}, y_{2:N}) + \omega_x} \left(1 + \frac{P_{Y|X}(y_1|x)}{\tilde{P}_{Y|Z}(y_1|z)} \frac{\hat{Z}(\tilde{P}_{Y|Z=z}, y_{1:N})}{\hat{Z}(P_{Y|X=x}, y_{1:N})} \right)^{-1}$$

$$\times \min\left(1, \frac{\hat{Z}(\tilde{P}_{Y|Z=z}, Y_{1:N})}{\hat{Z}(\tilde{P}_{Y|Z=z}, Y_{2:N}) + \omega_z} \cdot \frac{\hat{Z}(P_{Y|X=x}, y_{2:N}) + \omega_x}{\hat{Z}(P_{Y|X=x}, y_{1:N})} \right) dy_{2:N} \tag{415}$$

$$= N \int \frac{Q_Y(y_{2:N})}{\hat{Z}(P_{Y|X=x}, y_{1:N})} \left(1 + \frac{P_{Y|X}(y_1|x)}{\tilde{P}_{Y|Z}(y_1|z)} \frac{\hat{Z}(\tilde{P}_{Y|Z=z}, y_{1:N})}{\hat{Z}(P_{Y|X=x}, y_{1:N})} \right)^{-1}$$

$$\times \min\left(\frac{\hat{Z}(P_{Y|X=x}, y_{1:N})}{\hat{Z}(P_{Y|X=x}, y_{2:N}) + \omega_x}, \frac{\hat{Z}(\tilde{P}_{Y|Z=z}, Y_{1:N})}{\hat{Z}(\tilde{P}_{Y|Z=z}, Y_{2:N}) + \omega_z} \right) dy_{2:N} \tag{416}$$

$$\geq \int \frac{N Q_Y(y_{2:N})}{\hat{Z}(P_{Y|X=x}, y_{1:N})} \left(1 + \frac{P_{Y|X}(y_1|x)}{\tilde{P}_{Y|Z}(y_1|z)} \frac{\hat{Z}(\tilde{P}_{Y|Z=z}, y_{1:N})}{\hat{Z}(P_{Y|X=x}, y_{1:N})} \right)^{-1}$$

$$\times \left(\frac{\hat{Z}(P_{Y|X=x}, y_{1:N})}{\hat{Z}(P_{Y|X=x}, y_{2:N}) + \omega} \cdot \frac{\hat{Z}(\tilde{P}_{Y|Z=z}, Y_{1:N})}{\hat{Z}(\tilde{P}_{Y|Z=z}, Y_{2:N}) + \omega} \right) dy_{2:N} \tag{417}$$

with the last inequality follows the fact that $\omega > \max(\omega_x, \omega_z)$. The rest of the proof follows similar steps as in the proof of Proposition F.4. This completes the proof.

G ERS Matching with Batch Communication

Setup. We first describe the setup in the case where the selected batch index is communicated from the encoder to the decoder. The main difference between this and the setup in Section 5.3 is that the decoder (party B) will use the Gumbel-Max selection method instead of the ERS one, since it knows which batch the encoder index belongs to. Furthermore, we note this scheme requires a noiseless channel between the encoder and decoder, which is available in the distributed compression scenario. Similarly to Section 3.2.2, let $(X, Y, Z) \in \mathcal{X} \times \mathcal{Y} \times \mathcal{Z}$ with a joint distribution $P_{X,Y,Z}$. We use the same common randomness W as in Section 5.3, with the proposal distribution Q_Y requiring that the bounding condition hold for the tuple $(P_{Y|X=x}, Q_Y)$. The protocol is as follows:

1. The encoder receives the input $X = x \sim P_X$ and selects its value using ERS procedure:
$$K_A = \mathrm{ERS}(W; P_{Y|X=x}, Q_Y), \tag{418}$$
and sends the batch index $K_{1,A}$ to the decoder. It then sets $Y_A = Y_{K_A}$
2. Given $X = x, Y_A = y$, we generate $Z = z \sim P_{Z|X,Y}(.|x, y)$ and note that the Markov chain $Z - (X, Y_A) - W$ holds.
3. The decoder receives the batch index $K_{1,A}$ and $Z = z$ will use the Gumbel Max process to queries a sample from the common randomness W:
$$K_{1,B} = K_{1,A} \quad K_{2,B} = \mathrm{Gumbel}(B_{K_{1,A}}; \tilde{P}_{Y|Z=z}, Q_Y) \quad K_B = (K_{1,B} - 1)N + K_{2,B},$$
and output $Y_B = Y_{K_B}$. The procedure $\mathrm{Gumbel}(.)$ corresponds to Step 1,2 in Algoirhm 1.

Given the above setup, we have the following bound on the matching event $\{Y_A = Y_B\}$:

Proposition G.1. *Let $K_A, K_B, P_{Y|X}(.|X = x)$ and $\tilde{P}_{Y|Z}(.|z)$ defined above and set $P_Y^A = P_{Y|X=x}, P_Y^B = \tilde{P}_{Y|Z=z}$. For $N \geq 2$, we have:*
$$\Pr(Y_A = Y_B|Y_A = y, X = x, Z = z) \geq \left(1+\mu_1'(N)+\frac{P_Y^A(y)}{P_Y^B(y)}\left(1 + \mu_2'(N)\right)\right)^{-1}, \tag{419}$$
where $\mu_1'(N)$ and $\mu_2'(N)$ are defined as in Appendix G.1 and we note that $\mu_1'(N), \mu_2'(N) \to 0$ as $N \to \infty$ with rate N^{-1}under mild assumptions on the distributions $P_{Y|X}(y|x)$ and $Q_Y(.)$.

Proof. See Appendix G.2. □

G.1 Coefficients in Proposition G.1

We define the coefficient $\mu_1'(N)$ and $\mu_2'(N)$ in Proposition G.1.
$$\mu_1'(N) = \frac{3\omega}{N} \tag{420}$$
$$\mu_2'(N) = \frac{\omega}{N}\mathbb{I}_N(\omega, 2)d_2(Q_Y||P_{Y|X=x}) \tag{421}$$
where we define $\mathbb{I}_N(\omega, i) \triangleq (2\mathbb{1}_{N>i} + \omega\mathbb{1}_{N=i})$ as in Proposition F.3 and $\omega = \max_y \frac{P_{Y|X(y|x)}}{Q_Y(y)}$.

G.2 Proof of Proposition G.1

We now formally prove the bound Proposition G.1. First, we define:
$$\hat{Z}(P_{Y|X=x}, y_{1:N}) = \sum_{i=1}^{N} \frac{P_{Y|X}(y_i|x)}{Q_Y(y_i)}, \quad \hat{Z}(\tilde{P}_{Y|Z=z}, y_{1:N}) = \sum_{i=1}^{N} \frac{\tilde{P}_{Y|Z}(y_i|z)}{Q_Y(y_i)} \tag{422}$$

Recall that $K_{2,A}$ is the local index within the selected batch by party A and $Y_{K_{1,A},1:N}$ are the samples within the selected batch, we have:

$$
\begin{aligned}
&\Pr(Y_A = Y_B | Y_A = y_1, X = x, Z = z) && (423)\\
&= \Pr(Y_{K_A} = Y_{K_B} | Y_{K_A} = y_1, X = x, Z = z) && (424)\\
&\geq \Pr(K_{2,A} = K_{2,B} | Y_{K_A} = y_1, X = x, Z = z) && (425)\\
&= \sum_{i=1}^{N} \Pr(K_{2,A} = K_{2,B} = i | Y_{K_A} = y_1, X = x, Z = z) && (426)\\
&= N \Pr(K_{2,A} = K_{2,B} = 1 | Y_{K_A} = y_1, X = x, Z = z) \quad \text{(Due to Symmetry)} && (427)\\
&= N \Pr(K_{2,A} = K_{2,B} | Y_{K_A} = y_1, K_{2,A} = 1, X = x, Z = z)\\
&\qquad\qquad \times \Pr(K_{2,A} = 1 | Y_{K_A} = y_1, X = x, Z = z) && (428)\\
&= \Pr(K_{2,A} = K_{2,B} | Y_{K_A} = y_1, K_{2,A} = 1, X = x, Z = z) && (429)\\
&= \int_{-\infty}^{\infty} P(Y_{K_{1,A},2:N} = y_{2:N} | Y_{K_A} = y_1, K_{2,A} = 1, X = x, Z = z)\\
&\quad \times \Pr(K_{2,A} = K_{2,B} | Y_{K_A} = y_1, K_{2,A} = 1, Y_{K_{1,A},2:N} = y_{2:N}, X = x, Z = z) dy_{2:N}, && (430)
\end{aligned}
$$

where (429) is due to $\Pr(K_{2,A} = 1 | Y_{K_A} = y_1, X = x, Z = z) = N^{-1}$. Let $Y_{1:N} \sim Q$ are N i.i.d. proposal samples, then $\{Y_{K_A,1:N} = y_{1:N}\} = \{Y_{1:N} = y_{1:N}, A \text{ accepts } Y_{1:N}\}$ and we have:

$$
\begin{aligned}
&\Pr(K_{2,A} = K_{2,B} | Y_{K_A} = y_1, K_{2,A} = 1, Y_{K_{1,A},2:N} = y_{2:N}, X = x, Z = z) && (431)\\
&= 1 - \Pr(K_{2,B} \neq 1 | Y_{K_{1,A},1:N} = y_{1:N}, K_{2,A} = 1, Y_{K_A} = y_1, X = x, Z = z)\\
&= 1 - \Pr(\min_{j \neq 1} \frac{S_j}{\frac{\tilde{P}_{Y|Z}(y_j|z)}{\hat{Z}(\tilde{P}_{Y|Z=z}, y_{1:N})}} \leq \frac{S_1}{\frac{\tilde{P}_{Y|Z}(y_1|z)}{\hat{Z}(\tilde{P}_{Y|Z=z}, y_{1:N})}} | Y_{1:N} = y_{1:N}, A \text{ selects 1st index},\\
&\qquad\qquad A \text{ accepts } Y_{1:N}, Y_{K_A} = y_1, X = x, Z = z) && (432)\\
&= 1 - \Pr(\min_{j \neq 1} \frac{S_j}{\frac{\tilde{P}_{Y|Z}(y_j|z)}{\hat{Z}(\tilde{P}_{Y|Z=z}, y_{1:N})}} \leq \frac{S_1}{\frac{\tilde{P}_{Y|Z}(y_1|z)}{\hat{Z}(\tilde{P}_{Y|Z=z}, y_{1:N})}} | Y_{1:N} = y_{1:N}, A \text{ selects 1st index},\\
&\qquad\qquad A \text{ accepts } Y_{1:N}, Y_{K_A} = y_1, X = x) && (433)\\
&= 1 - \Pr\left(\min_{j \neq 1} \frac{S_j}{\frac{\tilde{P}_{Y|Z}(y_j|z)}{\hat{Z}(\tilde{P}_{Y|Z=z}, y_{1:N})}} \leq \frac{S_1}{\frac{\tilde{P}_{Y|Z}(y_1|z)}{\hat{Z}(\tilde{P}_{Y|Z=z}, y_{1:N})}} | Y_{1:N} = y_{1:N}, A \text{ selects 1st index}, X = x\right) && (434)\\
&\geq \left(1 + \frac{P_{Y|X}(y_1|x)}{\tilde{P}_{Y|Z}(y_1|z)} \frac{\hat{Z}(\tilde{P}_{Y|Z=z}, y_{1:N})}{\hat{Z}(P_{Y|X=x}, y_{1:N})}\right), && (435)
\end{aligned}
$$

where (433) is due to Markov property $Z - (X, Y) - W$,i.e. Z has no effects on the statistics of the exponential random variables. Line (434) is due to the fact that conditioning on A selected the 1st index, whether A selects $Y_{1:N}$ or not depends only on U. The final inequality is due to conditional matching lemma from [32].

Recall that $\omega = \max_y \frac{P_{Y|X(y|x)}}{Q_Y(y)}$, we have:

$$
\begin{aligned}
&P(Y_{K_{1,A},2:N} = y_{2:N} | Y_{K_A} = y_1, K_{2,A} = 1, X = x, Z = z) && (436)\\
&= P(Y_{K_{1,A},2:N} = y_{2:N} | Y_{K_A} = y_1, K_{2,A} = 1, X = x) && (437)\\
&= \frac{\bar{P}_{Y,K_{2,A}|X}(y_{1:N}, 1|x)}{P_{Y|X}(y_1|x) N^{-1}} && (438)\\
&= \frac{N Q_Y(y_{2:N})}{\Delta_{P_{Y|X=x}}(\hat{Z}(P_{Y|X=x}, y_{2:N}) + \omega)} && (439)
\end{aligned}
$$

where $\bar{P}_{Y,K_{2,A}|X}(y_{1:N}, 1|x)$ is the ERS target distribution (151) where we use $P_{Y|X}(.|x)$ as the target distribution and $\Delta_{P_{Y|X=x}} < 1$ is the normalized constant. We now shorthand $P^A_Y \triangleq P_{Y|X=x}$,

$P_Y^B \triangleq \tilde{P}(Y|Z=z)$ and $\Delta_{P_Y^A} \triangleq \Delta_{P_{Y|X=x}}$, and combining the two expressions, we have:

$$
\begin{aligned}
&\Pr(Y_A = Y_B | Y_A = y_1, X = x, Z = z) &(440)\\
&\geq \mathbb{E}_{Y_{2:N}\sim Q_Y}\left[\frac{N}{(\hat{Z}(P_Y^A, y_{2:N}) + \omega)\Delta_{P_Y^A}\left(1 + \frac{P_Y^A(y_1)}{P_Y^B(y1)}\frac{\hat{Z}(P_Y^B, y_{1:N})}{\hat{Z}(P_Y^A, y_{1:N})}\right)}\right] &(441)\\
&\geq \mathbb{E}_{Y_{2:N}\sim Q_Y}\left[\frac{N}{(\hat{Z}(P_Y^A, y_{2:N}) + \omega)\left(1 + \frac{P_Y^A(y_1)}{P_Y^B(y1)}\frac{\hat{Z}(P_Y^B, y_{1:N})}{\hat{Z}(P_Y^A, y_{1:N})}\right)}\right] \quad \text{(Since } \Delta_{P_Y^A} \leq 1\text{)} &(442)\\
&\geq \left(\mathbb{E}_{Y_{2:N}\sim Q_Y}\left[\frac{(\hat{Z}(P_Y^A, y_{2:N}) + \omega)}{N}\left(1 + \frac{P_Y^A(y_1)}{P_Y^B(y1)}\frac{\hat{Z}(P_Y^B, y_{1:N})}{\hat{Z}(P_Y^A, y_{1:N})}\right)\right]\right)^{-1} \text{(By Jensen's Inequality)} &(443)
\end{aligned}
$$

Since:

$$
\mathbb{E}_{Y_{2:N}\sim Q_Y}\left[\frac{\hat{Z}(P_Y^A, y_{2:N}) + \omega}{N}\right] \leq \frac{N-1}{N} + \frac{\omega}{N} \tag{444}
$$

and:

$$
\begin{aligned}
&\mathbb{E}_{Y_{2:N}\sim Q_Y}\left[\left(\frac{\hat{Z}(P_Y^A, y_{2:N}) + \omega}{N}\right)\frac{\hat{Z}(P_Y^B, y_{1:N})}{\hat{Z}(P_Y^A, y_{1:N})}\right] &(445)\\
&= \mathbb{E}_{Y_{2:N}\sim Q_Y}\left[\frac{\hat{Z}(P_Y^A, y_{2:N})}{N}\frac{\hat{Z}(P_Y^B, y_{1:N})}{\hat{Z}(P_Y^A, y_{1:N})} + \frac{\omega}{N}\frac{\hat{Z}(P_Y^B, y_{1:N})}{\hat{Z}(P_Y^A, y_{1:N})}\right] &(446)\\
&\leq \frac{N-1}{N} + \frac{P_Y^B(y_1)/Q_Y(y_1)}{N} + \frac{\omega}{N}\mathbb{E}_{Y_{2:N}\sim Q_Y}\left[\frac{\hat{Z}(P_Y^B, y_{1:N})}{\hat{Z}(P_Y^A, y_{1:N})}\right] &(447)
\end{aligned}
$$

where we have:

$$
\begin{aligned}
\mathbb{E}_{Y_{2:N}\sim Q_Y}\left[\frac{\hat{Z}(P_Y^B, y_{1:N})}{\hat{Z}(P_Y^A, y_{1:N})}\right] &= \mathbb{E}_{Y_{2:N}\sim Q_Y}\left[\frac{P_Y^B(y_1)/Q_Y(y_1)}{\hat{Z}(P_Y^A, y_{1:N})} + \frac{\hat{Z}(P_Y^B, y_{2:N})}{\hat{Z}(P_Y^A, y_{1:N})}\right] &(448)\\
&\leq \mathbb{E}_{Y_{2:N}\sim Q_Y}\left[\frac{P_Y^B(y_1)/Q_Y(y_1)}{P_Y^A(y_1)/Q_Y(y_1)}\right] + \mathbb{E}_{Y_{2:N}\sim Q_Y}\left[\frac{\hat{Z}(P_Y^B, y_{2:N})}{\hat{Z}(P_Y^A, y_{2:N})}\right] &(449)\\
&\leq \frac{P_Y^B(y_1)}{P_Y^A(y_1)} + \mathbb{I}_N(\omega, 2)d_2(Q_Y||P_Y^A) &(450)
\end{aligned}
$$

Then, combining (450) into (447), then combine with (444) into the term (443), we have:

$$
\begin{aligned}
&\Pr(Y_A = Y_B | Y_A = y_1, X = x, Z = z) &(451)\\
&\geq \left(1 + \frac{\omega}{N} + \frac{P_Y^A(y_1)}{P_Y^B(y_1)}\left(\frac{N-1}{N} + \frac{P_Y^B(y_1)/Q_Y(y_1)}{N} + \frac{\omega}{N}\left(\frac{P_Y^B(y_1)}{P_Y^A(y_1)} + \mathbb{I}_N(\omega, 2)d_2(Q_Y||P_Y^A)\right)\right)\right)^{-1} &(452)\\
&= \left(1 + \frac{\omega}{N} + \frac{P_Y^A(y_1)}{P_Y^B(y_1)}\left(\frac{N-1}{N} + \frac{P_Y^B(y_1)/Q_Y(y_1)}{N} + \frac{\omega}{N}\left(\frac{P_Y^B(y_1)}{P_Y^A(y_1)} + \mathbb{I}_N(\omega, 2)d_2(Q_Y||P_Y^A)\right)\right)\right)^{-1} &(453)\\
&\geq \left(1 + \frac{3\omega}{N} + \frac{P_Y^A(y_1)}{P_Y^B(y_1)}\left(1 + \frac{\omega}{N}\mathbb{I}_N(\omega, 2)d_2(Q_Y||P_Y^A)\right)\right)^{-1} &(454)\\
&= \left(1 + \mu_1'(N) + \frac{P_Y^A(y_1)}{P_Y^B(y_1)}(1 + \mu_2'(N))\right)^{-1}, &(455)
\end{aligned}
$$

where we repeatly use the fact that $P_Y^A(y)/Q_Y(y) \leq \omega$. This completes the proof.

H Proof of Proposition 5.6

Algorithm 2: Wyner-Ziv Distributed Compression Protocol

Encoder: Receives $X = x$ and W, performs:
1. Select $K_A = \mathrm{ERS}(W; P_{Y'|X=x}, Q_{Y'})$; 2. Sends $(K_{1,A}, V_{K_A})$ to the decoder.

Decoder: Receives $Z = (V_{K_A}, K_{1,A}, X')$ and W, performs:
1. Keep batch $K_{1,A}$; 2. Remove all j where $V_{K_{1,A},j} \neq V_{K_A}$; 3. Select K_B with $P_{Y'|X'=x'}$.

Main Proof. We remind the protocol in Algorithm 2. The encoder and decoder's target distribution for this case are:

$$P_Y^A(y, v) = P_{Y|X}(y|x)P_V(v) \quad P_Y^B(y, v) = P_{Y|X'}(y|x)\mathbb{I}_V(v) \tag{456}$$

For a sufficient large batch size N and apply Proposition G.1, we have:

$$\begin{aligned}
&\Pr(Y'_{K_A} \neq Y'_{K_B} | (Y'_{K_A}, V_{K_A}) = (y', v), X = x, Z = (x', v)) && (457)\\
&= \Pr((Y'_{K_A}, V_{K_A}) \neq (Y'_{K_B}, V_{K_B}) | (Y'_{K_A}, V_{K_A}) = (y', v), X = x, Z = (x', v)) && (458)\\
&\leq 1 - \left(1 + \epsilon + \frac{P_{Y'|X}(y'|x)P_V(v)}{P_{Y'|X'}(y'|x')\mathbb{I}_v(v)}(1+\epsilon)\right)^{-1} && (459)\\
&\leq 1 - \left(1 + \epsilon + \mathcal{V}^{-1}(1+\epsilon)\frac{P_{Y'|X}(y'|x)}{P_{Y'|X'}(y'|x')}\right)^{-1} && (460)\\
&= 1 - \left(1 + \epsilon + \mathcal{V}^{-1}(1+\epsilon)\frac{P_{Y'|X}(y'|x)}{P'_Y(y')}\frac{P'_Y(y')}{P_{Y'|X'}(y'|x')}\right)^{-1} && (461)\\
&= 1 - \left(1 + \epsilon + \mathcal{V}^{-1}(1+\epsilon)2^{i_{Y';X}(y';x) - i_{Y';X'}(y';x')}\right)^{-1} && (462)
\end{aligned}$$

Finally, taking the expectation of both sides yields the final result.

Coding Cost. In terms of the bound on r, recall the following bound on batch acceptance probability:

$$\Delta = \mathbb{E}_{Y_{1:N} \sim P_Y(.)}\left[\frac{N}{\bar{Z}(1, Y'_{1:N})}\right] \geq \frac{N}{\mathbb{E}_{Y'_{1:N} \sim P_Y(.)}[\bar{Z}(1)]} = \frac{N}{N - 1 + \omega} \tag{463}$$

Here for $N = \omega$, we have $\Delta > \frac{1}{2}$ and thus the chunk size $L = \lfloor \Delta^{-1} \rfloor$ in the ERS coding scheme is 1 and thus do not need to send $\hat{K}_1$. Using the fact that $\mathbb{E}[\log L] \leq 1$, we have $r \leq H[L] + 1 = 4$bits by entropy coding with Zipf distribution [25].

Compressing Multiple Samples. When compressing n samples jointly, let the rate per sample (without the overhead for batch communication) be r' where $\log(V) = nr'$ consider the following approximation:

$$\sum_{i=1}^{n} i(y'_i; x_i) - i(y'_i; x'_i) \approx nI(X; Y'|X'),$$

Then we have:

$$\begin{aligned}
2^{\sum_{i=1}^{n}[i(y'_i;x_i) - i(y'_i;x'_i)] - \log(V)} &\approx 2^{nI(X;Y'|X') - \log(V)} && (464)\\
&= 2^{n(I(X;Y'|X') - r')}, && (465)
\end{aligned}$$

and thus, if $r' > I(X; Y'|X')$, by increasing n we reduces the mismatching probability while maintaining the compression rate per sample. We visualize this in the experimental results with $N = 2^{19}$ in Figure 9.

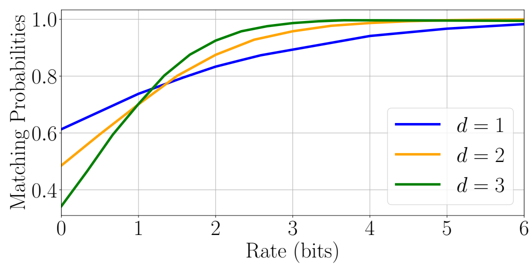

Figure 9: Matching Probabilities with $N = 2^{19}$ and jointly compressing $1, 2, 3$ i.i.d. samples respectively. Target distortion $\sigma^2_{Y'|X} = 0.008$ for every samples.

I Feedback Scheme

In distributed compression, decoding errors can lead to significant average reconstruction distortion. To address this, feedback communication from the decoder can be employed to correct errors and enhance rate-distortion performance, as proposed in [32]. The feedback mechanism is identical for both ERS and IML, except that ERS additionally transmits the batch index to the decoder.

We recall that $K_{1,A}$ and $K_{2,A}$ denote the batch index and local index, respectively, of samples selected by party A through the ERS sample selection. On the other hand, party B uses Gumbel-Max selection process to output its selected local index $K_{2,B}$ within the $K_{1,A}$ batch, then the ERS process can be described as follows:

1. *Index Selection.* After transmitting the batch index $K_{1,A}$, the encoder sends the $\log_2(\mathcal{V})$ least significant bits (LSB) of the selected index $K_{2,A}$ to the decoder.
2. *Decoding and Feedback.* The decoder outputs $K_{2,B}$ and sends the $\log_2(N/\mathcal{V})$ most significant bits (MSB) of K_B to the encoder.
3. *Re-transmission.* Based on the received MSB feedback, if the index is correct, the encoder responds with an acknowledgment bit, say 1. Otherwise, it sends 0 along with the MSB of its selection to the decoder.

We note that, in this context, using LSB instead of random bits in step 1 does not yield a noticeable difference in performance. For the rate-distortion analysis, the rate is computed based on the total length of messages transmitted during index selection and re-transmission, including any acknowledgment messages. However, the rate of the feedback message is excluded from this calculation, which can be justified in scenarios with asymmetric communication costs in the forward and reverse directions, such as in wireless channels.

J Neural Contrastive Estimator

In our ERS scheme, the selection rule requires estimating the following ratio at the decoder side:

$$\tilde{K}_B = \operatorname*{argmin}_{1 \le k \le N} \frac{S_{ik}}{\frac{P_{Y|X'}(y|x')\mathbb{I}_V(v)}{Q_Y(Y_{ik})\mathbb{V}^{-1}}}, \text{ where } i = K_{1,A}, \tag{466}$$

where the normalization term can be ignored as it is the same for every sample in the batch $K_{1,A}$. Our goal is to learn the ratio $P_{Y|X'}(Y_{ik}|x')/Q_Y(Y_{ik})$ from data. In particular, we can access the data samples from the joint distribution $P_{X,Y,X'}$.

To this end, we construct a binary neural classifier $h'(y, x') = \frac{1}{1+\exp[-h(y,x')]}$ which classifies if the input (y, x') is distributed according to the marginal distribution $Q_Y(.) \times P_{X'}(.)$ (positive samples) or the joint $P_{Y,X'}$ (negative samples). Once converged, we can use the logits value $h(y, x')$ to compute the log of the ratio of interest [19]. In particular:

$$h(y, x') \approx -\log \frac{P_{Y|X'}(Y_{ik}|x')}{Q_Y(y)} \tag{467}$$

This allows us to estimate the ratio without needing to obtain the exact ratio's value. Finally, to generate the positive samples, we simply generate $Y \sim Q_Y(.)$ and get a random X' from the training data. For negative samples, we generate the data according to the Markov sequence $X' - X - Y$. The ratio between the two labels should be the same.

K Distributed Compression with MNIST

K.0.1 Training Details

β-VAE Architecture. We adopt a setup similar to [32]. Our neural encoder-decoder model comprises an encoder network $y = \mathrm{enc}(x)$, a projection network $\mathrm{proj}(x')$, and a decoder network $\hat{x} = \mathrm{dec}(y, \mathrm{proj}(x'))$, as detailed in Table 1. The encoder network converts an image into two vectors of size 3 (total 6D output), with the first vector representing the output mean $\mu(x)$ and the second representing the output variance $\sigma^2(x)$. Here, we define $p_{Y|X}(.|x) = \mathcal{N}(\mu(x), \sigma^2(x))$ and use the prior distribution $p_Y(.) = \mathcal{N}(0, 1)$. At the decoder side (party B), the projection network first maps the side information image X' to a 128-dimensional vector, which is then combined with a 3-dimensional vector from the encoder. This concatenated vector is input to the decoder network, producing a reconstructed output of size 28×28.

Table 1: Architecture of the encoder, projection network, and decoder for distributed MNIST image compression. Convolutional and transposed convolutional layers are denoted as "conv" and "upconv," respectively, with specifications for the number of filters, kernel size, stride, and padding. For "upconv," an additional output padding parameter is included.

(a) Encoder
Input $28 \times 28 \times 1$
conv (128:3:1:1), ReLU
conv (128:3:2:1), ReLU
conv (128:3:2:1), ReLU
Flatten
Linear (6272, 512), ReLU
Linear (512, 6)

(b) Projection Network
Input $14 \times 14 \times 1$
conv (32:3:1:1), ReLU
conv (64:3:2:1), ReLU
conv (128:3:2:1), ReLU
Flatten
Linear (2048, 512), ReLU
Linear (512, 128)

(c) Decoder Network
Input-(3+128)
Linear-(132, 512), ReLU
upconv (64:3:2:1:1), ReLU
upconv (32:3:2:1:1), ReLU
upconv (1:3:1:1), Tanh

Loss Function We train our β-VAE network by optimizing the following rate-distortion loss function:

$$\mathcal{L} = \beta(X - \hat{X})^2 + E_X[D_{\mathrm{KL}}(p_{Y|X}(.|v)||p_Y(.))] \tag{468}$$

where we vary β for different rate-distortion tradeoff.Each model is trained for 30 epochs on an NVIDIA RTX A4500, requiring approximately 30 minutes per model. We use random horizontal flipping and random rotation within the range $\pm 15^{\mathrm{o}}$. We use the following values of $\beta \in \{0.225, 0.28, 0.31, 0.4\}$ that corresponds to the target distortions $\{6.6, 6.3, 6.1, 5.8\} \times 10^{-2}$ in Figure 6.

Neural Contrastive Estimator Network. The neural estimator network comprises two subnetworks. The first subnetwork projects the side information into a 128-dimensional embedding. The second subnetwork combines this 128D embedding with a 4D embedding, derived from either $p_{Y|X}$ or the prior p_Y, and outputs the probability that X', Y originate from the joint or marginal distributions. The model is trained for 100 epochs.

Table 2: Neural Estimator Networks for Distributed Image Compression.

(a) Projection Network
Input $14 \times 14 \times 1$
conv (32:3:1:1), ReLU
conv (64:3:2:1), ReLU
conv (128:3:2:1), ReLU
Flatten
Linear (2048, 512), ReLU
Linear (512, 128)

(b) Combine and Classify
Input $128 + 3$
Linear (132, 128), l-ReLU
Linear (128,128), l-ReLU
Linear (128,128), l-ReLU
Linear (128, 1)

$\log \mathcal{V}$	N	N^*	Target dB
9.6	0.6e6	1.0e6	−21.5dB
10.6	0.7e6	1.1e6	−22dB
11.6	0.8e6	1.5e6	−22.5dB
12.6	1.04	1.6e6	−23dB

Table 3: Details for ERS Gaussian Experiment in Figure 5 (right)

L Wyner-Ziv Gaussian Experiment

In Figure 5 (left), the batch size of ERS are $N \in \{2^{19}, 2^{20}\}$ respectively for the average number of proposals $N^* \in \{1.1, 1.6\} \times 10^6$. For Figure 5 (right), details for ERS are shown in Table 3.