Exercise 5.1 Min-entropy and Shannon entropy as Rényi entropies [EE5139]

Both the min-entropy and the Shannon entropy are limiting cases of the following family of Rényi entropies:

$$H_\alpha(X) = \frac{1}{1-\alpha} \log \sum_x P(x)^\alpha, \qquad \alpha \in (0,1) \cup (1, +\infty)\,. \tag{1}$$

a.) To verify this, compute the limit of the above quantities for $\alpha \to \{0_+, 1, +\infty\}$. (Here, by saying $\alpha \to 0_+$, we mean α "approaching 0 from right-hand side".)

Solution:

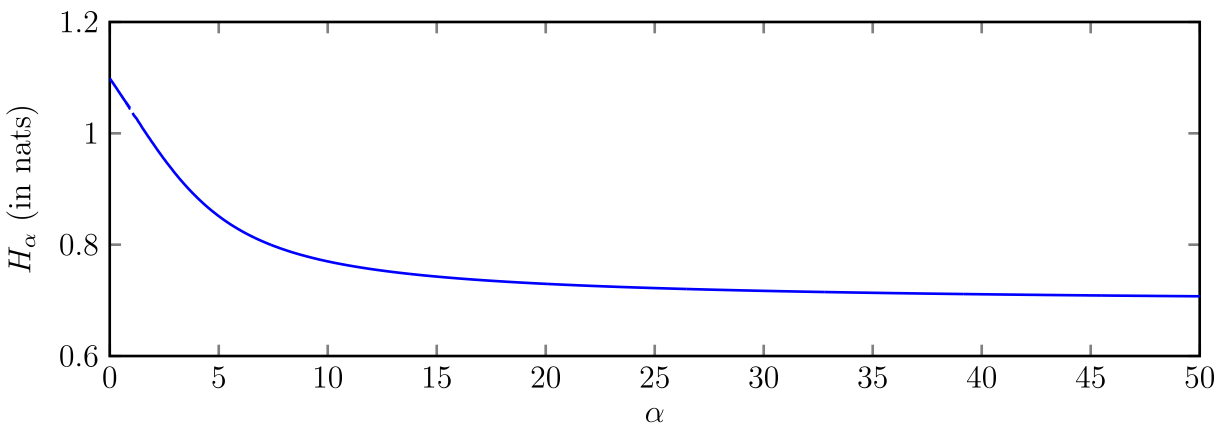
$$\begin{aligned}
\lim_{\alpha \to 0_+} H_\alpha(X) &= \log \left(\sum_x \lim_{\alpha \to 0_+} P(x)^\alpha \right) \\
&= \log |\{x : P(x) > 0\}| = H_{\max}(X). \\
\lim_{\alpha \to 1} H_\alpha(X) &= \frac{\left(\sum_x P(x)\right)^{-1} \cdot \sum_x \log P(x) \cdot P(x)^\alpha}{-1} && \blacktriangleright \text{ L'Hôpital's rule} \\
&= -\sum_x P(x) \cdot \log P(x) = H(X). \\
\lim_{\alpha \to +\infty} H_\alpha(X) &= -\lim_{\alpha \to +\infty} \log \left[\left(\sum_x P_{\max}^\alpha \cdot \left(\frac{P(x)}{P_{\max}} \right)^\alpha \right)^{\frac{1}{\alpha-1}} \right] && \blacktriangleright P_{\max} \triangleq \max_x P(x) \\
&= -\lim_{\alpha \to +\infty} \frac{\alpha}{\alpha - 1} \cdot \log P_{\max} + \frac{1}{\alpha-1} \cdot \log \left(\sum_x \left(\frac{P(x)}{P_{\max}} \right)^\alpha \right) \\
&= \log P_{\max} = H_{\min}(X). && \blacktriangleright 1 \le \sum_x \left(\frac{P(x)}{P_{\max}} \right)^\alpha \le |\mathcal{X}|
\end{aligned}$$

□

b.) Plot the Rényi entropy as a function of α for the random variable X distributed as

x	0	1	2
$P(x)$	1/2	1/4	1/4

Solution: We plot the Rényi entropy of the above random variable in nats.



$\square$

c.) Show that, for any random variable $X \in \mathcal{X}$ and any pmf $P(x)$, the Rényi entropy is monotonically non-increasing in the parameter α. Argue how this yields an alternative proof of the fact that $H_{\min}(X) \leq H(X) \leq \log|\mathcal{X}|$.

Solution: Consider the derivative of H_α w.r.t α

$$\frac{\mathrm{d}}{\mathrm{d}\alpha} H_\alpha(X) = \frac{1}{(1-\alpha)^2} \cdot \left\{ (1-\alpha) \cdot \frac{\sum_x \log P(x) \cdot P(x)^\alpha}{\sum_x P(x)^\alpha} + \log \sum_x P(x)^\alpha \right\}.$$

Note that $\lim_{\alpha \to 1} H_\alpha$ exists from both sides. It suffices to show $\frac{\mathrm{d}}{\mathrm{d}\alpha} H_\alpha(X)$ to be non-positive for all $\alpha \in (0,1) \cup (1,\infty)$. By letting

$$f(\alpha) \triangleq (1-\alpha) \cdot \frac{\sum_x \log P(x) \cdot P(x)^\alpha}{\sum_x P(x)^\alpha} + \log \sum_x P(x)^\alpha$$

it suffices to show f to be non-positive for $\alpha > 0$.

However, noticing that function $g : t \mapsto t \cdot \log t$ is convex for $t > 0$, we have (for each α)

$$\begin{aligned} f(\alpha) &= \frac{\left(\sum_x P(x) \cdot P(x)^{\alpha-1}\right) \log\left(\sum_x P(x) \cdot P(x)^{\alpha-1}\right) - \sum_x P(x) \cdot \left(P(x)^{\alpha-1} \log P(x)^{\alpha-1}\right)}{\sum_x P(x)^\alpha} \\ &= \frac{g(\sum_x P(x) \cdot t_x) - \sum_x P(x) \cdot g(t_x)}{\sum_x P(x)^\alpha} \leq 0, \end{aligned}$$

where $t_x \triangleq P(x)^{\alpha-1}$. Thus, we have finished the proof. $\square$

d.) Compute the min-entropy $H_{\min}(X|Y)$ of the joint random variables (X, Y) distributed as

$P(x,y)$		$X=0$	$X=1$	$X=2$
Y	0	1/6	1/12	1/12
	1	1/12	1/6	1/12
	2	1/12	1/12	1/6

Solution: $H_{\min}(X|Y) = 1$. $\square$

Exercise 5.2 Distributions with a large entropy gap [all]

It is possible to construct distributions that have a large gap between min-entropy and Shannon entropy. This shows that controlling the Shannon entropy or the mutual information is not sufficient for most cryptographic tasks.

a.) Given $\epsilon \in (0,1)$, construct a sequence of random variables $(X_2, X_3, \ldots, X_n, \ldots)$ where $X_n \in \{0, 1, \ldots, n-1\}$, such that

$$\left.\begin{aligned} H(X_n) &\geq (1-\epsilon)\log n \\ H_{\min}(X_n) &= C, \end{aligned}\right\} \forall n \geq N$$

for some $N \in \mathbb{N}$ and some constant $C > 0$.

Solution: For each $n = 2, 3, 4, \ldots$, we consider the following distribution

$$P_{X_n}(x) = \begin{cases} \epsilon & \text{if } x = 0, \\ \frac{1-\epsilon}{n-1} & \text{otherwise.} \end{cases}$$

In this case,

$$H(X_n) = H(\epsilon) + (1-\epsilon)\cdot\log(n-1) \geq (1-\epsilon)\cdot\log n \qquad \forall n \geq \frac{1-\epsilon}{H(\epsilon)} + 1,$$
$$H_{\min}(X_n) = -\log\epsilon \qquad \forall n \geq \frac{1-\epsilon}{\epsilon} + 1.$$

Thus the construction satisfies the requirements by letting $N = \left\lceil \max\{\frac{1-\epsilon}{H(\epsilon)} + 1, \frac{1-\epsilon}{\epsilon} + 1\}\right\rceil$, and $C = -\log\epsilon$. □

b.) Given $\epsilon \in (0,1)$, construct a sequence of random variables $((X_2, Y_2), (X_3, Y_3), \ldots, (X_n, Y_n), \ldots)$, where $X_n,\, Y_n \in \{0, 1, \ldots, n-1\}$, such that

$$H(X_n) = H_{\min}(X_n) = \log n \qquad \forall n$$
$$\left.\begin{aligned} H(X_n|Y_n) &\geq (1-\epsilon)\log n \\ H_{\min}(X_n|Y_n) &= C \end{aligned}\right\} \forall n \geq N$$

for some $N \in \mathbb{N}$ and some constant $C > 0$.

Solution: For each $n = 2, 3, 4, \ldots$, we consider the following distribution

$$P_{Y_n}(y) = \frac{1}{n}$$
$$P_{X_n|Y_n}(x|y) = \begin{cases} \epsilon & \text{if } x = y, \\ \frac{1-\epsilon}{n-1} & \text{otherwise.} \end{cases}$$

In this case, $P_{X_n}(x) = \sum_y P_{Y_n}(y)\cdot P_{X_n|Y_n}(x|y) = 1/n$. Thus, $H(X_n|Y_n) \geq (1-\epsilon)\log n$. Additionally,

$$H(X_n|Y_n) = H(\epsilon) + (1-\epsilon)\cdot\log(n-1) \geq (1-\epsilon)\cdot\log n \qquad \forall n \geq \frac{1-\epsilon}{H(\epsilon)} + 1,$$
$$H_{\min}(X_n|Y_n) = -\log\epsilon \qquad \forall n \geq \frac{1-\epsilon}{\epsilon} + 1.$$

Thus the construction satisfies the requirements by letting $N = \left\lceil \max\{\frac{1-\epsilon}{H(\epsilon)} + 1, \frac{1-\epsilon}{\epsilon} + 1\}\right\rceil$, and $C = -\log\epsilon$. □

Exercise 5.3 Typical sets [all]

Consider a DMS with a two symbol alphabet $\{a, b\}$ where $p_X(a) = 2/3$ and $p_X(b) = 1/3$. Let $X^n = (X_1, \ldots, X_n)$ be a string of symbols emitted by the source with $n = 100{,}000$. Let $W(X_j)$ be the suprisal for the j-th source output, i.e., $W(X_j) = -\log 2/3$ for $X_j = a$ and $-\log 1/3$ for $X_j = b$. Define $W(X^n) = \sum_{j=1}^{n} W(X_j)$.

a.) Find the variance of $W(X_j)$. For $\epsilon = 0.01$, evaluate a bound on the probability of the typical set $\mathcal{A}_\epsilon^{(n)}$ using Chebyshev's inequality.

Solution: For notational convenience, we will denote the log pmf random variable by W. Now, note that W takes on values $-\log 2/3$ with probability $2/3$ and $-\log 1/3$ with probability $1/3$. Hence,

$$\mathrm{Var}(W) = \mathbb{E}[W^2] - \mathbb{E}[W]^2 = \frac{2}{9}.$$

The bound on the typical set, as derived using Chebyshev's inequality is

$$\Pr(X^n \in A_\epsilon^{(n)}) \geq 1 - \frac{\sigma_W^2}{n\epsilon^2}.$$

Substituting the values of $n = 10^5$ and $\epsilon = 0.01$, we obtain

$$\Pr(X^n \in A_\epsilon^{(n)}) \geq 1 - \frac{1}{45} = \frac{44}{45}$$

Loosely speaking this means that if we were to look at sequences of length 100,000 generated from our DMS, more than 97% of the time the sequence will be typical. □

b.) Let N_a be the number of a's in the string $X^n = (X_1, \ldots, X_n)$. The random variable (rv) N_a is the sum of n iid rv's. Show what these rv's are.

Solution: The rv N_a is the sum of n iid rv's Y_i , $N_a = \sum_{i=1}^n Y_i$ where Y_i's are Bernoulli with $\Pr(Y_i = 1) = 2/3$. □

c.) Express the rv $W(X^n)$ as a function of the rv N_a. Note how this depends on n.

Solution: The probability of a particular sequence X^n with N_a number of a's $(2/3)^{N_a}(1/3)^{n-N_a}$. Hence,

$$W(X^n) = -\log p_{X^n}(x^n) = -\log[(2/3)^{N_a}(1/3)^{n-N_a}] = n\log 3 - N_a.$$

□

d.) Express the typical set in terms of bounds on N_a. Use Chebyshev's inequality to derive bounds on the probability of the typical set, using properties of N_a instead of $W(X_j)$.
Hint: You may write $\mathcal{A}_\epsilon^{(n)} = \{x^n : \alpha < N_a < \beta\}$ and calculate α and β.

Solution: For a sequence X^n to be typical, it must satisfy

$$\left| -\frac{1}{n}\log p_{X^n}(x^n) - H(X) \right| < \epsilon$$

From (a) the source entropy is $H(X) = \mathbb{E}[W(X)] = \log 3 - 2/3$ and substituting in ϵ and $W(X^n)$ from part (d), we get

$$\left| \frac{N_a}{n} - \frac{2}{3} \right| \leq 0.01$$

Note the intuitive appeal of this condition! It says that for a sequence to be typical, the proportion of a's in that sequence will be very close to the probability that the DMS generates an a. Plugging in the value of n in the above equation, we get the bounds on

$$65,667 \leq N_a \leq 67,666.$$

□

e.) Find $\Pr(N_a = i)$ for $i = 0, 1, 2$. Find the probability of each individual string x^n for those values of i. Find the particular string x^n that has maximum probability over all sample values of X^n. What are the next most probable n-strings. Give a brief discussion of why the most probable n-strings are not regarded as typical strings.

Solution:

$$\Pr(N_a = 0) = \left(\frac{1}{3}\right)^n, \ \Pr(x^n = b^n) = \left(\frac{1}{3}\right)^n.$$

$$\Pr(N_a = 1) = n\left(\frac{2}{3}\right)\left(\frac{1}{3}\right)^{n-1}, \quad \text{the probability of each } x^n \text{ is } \left(\frac{2}{3}\right)\left(\frac{1}{3}\right)^{n-1}.$$

$$\Pr(N_a = 2) = \frac{n(n-1)}{2}\left(\frac{2}{3}\right)^2\left(\frac{1}{3}\right)^{n-2}, \quad \text{the probability of each } x^n \text{ is } \left(\frac{2}{3}\right)^2\left(\frac{1}{3}\right)^{n-2}.$$

The particular string x^n with maximum probability is a^n:

$$\Pr(x^n = a^n) = \left(\frac{2}{3}\right)^n.$$

The next most probable n-strings are x^n with 1 symbol b and $(n-1)$ symbol a's.

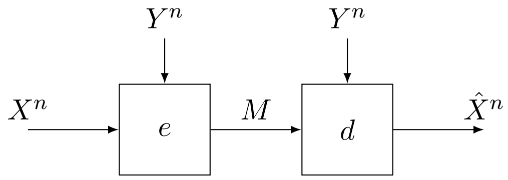
The typical strings should have around $\frac{2}{3}n$ symbol a's and around $\frac{1}{3}n$ symbol b's. The most probable n-strings are usually far from this situation. $\square$

Exercise 5.4 Source coding with side information [EE5139]

Consider a memoryless source $(\boldsymbol{X}, \boldsymbol{Y})$ that produces in each iteration two random variables, X_i and Y_i, where X_i is private information and Y_i is public information. The pairs (X_i, Y_i) follow a joint distribution P_{XY} and are i.i.d.. We are looking for a fixed-length block code that compresses the private information $X^n = (X_1, X_2, \ldots, X_n)$ using the public information $Y^n = (Y_1, Y_2, \ldots, Y_n)$ such that the code can be decoded asymptotically error-free with help of the public information.
An $(n, 2^L)$-code for such a source is given by an encoder, $e : (X^n, Y^n) \to M$, and decoder, $d : (M, Y^n) \to \hat{X}^n$, as illustrated in the figure below. The codeword $M \in \{0,1\}^L$ is a binary string of length L. We define $R^*(\boldsymbol{X}|\boldsymbol{Y})$ as the infimum over all rates R such that there exists a sequence of $(n, 2^{nR})$-codes satisfying

$$\lim_{n\to\infty} \Pr\left[X^n \neq \hat{X}^n\right] = 0, \quad \text{where} \quad \hat{X}^n = d_n(e_n(X^n, Y^n), Y^n) \tag{2}$$

is a function of both X^n and Y^n. We want to establish that $R^*(\boldsymbol{X}|\boldsymbol{Y}) = H(X|Y)$.



a.) Determine $R^*(\boldsymbol{X}|\boldsymbol{Y})$, by intuitive or formal arguments, for the simple cases where

i.) X and Y are independent,
Solution: $R^*(\boldsymbol{X}|\boldsymbol{Y}) = H(X)$. If X and Y are independent, it means Y does not provide useful side information about X. $\square$

ii.) $X = Y$,
Solution: $R^*(\boldsymbol{X}|\boldsymbol{Y}) = 0$. $X = Y$ means that the decoder can exactly recover X^n simply from the side information Y^n. $\square$

b.) By explicitly constructing a code for the source (X, Y) using codes for the sources Y and X (with side information Y), show that $R^*(\boldsymbol{X}, \boldsymbol{Y}) \leq R^*(\boldsymbol{X}|\boldsymbol{Y}) + R^*(\boldsymbol{Y})$.

Solution: A code for the source (X, Y) can be constructed by concatenating the codes for sources Y and X (with side information Y). The error probability is given by

$$\begin{aligned}\Pr\left[(\hat{X}^n, \hat{Y}^n) \neq (X^n, Y^n)\right] &= \Pr\left[\hat{Y}^n \neq Y^n \vee \hat{X}^n \neq X^n\right] \\ &\leq \Pr\left[\hat{Y}^n \neq Y^n\right] + \Pr\left[\hat{X}^n \neq X^n\right] \to 0, \text{ as } n \to \infty.\end{aligned}$$

Let L_Y and $L_{X|Y}$ denote the number of bits of the optimal (shortest) codes for the sources Y and X with side information Y, respectively. Then we have

$$R^*(\boldsymbol{X}, \boldsymbol{Y}) \leq \frac{L_{X|Y} + L_Y}{n} = R^*(\boldsymbol{X}|\boldsymbol{Y}) + R^*(\boldsymbol{Y}).$$

$\square$

c.) Show that the converse, $R^*(\boldsymbol{X}|\boldsymbol{Y}) \geq H(X|Y)$ using Fano's inequality. **Hint:** You will also need the following sequence of inequalities, which needs to be verified.

$$
\begin{aligned}
H(X^n|\hat{X}^n) &\geq H(X^n|Y^n M) && (3)\\
&= H(X^n M|Y^n) - H(M|Y^n) && (4)\\
&\geq H(X^n M|Y^n) - L && (5)\\
&\geq H(X^n|Y^n) - L\,. && (6)
\end{aligned}
$$

Solution: We first verify the sequence of inequalities. Eq. (3) is the data-processing inequality applied to the fact that $\hat{X}^n$ is computed from Y^n and M. Eq. (4) is the chain rule for conditional entropy. Eq. (5) follows from the dimension bound for $|M| \leq 2^L$. Finally, Eq. (6) uses the chain rule and the fact that $H(M|X^n Y^n) \geq 0$.

Consider now any sequence of $(n, 2^L)$-codes that satisfy $\epsilon_n = \Pr[\hat{X}^n \neq X^n] \to 0$ in the limit $n \to \infty$.

By Fano's inequality and using the given sequence of inequalities, we have

$$H(\epsilon_n) + \epsilon_n n \log|\mathcal{X}| \geq H(X^n|\hat{X}^n) \geq H(X^n|Y^n) - L.$$

Hence, as $n \to \infty$,

$$
\begin{aligned}
\frac{L}{n} &\geq \frac{1}{n}(H(X^n|Y^n) - H(\epsilon_n) - \epsilon_n n \log|\mathcal{X}|)\\
&\geq \frac{1}{n}H(X^n|Y^n) - \frac{1}{n} - \epsilon_n \log|\mathcal{X}|\\
&= H(X|Y) - \frac{1}{n} - \epsilon_n \log|\mathcal{X}|\\
&\to H(X|Y).
\end{aligned}
$$

Since this holds for any sequence of codes, we conclude that $R^*(\boldsymbol{X}|\boldsymbol{Y}) \geq H(X|Y)$. $\square$

d.) Give a formal proof or a sketch of a proof that $R^*(\boldsymbol{X}|\boldsymbol{Y}) \leq H(X|Y)$. **Hint:** Consider the typical set

$$\mathcal{A}_\epsilon^{(n)}(\boldsymbol{X}|\boldsymbol{Y}) := \left\{(x^n, y^n) \in \mathcal{X}^n \times \mathcal{Y}^n : \left|\frac{1}{n}\log\frac{1}{P_{X^n|Y^n}(x^n|y^n)} - H(X|Y)\right| \leq \epsilon\right\}. \quad (7)$$

Solution: For any $\epsilon > 0$, define a typical set for $(\boldsymbol{X}, \boldsymbol{Y})$,

$$\mathcal{A}_\epsilon^{(n)}(\boldsymbol{X}|\boldsymbol{Y}) \triangleq \left\{(x^n, y^n) \in \mathcal{X}^n \times \mathcal{Y}^n : \left|\frac{1}{n}\log\frac{1}{P_{X^n|Y^n}(x^n|y^n)} - H(X|Y)\right| \leq \epsilon\right\},$$

where $P_{X^n|Y^n}(x^n|y^n) = \prod_{i=1}^n P_{X|Y}(x_i|y_i)$ for all $(x^n, y^n) \in \mathcal{X}^n \times \mathcal{Y}^n$.

By definition, for any $(x^n, y^n) \in \mathcal{A}_\epsilon^{(n)}(\boldsymbol{X}|\boldsymbol{Y})$, we can establish that

$$
\begin{aligned}
& P_{X^n|Y^n}(x^n|y^n) \geq 2^{-n(H(X|Y)+\epsilon)}\\
\implies\ & 1 \geq \sum_{x^n, y^n \in \mathcal{A}_\epsilon^{(n)}(\boldsymbol{X}|\boldsymbol{Y})} P_{X^n|Y^n}(x^n|y^n) \geq |\mathcal{A}_\epsilon^{(n)}(\boldsymbol{X}|\boldsymbol{Y})| 2^{-n(H(X|Y)+\epsilon)}\\
\implies\ & |\mathcal{A}_\epsilon^{(n)}(\boldsymbol{X}|\boldsymbol{Y})| \leq 2^{nH(X|Y)+\epsilon}.
\end{aligned}
$$

Furthermore, let $Z_i = \log \frac{1}{P_{X|Y}(X_i|Y_i)} - H(X|Y)$ and we have

$$\begin{aligned}
\Pr\left[(X^n, Y^n) \in \mathcal{A}_\epsilon^{(n)}(\boldsymbol{X}|\boldsymbol{Y})\right] &= \Pr\left[\left|\frac{1}{n}\log\frac{1}{P_{X^n|Y^n}(X^n|Y^n)} - H(X|Y)\right| \le \epsilon\right] \\
&= \Pr\left[\left|\frac{1}{n}\log\frac{1}{\prod_{i=1}^n P_{X|Y}(X_i|Y_i)} - H(X|Y)\right| \le \epsilon\right] \\
&= \Pr\left[\left|\frac{1}{n}\sum_{i=1}^n \log\frac{1}{P_{X|Y}(X_i|Y_i)} - H(X|Y)\right| \le \epsilon\right] \\
&= 1 - \Pr\left[\left|\frac{1}{n}\sum_{i=1}^n \log\frac{1}{P_{X|Y}(X_i|Y_i)} - H(X|Y)\right| > \epsilon\right] \\
&= 1 - \Pr\left[\left|\frac{1}{n}\sum_{i=1}^n Z_i\right| > \epsilon\right].
\end{aligned}$$

Since Z_i are i.i.d and zero mean, by the weak law of large numbers, we have

$$\lim_{n\to\infty}\Pr\left[(X^n, Y^n) \in \mathcal{A}_\epsilon^{(n)}(\boldsymbol{X}|\boldsymbol{Y})\right] = 1 - \lim_{n\to\infty}\Pr\left[\left|\frac{1}{n}\sum_{i=1}^n Z_i\right| > \epsilon\right] = 1.$$

Encoder e:

$$e(x^n|y^n) = \begin{cases} m(x^n|y^n) & (x^n, y^n) \in \mathcal{A}_\epsilon^{(n)}(\boldsymbol{X}|\boldsymbol{Y}), \\ 0^L & (x^n, y^n) \notin \mathcal{A}_\epsilon^{(n)}(\boldsymbol{X}|\boldsymbol{Y}). \end{cases}$$

Decoder d: Given y^n, m, output any x^n such that $m = m(x^n|y^n)$.

Then the error probability is given by

$$\Pr\left[\hat{X}^n \ne X^n\right] = 1 - \Pr\left[(X^n, Y^n) \in \mathcal{A}_\epsilon^{(n)}(\boldsymbol{X}|\boldsymbol{Y})\right] \to 0 \quad \text{for } n \to \infty.$$

This implies that $R = \frac{L}{n} \le H(X|Y) + \epsilon$ is achievable and thus

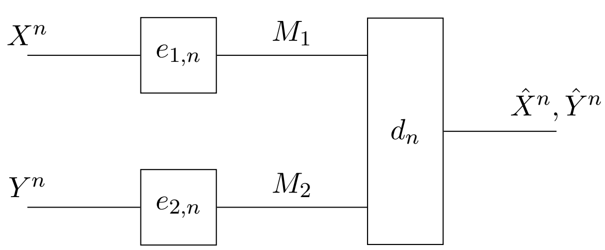
$$R^*(\boldsymbol{X}|\boldsymbol{Y}) \le H(X|Y).$$

□

Exercise 5.5 Achievability for the Slepian–Wolf coding problem [EE6139]

We return to the Exercise 4.4 from the last homework. Let X and Y be a pair of jointly distributed random variables. (X is distributed on finite set $\mathcal{X}$, and Y is distributed on finite set $\mathcal{Y}$.) An $(n, 2^{nL_1}, 2^{nL_2})$-separately-encoded-jointly-decoded source code consists of a pair of encoders e_1, e_2, and a decoder d, where

- $e_1 : \mathcal{X}^n \to \{0,1\}^{nL_1}$,
- $e_2 : \mathcal{Y}^n \to \{0,1\}^{nL_2}$, and
- $d : \{0,1\}^{nL_1} \times \{0,1\}^{nL_2} \to \mathcal{X}^n \times \mathcal{Y}^n$.



The rate pair (R_1, R_2) is said to be achievable for DMS (X, Y) if there exists a sequence of $(n, 2^{nL_1}, 2^{nL_2})$-codes with encoders $e_{1,n}$, $e_{2,n}$ and decoder d_n such that

$$\lim_{n\to\infty} P\{(\hat{X}^n, \hat{Y}^n) \neq (X^n, Y^n)\} = 0$$

where

$$(\hat{X}^n, \hat{Y}^n) = d_n(M_1, M_2),\ M_1 = e_{1,n}(X^n),\ \text{and } M_2 = e_{2,n}(Y^n)$$

are the reconstructed source and codewords respectively.
This time, we are interested in the achievability of the problem.

a.) **An alternative for typical sequences** Given $n \in \mathbb{N}$ and $\epsilon \in (0,1)$, we define the set of Y-sequences as

$$\mathcal{T}_\epsilon^{(n)}(Y) \triangleq \left\{ \boldsymbol{y} \in \mathcal{Y}^n : \left| \frac{\sum_{i=1}^n \delta_{y,y_i}}{n} - p_Y(y) \right| < \left\lceil \sqrt{\frac{|\mathcal{Y}|}{\epsilon}} \right\rceil \sqrt{\frac{p_Y(y)(1-p_Y(y))}{n}} \quad \forall y \in \mathcal{Y} \right\}.$$

Show that

i.) $P\left[Y^n \in \mathcal{T}_\epsilon^{(n)}(Y)\right] \geq 1 - \epsilon$

Solution:

$$\begin{aligned} P\left[Y^n \notin \mathcal{T}_\epsilon^{(n)}(Y)\right] &= P\left[\left| \frac{\sum_{i=1}^n \delta_{y,Y_i}}{n} - p_Y(y) \right| \geq \left\lceil \sqrt{\frac{|\mathcal{Y}|}{\epsilon}} \right\rceil \sqrt{\frac{p_Y(y)(1-p_Y(y))}{n}} \quad \exists y \in \mathcal{Y}\right] \\ &\leq \sum_{y\in\mathcal{Y}} P\left[\left| \frac{\sum_{i=1}^n \delta_{y,Y_i}}{n} - p_Y(y) \right| \geq \left\lceil \sqrt{\frac{|\mathcal{Y}|}{\epsilon}} \right\rceil \sqrt{\frac{p_Y(y)(1-p_Y(y))}{n}}\right] \\ &\leq \sum_{y\in\mathcal{Y}} \left(\sqrt{\frac{|\mathcal{Y}|}{\epsilon}}\right)^{-2} = \epsilon, \end{aligned}$$

where we have used Chebyshev's inequality in the last line. □

ii.) There exists some $A > 0$ independent from n and ϵ such that

$$2^{-nH(Y) - A\sqrt{n/\epsilon}} < p_{Y^n}(\boldsymbol{y}) < 2^{-nH(Y) + A\sqrt{n/\epsilon}}$$

for all $\boldsymbol{y} \in \mathcal{T}_\epsilon^{(n)}(Y)$.

Solution: Note that for any n-length vector $\boldsymbol{y}$, we have

$$\log p_{Y^n}(\boldsymbol{y}) = \log \prod_{y\in\mathcal{Y}} p_Y(y)^{\sum_{i=1}^n \delta_{y,y_i}} = \sum_{y\in\mathcal{Y}} \left(\sum_{i=1}^n \delta_{y,y_i}\right) \cdot \log p_Y(y)$$

For $\boldsymbol{y} \in \mathcal{T}_\epsilon^{(n)}(Y)$, note that

$$\begin{aligned} \sum_{i=1}^n \delta_{y,y_i} &\in \left(np_Y(y) - n \left\lceil \sqrt{\frac{|\mathcal{Y}|}{\epsilon}} \right\rceil \sqrt{\frac{p_Y(y)(1-p_Y(y))}{n}}, \right. \\ &\qquad \left. np_Y(y) + n \left\lceil \sqrt{\frac{|\mathcal{Y}|}{\epsilon}} \right\rceil \sqrt{\frac{p_Y(y)(1-p_Y(y))}{n}} \right). \\ &\subset \left(np_Y(y) - f(y)\sqrt{\frac{n}{\epsilon}}, np_Y(y) + f(y)\sqrt{\frac{n}{\epsilon}} \right) \end{aligned}$$

where $f(y) \triangleq 2\sqrt{|\mathcal{Y}| \cdot p_Y(y)(1-p_Y(y))}$. Thus, by defining

$$A \triangleq -\sum_{y\in\mathcal{Y}} f(y) \cdot \log p_Y(y),$$

we have

$$-nH(Y) - A\sqrt{\frac{n}{\epsilon}} < \log p_{Y^n}(\boldsymbol{y}) < -nH(Y) + A\sqrt{\frac{n}{\epsilon}},$$

which are equivalent to the to-be-proven inequalities. $\square$

iii.) $\lim_{n\to\infty} \frac{1}{n} \log_2 \left|\mathcal{T}_\epsilon^{(n)}(Y)\right| = H(Y)$.

Solution: Note that

$$\left|\mathcal{T}_\epsilon^{(n)}(Y)\right| \cdot \min_{\boldsymbol{y}\in\mathcal{T}_\epsilon^{(n)}(Y)} p_{Y^n}(\boldsymbol{y}) \leq P\left[Y^n \in \mathcal{T}_\epsilon^{(n)}(Y)\right] \leq 1,$$

$$\left|\mathcal{T}_\epsilon^{(n)}(Y)\right| \cdot \max_{\boldsymbol{y}\in\mathcal{T}_\epsilon^{(n)}(Y)} p_{Y^n}(\boldsymbol{y}) \geq P\left[Y^n \in \mathcal{T}_\epsilon^{(n)}(Y)\right] \geq 1-\epsilon.$$

Combining with the results from ii.), we have

$$(1-\epsilon)\cdot 2^{nH(Y)-A\sqrt{n/\epsilon}} \leq \left|\mathcal{T}_\epsilon^{(n)}(Y)\right| \leq 2^{nH(Y)+A\sqrt{n/\epsilon}},$$

or, equivalently,

$$\log(1-\epsilon) + nH(Y) - A\sqrt{n/\epsilon} \leq \log\left|\mathcal{T}_\epsilon^{(n)}(Y)\right| \leq nH(Y) + A\sqrt{n/\epsilon}.$$

Thus,

$$\left|\frac{1}{n}\log\left|\mathcal{T}_\epsilon^{(n)}(Y)\right| - H(Y)\right| \leq \frac{|\log(1-\epsilon)|}{n} + \frac{A}{\sqrt{\epsilon}\sqrt{n}} \to 0$$

as $n\to\infty$. $\square$

b.) **Position-based coding** Given positive integer n and $\epsilon > 0$, let $M_X \triangleq \left\lfloor 2^{n(H(Y|X)+\epsilon)}\right\rfloor$, and let M be another positive integer. Let $\{\boldsymbol{X}_{i,j}\}_{i,j}$ be a set of i.i.d. random variables on $\mathcal{X}^n$, where $i \in \{1,\ldots,M_X\}$, $j\in\{1,\ldots,M\}$, and

$$p_{\boldsymbol{X}_{i,j}}(\boldsymbol{x}) = \prod_{k=1}^{n} p_X(x_k)$$

for each (i,j).

i.) Suppose $I(X,Y) > \frac{1}{2}\epsilon$, and let $M = \left\lfloor 2^{n(I(X,Y)-\frac{1}{2}\epsilon)}\right\rfloor$. Prove that, for n large enough,

$$P\left[X^n \neq \boldsymbol{X}_{i,j}\ \forall(i,j)\right] < 2\epsilon.$$

Solution: Considering the set $\mathcal{T}_\epsilon^n(X)\subset\mathcal{X}^n$, and that $\{\boldsymbol{X}_{i,j}\}_{i,j}$ are i.i.d., we have

$$\begin{aligned}
&P\left[X^n \neq \boldsymbol{X}_{i,j}\ \forall(i,j)\right] = P\left[X^n \neq \boldsymbol{X}_{i,j}\ \forall(i,j)|X^n=\boldsymbol{x}\right]\cdot P[X^n=\boldsymbol{x}]\\
&= \sum_{\boldsymbol{x}\in\mathcal{X}^n} p_{X^n}(\boldsymbol{x})\cdot\prod_{i,j}(1-p_{\boldsymbol{X}_{i,j}}(\boldsymbol{x}))\\
&= \sum_{\boldsymbol{x}\in\mathcal{X}^n} p_{X^n}(\boldsymbol{x})\cdot(1-p_{X^n}(\boldsymbol{x}))^{MM_X}\\
&= \sum_{\boldsymbol{x}\in\mathcal{T}_\epsilon^n(X)} p_{X^n}(\boldsymbol{x})\cdot(1-p_{X^n}(\boldsymbol{x}))^{MM_X} + \sum_{\boldsymbol{x}\in\mathcal{X}^n\setminus\mathcal{T}_\epsilon^n(X)} p_{X^n}(\boldsymbol{x})\cdot(1-p_{X^n}(\boldsymbol{x}))^{MM_X}\\
&\leq \sum_{\boldsymbol{x}\in\mathcal{T}_\epsilon^n(X)} p_{X^n}(\boldsymbol{x})\cdot(1-p_{X^n}(\boldsymbol{x}))^{MM_X} + \epsilon.
\end{aligned}$$

Note that for any $\boldsymbol{x} \in \mathcal{T}_\epsilon^n(X)$, we have $p_{X^n}(\boldsymbol{x}) \geq 2^{-nH(X)-A\sqrt{n/\epsilon}}$ for some A independent from n and ϵ. In this case,

$$\begin{aligned}\sum_{\boldsymbol{x}\in\mathcal{T}_\epsilon^n(X)} p_{X^n}(\boldsymbol{x}) \cdot (1-p_{X^n}(\boldsymbol{x}))^{MM_X} &\leq \left[1-2^{-nH(X)-A\sqrt{n/\epsilon}}\right]^{MM_X} \cdot \sum_{\boldsymbol{x}\in\mathcal{T}_\epsilon^n(X)} p_{X^n}(\boldsymbol{x}) \\ &\leq \left[1-2^{-nH(X)-A\sqrt{n/\epsilon}}\right]^{MM_X}.\end{aligned}$$

Furthermore, denoting the quantity on the right-hand side of the above line by Z, we have

$$\begin{aligned}\log Z &= MM_X \log\left[1-2^{-nH(X)-A\sqrt{n/\epsilon}}\right]^{MM_X} \\ &\leq 2^{n(I(X,Y)-\frac{1}{2}\epsilon)} \cdot 2^{n(H(Y|X)+\epsilon)} \cdot \log\left[1-2^{-nH(X)-A\sqrt{n/\epsilon}}\right]^{MM_X} \\ &\leq -2^{n(I(X,Y)-\frac{1}{2}\epsilon)} \cdot 2^{n(H(Y|X)+\epsilon)} \cdot 2^{-nH(X)-A\sqrt{n/\epsilon}} \\ &= -2^{n\epsilon/2-A\sqrt{n/\epsilon}}.\end{aligned}$$

Notice that for fixed ϵ, above tends to $-\infty$ as $n \to \infty$. It must hold that $-2^{n\epsilon/2-A\sqrt{n/\epsilon}} < \log\epsilon$ (and thus $Z < \epsilon$) for n large enough. Therefore,

$$P\left[X^n \neq \boldsymbol{X}_{i,j}\ \forall(i,j)\right] \leq Z + \epsilon < 2\epsilon$$

for n large enough. □

ii.) Let A and B be a pair of random variable denoting the "smallest" indices a, b such that that $X^n = \boldsymbol{X}_{a,b}$. Namely,

$$p_{A,B|X^n,\{\boldsymbol{X}_{i,j}\}}(a,b|\boldsymbol{x},\{\boldsymbol{x}_{i,j}\}) = \begin{cases} 1 & \text{if} \begin{array}{ll} \boldsymbol{x} = \boldsymbol{x}_{a,b} & \\ \boldsymbol{x} \neq \boldsymbol{x}_{i,j} & \forall i < a \\ \boldsymbol{x} \neq \boldsymbol{x}_{a,j} & \forall j < b \end{array} \\ 0 & \text{otherwise} \end{cases}.$$

We take the convention that $(A,B) = (\infty,\infty)$ if $X^n \neq \boldsymbol{X}_{i,j}$ for all i, j. Prove that

$$P\left[A<\infty,\, B<\infty,\, p_{Y^n|X^n}(Y^n|\boldsymbol{X}_{A,j}) \geq p_{Y^n|X^n}(Y^n|X^n)\ \exists j \neq B\right] < \epsilon$$

for n large enough.

Solution: Firstly, we can rewrite above probability into

$$\begin{aligned}&P\left[A<\infty,\, B<\infty,\, p_{Y^n|X^n}(Y^n|\boldsymbol{X}_{A,j}) \geq p_{Y^n|X^n}(Y^n|X^n)\ \exists j \neq B\right] \\ &= \sum_{\boldsymbol{x},\boldsymbol{y}} p_{X^nY^n}(\boldsymbol{x},\boldsymbol{y}) \sum_{a=1}^{M_X} p_{A|X^n,Y^n}(a|\boldsymbol{x},\boldsymbol{y}) \sum_{b=1}^{M} p_{B|A,X^n,Y^n}(b|a,\boldsymbol{x},\boldsymbol{y}) \cdot \\ &\quad P\left[p_{Y^n|X^n}(\boldsymbol{y}|\boldsymbol{X}_{a,j}) \geq p_{Y^n|X^n}(\boldsymbol{y}|\boldsymbol{x})\ \exists j \neq b | X^n = \boldsymbol{x}, Y^n = \boldsymbol{y}, A = a, B = b\right] \\ &= \sum_{\boldsymbol{x},\boldsymbol{y}} p_{X^nY^n}(\boldsymbol{x},\boldsymbol{y}) \cdot \sum_{a=1}^{M_X} (1-p_{X^n}(\boldsymbol{x}))^{(a-1)M} \cdot \sum_{b=1}^{M} (1-p_{X^n}(\boldsymbol{x}))^{b-1} p_{X^n}(\boldsymbol{x}) \cdot C_{a,b}(\boldsymbol{x},\boldsymbol{y})\end{aligned}$$

where

$$C_{a,b}(\boldsymbol{x},\boldsymbol{y}) \triangleq P\left[\begin{array}{c} p_{Y^n|X^n}(\boldsymbol{y}|\boldsymbol{X}_{a,j}) \geq p_{Y^n|X^n}(\boldsymbol{y}|\boldsymbol{x}) \\ \exists j \neq b \end{array} \middle| \begin{array}{ll} X^n = \boldsymbol{x},\, Y^n = \boldsymbol{y} & \\ \boldsymbol{X}_{i,j} \neq \boldsymbol{x} & \forall i < a \\ \boldsymbol{X}_{a,j} \neq \boldsymbol{x} & \forall j < b \end{array}\right].$$

For any $\rho \in (0,1)$, we can bound $C_{a,b}(\boldsymbol{x},\boldsymbol{y})$ as

$$C_{a,b}(\boldsymbol{x},\boldsymbol{y}) \leq \left\{ \sum_{j\neq b} P\left[p_{Y^n|X^n}(\boldsymbol{y}|\boldsymbol{X}_{a,j}) \geq p_{Y^n|X^n}(\boldsymbol{y}|\boldsymbol{x})|\cdots\right] \right\}^{\rho}.$$

Note that, for $j > b$,

$$\begin{aligned}
&P\left[p_{Y^n|X^n}(\boldsymbol{y}|\boldsymbol{X}_{a,j}) \geq p_{Y^n|X^n}(\boldsymbol{y}|\boldsymbol{x})|\cdots\right] \\
=&P\left[p_{Y^n|X^n}(\boldsymbol{y}|\boldsymbol{X}_{a,j}) \geq p_{Y^n|X^n}(\boldsymbol{y}|\boldsymbol{x})|X^n=\boldsymbol{x}, Y^n=\boldsymbol{y}\right] \\
=&\sum_{\tilde{\boldsymbol{x}}\in\mathcal{X}^n: p_{Y^n|X^n}(\boldsymbol{y}|\tilde{\boldsymbol{x}})\geq p_{Y^n|X^n}(\boldsymbol{y}|\boldsymbol{x})} p_{X^n}(\tilde{\boldsymbol{x}}).
\end{aligned}$$

Whereas, for $j < b$,

$$\begin{aligned}
&P\left[p_{Y^n|X^n}(\boldsymbol{y}|\boldsymbol{X}_{a,j}) \geq p_{Y^n|X^n}(\boldsymbol{y}|\boldsymbol{x})|\cdots\right] \\
=&P\left[p_{Y^n|X^n}(\boldsymbol{y}|\boldsymbol{X}_{a,j}) \geq p_{Y^n|X^n}(\boldsymbol{y}|\boldsymbol{x})|X^n=\boldsymbol{x}, Y^n=\boldsymbol{y}, \boldsymbol{X}_{a,j}\neq\boldsymbol{x}\right] \\
=&\frac{P\left[p_{Y^n|X^n}(\boldsymbol{y}|\boldsymbol{X}_{a,j}) \geq p_{Y^n|X^n}(\boldsymbol{y}|\boldsymbol{x}), \boldsymbol{X}_{a,j}\neq\boldsymbol{x}|X^n=\boldsymbol{x}, Y^n=\boldsymbol{y}\right]}{P\left[\boldsymbol{X}_{a,j}\neq\boldsymbol{x}|X^n=\boldsymbol{x}, Y^n=\boldsymbol{y}\right]} \\
=&\frac{P\left[p_{Y^n|X^n}(\boldsymbol{y}|\boldsymbol{X}_{a,j}) \geq p_{Y^n|X^n}(\boldsymbol{y}|\boldsymbol{x}), \boldsymbol{X}_{a,j}\neq\boldsymbol{x}|X^n=\boldsymbol{x}, Y^n=\boldsymbol{y}\right]}{P\left[\boldsymbol{X}_{a,j}\neq\boldsymbol{x}\right]} \\
\leq&\frac{P\left[p_{Y^n|X^n}(\boldsymbol{y}|\boldsymbol{X}_{a,j}) \geq p_{Y^n|X^n}(\boldsymbol{y}|\boldsymbol{x})|X^n=\boldsymbol{x}, Y^n=\boldsymbol{y}\right]}{P\left[\boldsymbol{X}_{a,j}\neq\boldsymbol{x}\right]} \\
=&(1-p_{X^n}(\boldsymbol{x}))^{-1}\cdot\sum_{\tilde{\boldsymbol{x}}\in\mathcal{X}^n: p_{Y^n|X^n}(\boldsymbol{y}|\tilde{\boldsymbol{x}})\geq p_{Y^n|X^n}(\boldsymbol{y}|\boldsymbol{x})} p_{X^n}(\tilde{\boldsymbol{x}}).
\end{aligned}$$

Hence, following holds for all $s \geq 0$

$$\begin{aligned}
C_{a,b}(\boldsymbol{x},\boldsymbol{y}) &\leq \left\{ \left[(b-1)(1-p_{X^n}(\boldsymbol{x}))^{-1}+M-b\right]\cdot \sum_{\substack{\tilde{\boldsymbol{x}}\in\mathcal{X}^n: \\ p_{Y^n|X^n}(\boldsymbol{y}|\tilde{\boldsymbol{x}})\geq p_{Y^n|X^n}(\boldsymbol{y}|\boldsymbol{x})}} p_{X^n}(\tilde{\boldsymbol{x}}) \right\}^{\rho} \\
&\leq \left[(b-1)(1-p_{X^n}(\boldsymbol{x}))^{-1}+M-b\right]^{\rho}\cdot\left[\sum_{\tilde{\boldsymbol{x}}\in\mathcal{X}^n} p_{X^n}(\tilde{\boldsymbol{x}})\left(\frac{p_{Y^n|X^n}(\boldsymbol{y}|\tilde{\boldsymbol{x}})}{p_{Y^n|X^n}(\boldsymbol{y}|\boldsymbol{x})}\right)^{s}\right]^{\rho},
\end{aligned}$$

since $\frac{p_{Y^n|X^n}(\boldsymbol{y}|\tilde{\boldsymbol{x}})}{p_{Y^n|X^n}(\boldsymbol{y}|\boldsymbol{x})} \geq 1$ for all $\tilde{\boldsymbol{x}} \in \mathcal{X}^n$ such that $p_{Y^n|X^n}(\boldsymbol{y}|\tilde{\boldsymbol{x}}) \geq p_{Y^n|X^n}(\boldsymbol{y}|\boldsymbol{x})$. Substituting above bound on $C_{a,b}(\boldsymbol{x},\boldsymbol{y})$ into the expression for the targeting probability, we have

$$\begin{aligned}
&P\left[A<\infty,\, B<\infty,\, p_{Y^n|X^n}(Y^n|\boldsymbol{X}_{A,j}) \geq p_{Y^n|X^n}(Y^n|X^n)\ \exists j\neq B\right] \\
\leq&\sum_{\boldsymbol{x},\boldsymbol{y}} p_{X^nY^n}(\boldsymbol{x},\boldsymbol{y})\cdot\sum_{a=1}^{M_X}(1-p_{X^n}(\boldsymbol{x}))^{(a-1)M}\cdot\sum_{b=1}^{M}(1-p_{X^n}(\boldsymbol{x}))^{b-1}p_{X^n}(\boldsymbol{x})\cdot \\
&\left[(b-1)(1-p_{X^n}(\boldsymbol{x}))^{-1}+M-b\right]^{\rho}\cdot\left[\sum_{\tilde{\boldsymbol{x}}\in\mathcal{X}^n} p_{X^n}(\tilde{\boldsymbol{x}})\left(\frac{p_{Y^n|X^n}(\boldsymbol{y}|\tilde{\boldsymbol{x}})}{p_{Y^n|X^n}(\boldsymbol{y}|\boldsymbol{x})}\right)^{s}\right]^{\rho}.
\end{aligned}$$

We make the following claim, and defer its proof to the very end.
Claim. For $\alpha \in [0,1)$, $\beta \in [0,1]$ and m, n being positive integers, it holds that

$$\sum_{i=1}^{m}\alpha^{(i-1)n}\cdot\sum_{j=1}^{n}\alpha^{j-1}\cdot\left(\frac{j-1}{\alpha}+n-j\right)^{\beta} \leq \frac{n^{\beta}}{1-\alpha}.$$

Using above claim, we have

$$
\begin{aligned}
&P\left[A<\infty,\, B<\infty,\, p_{Y^n|X^n}(Y^n|\boldsymbol{X}_{A,j}) \ge p_{Y^n|X^n}(Y^n|X^n)\ \exists j \neq B\right] \\
\le &\sum_{\boldsymbol{x},\boldsymbol{y}} p_{X^nY^n}(\boldsymbol{x},\boldsymbol{y}) \cdot M^\rho \cdot \left[\sum_{\tilde{\boldsymbol{x}}\in\mathcal{X}^n} p_{X^n}(\tilde{\boldsymbol{x}}) \left(\frac{p_{Y^n|X^n}(\boldsymbol{y}|\tilde{\boldsymbol{x}})}{p_{Y^n|X^n}(\boldsymbol{y}|\boldsymbol{x})}\right)^s\right]^\rho \\
= &\sum_{\boldsymbol{x},\boldsymbol{y}} p_{X^n}(\boldsymbol{x}) \cdot \left[p_{Y^n|X^n}(\boldsymbol{y}|\boldsymbol{x})\right]^{1-\rho s} \cdot M^\rho \cdot \left[\sum_{\tilde{\boldsymbol{x}}\in\mathcal{X}^n} p_{X^n}(\tilde{\boldsymbol{x}}) \left(p_{Y^n|X^n}(\boldsymbol{y}|\tilde{\boldsymbol{x}})\right)^s\right]^\rho .
\end{aligned}
$$

By picking $s = 1/(1+\rho)$, above can be rewritten as

$$
\begin{aligned}
P[\cdots] &\le M^\rho \sum_{\boldsymbol{y}} \left(\sum_{\boldsymbol{x}} p_{X^n}(\boldsymbol{x}) p_{Y^n|X^n}(\boldsymbol{y}|\boldsymbol{x})^{\frac{1}{1+\rho}}\right)^{1+\rho} \\
&= M^\rho \cdot \left[\sum_y \left(\sum_x p_X(x) p_{Y|X}(y|x)^{\frac{1}{1+\rho}}\right)^{1+\rho}\right]^n \\
&\le 2^{n\rho(I(X,Y)-\frac{1}{2}\epsilon)} \cdot \left[\sum_y \left(\sum_x p_X(x) p_{Y|X}(y|x)^{\frac{1}{1+\rho}}\right)^{1+\rho}\right]^n \\
&= 2^{-n\rho(V(\rho)/\rho - I(X,Y)+\frac{1}{2}\epsilon)},
\end{aligned}
$$

where

$$
V(\rho) \triangleq -\log_2 \left\{\sum_y \left(\sum_x p_X(x) p_{Y|X}(y|x)^{\frac{1}{1+\rho}}\right)^{1+\rho}\right\}.
$$

Now, notice that V is differentiable in an open neighborhood around $\rho = 0$ and that $V(0) = 0$. Define function E in this neighborhood excluding the point $\rho = 0$ as $E(\rho) \triangleq V(\rho)/\rho$. Function E must be continuous in this deleted neighborhood, and is continuous in the neighborhood around $\rho = 0$ by extending to point 0 via limit. Namely,

$$
E(0) \triangleq \lim_{\rho\to 0} E(\rho) = \lim_{\rho\to 0} \frac{V(\rho)}{\rho} = \left.\frac{\mathrm{d}}{\mathrm{d}\rho}\right|_{\rho=0} V(\rho) = I(X,Y).
$$

Thus, we can pick some $\rho_0 > 0$ such that $V(\rho_0)/\rho_0 - I(X,Y) > -\epsilon/4$. In this case,

$$
P[\cdots] \le 2^{-n\rho_0(V(\rho_0)/\rho_0 - I(X,Y)+\frac{1}{2}\epsilon)} < 2^{-\frac{1}{4}n\rho\epsilon} \to 0
$$

as $n \to \infty$. Therefore, for n large enough, $P[\cdots] < \epsilon$.

Proof of the claim. Firstly, note that the function $x \mapsto x^\beta$ is concave. By Jensen's inequality, we have

$$
\begin{aligned}
\frac{\alpha-1}{\alpha^n-1} \cdot \sum_{j=1}^n \alpha^{j-1} \cdot \left(\frac{j-1}{\alpha} + n - j\right)^\beta &\le \left(\frac{\alpha-1}{\alpha^n-1} \cdot \sum_{j=1}^n \alpha^{j-1} \cdot \left(\frac{j-1}{\alpha} + n - j\right)\right)^\beta \\
&= \left(\frac{\alpha^{n-1}-1}{\alpha^n-1} \cdot n\right)^\beta .
\end{aligned}
$$

Since $\alpha < 1$, we further have

$$
\sum_{j=1}^n \alpha^{j-1} \cdot \left(\frac{j-1}{\alpha} + n - j\right)^\beta \le \frac{\alpha^n-1}{\alpha-1} \cdot n^\beta \cdot \left(\frac{\alpha^{n-1}-1}{\alpha^n-1}\right)^\beta \le \frac{\alpha^n-1}{\alpha-1} \cdot n^\beta .
$$

Also note that $\sum_{i=1}^{m} \alpha^{(i-1)n} = \frac{\alpha^{nm}-1}{\alpha^n-1}$. Therefore,

$$\sum_{i=1}^{m} \alpha^{(i-1)n} \cdot \sum_{j=1}^{n} \alpha^{j-1} \cdot \left(\frac{j-1}{\alpha} + n - j\right)^{\beta} \leq \frac{\alpha^{nm}-1}{\alpha^n-1} \cdot \frac{\alpha^n-1}{\alpha-1} \cdot n^{\beta} \leq \frac{n^{\beta}}{1-\alpha}.$$

□

c.) Based on the arguments in a.) and b.), show that, for any δ_1, $\delta_2 > 0$, the following rates are achievable

$$R_1 = H(X|Y) + \delta_1, \tag{8}$$
$$R_2 = H(Y) + \delta_2. \tag{9}$$

Solution: Let $\epsilon \in (0, \min\{1, \delta_1\})$ be arbitrarily picked.

To encode $\boldsymbol{y} \in \mathcal{Y}^n$: We firstly prepare the set of Y-sequences $\mathcal{T}_\epsilon^{(n)}(Y)$ and index the elements in this set by $\{1, \ldots, M_Y\}$, where $M_Y = \left|\mathcal{T}_\epsilon^{(n)}(Y)\right|$. Namely, $\mathcal{T}_\epsilon^{(n)}(Y) = \{\hat{\boldsymbol{y}}_1, \ldots, \hat{\boldsymbol{y}}_{M_Y}\}$. If $\boldsymbol{y}$ is in $\mathcal{T}_\epsilon^{(n)}(Y)$, we encode it by its index in $\mathcal{T}_\epsilon^{(n)}(Y)$; otherwise we encode it to something fixed. More precisely,

$$e_Y^{(n)} : \boldsymbol{y}^n \mapsto \begin{cases} \mathsf{index}_{\mathcal{T}_\epsilon^{(n)}(Y)}(\boldsymbol{y}) & \text{if } \boldsymbol{y} \in \mathcal{T}_\epsilon^{(n)}(Y) \\ 1 & \text{otherwise} \end{cases}.$$

By a.)iii.), for n large enough, $M_Y \leq 2^{n(H(Y)+\delta_2)} = 2^{nR_2}$. Thus, to transmit the encoded message, we need at most nR_2 bits. Upon receiving $m_2 \in \{1, \ldots, M_Y\}$, we pick the m_2-th element in $\mathcal{T}_\epsilon^{(n)}(Y)$ as the decoded message. Namely,

$$d_Y^{(n)} : m_2 \mapsto \hat{\boldsymbol{y}}_{m_2}.$$

Denoting the output by random variable $\hat{Y}^{(n)}$, we have $\hat{Y}^{(n)} = d_Y^{(n)}(e_Y^{(n)}(Y^n))$. By a.)i.), we know

$$P\left[\hat{Y}^{(n)} \neq Y^n\right] \leq P\left[Y^n \notin \mathcal{T}_\epsilon^{(n)}(Y)\right] \leq \epsilon$$

for n large enough.

To transmit X^n, we construct following *randomized* encoders and decoders based on auxiliary random variables $\{\boldsymbol{X}_{i,j}\}_{i,j}$ where $i \in \{1, \ldots, M_X\}$, $j \in \{1, \ldots, M\}$, and $\boldsymbol{X}_{i,j} \in \mathcal{X}^n$ are i.i.d. random variables for each (i, j) and have the same distribution as X^n. Here, we pick $M_X = \lfloor 2^{n(H(Y|X)+\epsilon)} \rfloor$ and $M = \left\lfloor 2^{n(I(X,Y)-\frac{1}{2}\epsilon)} \right\rfloor$.

To encode $\boldsymbol{x} \in \mathcal{X}^n$: We *try* to find the "smallest" (a, b) such that $\boldsymbol{X}_{a,b} = \boldsymbol{x}$. If such (a, b) exists, we use a as the encoded message; otherwise we encode $\boldsymbol{x}$ to something fixed. Namely, given a realization of $\{\boldsymbol{X}_{i,j}\}_{i,j}$ as $\{\tilde{\boldsymbol{x}}_{i,j}\}_{i,j}$,

$$e_X^{(n)}(\{\tilde{\boldsymbol{x}}_{i,j}\}_{i=1,\ldots,M_X; j=1,\ldots,M}) : \boldsymbol{x}^n \mapsto \begin{cases} & \exists b \text{ s.t. } \boldsymbol{x}_{a,b} = \boldsymbol{x} \\ a & \text{if} \quad \forall i < a, \forall j, \boldsymbol{x}_{i,j} \neq \boldsymbol{x} \\ & \forall j < b, \boldsymbol{x}_{a,j} \neq \boldsymbol{x} \\ 1 & \text{otherwise} \end{cases}.$$

Since $\epsilon \leq \delta_1$, we have $M_X \leq 2^{nR_1}$, i.e., transmitting above encoded message requires at most nR_1 bits. Upon receiving $m_1 \in \{1, \ldots, M_X\}$ *and* $m_2 \in \{1, \ldots, M_Y\}$, we pick one of the M sequences in $\{\boldsymbol{X}_{m_1,j}\}_{j=1}^{M}$ that maximizes $p_{Y^n|X^n}(\hat{\boldsymbol{y}}|\boldsymbol{X}_{m_1,j})$ as the decoded message, where $\hat{\boldsymbol{y}}$ is the decoded message for $\boldsymbol{y}$ from m_2. Nemely,

$$d_X^{(n)}(\{\tilde{\boldsymbol{x}}_{i,j}\}_{i=1,\ldots,M_X; j=1,\ldots,M}) : (m_1, m_2) \mapsto \underset{\tilde{\boldsymbol{x}} \in \{\tilde{\boldsymbol{x}}_{m_1,j} | j=1,\ldots,M\}}{\operatorname{argmax}} p_{Y^n|X^n}(d_Y^{(n)}(m_2)|\tilde{\boldsymbol{x}})$$

We denote the output by random variable $\hat{X}^{(n)}$. Conditioning on $\hat{Y}^{(n)} = Y^n$, it is clear that $\hat{X}^{(n)} = X^n$ as long as the encoder managed to find some $\boldsymbol{X}_{M_1,\tilde{M}_1}$ equal to X^n, and $p_{Y^n|X^n}(Y^n|X^n) > p_{Y^n|X^n}(Y^n|\boldsymbol{X}_{M_2,j})$ for all $j \neq \tilde{M}_1$. Combining this observation with b.)i.) and b.)ii.), we have, for n large enough,

$$\begin{aligned}
&P[\hat{X}^{(n)} \neq X^n|\hat{Y}^{(n)} = Y^n] \\
\leq &P\left[X^n \neq \boldsymbol{X}_{i,j}\ \forall (i,j)\right] + P\left[M_1, \tilde{M}_1 < \infty, p_{Y^n|X^n}(Y^n|\boldsymbol{X}_{M_1,j}) \geq p_{Y^n|X^n}(Y^n|X^n)\ \exists j \neq \tilde{M}_1\right] \\
< &3\epsilon
\end{aligned}$$

In summary, we have

$$P\left[\hat{X}^{(n)} \neq X^n \text{ or } \hat{Y}^{(n)} \neq Y^n\right] = P\left[\hat{Y}^{(n)} \neq Y^n\right] + P\left[\hat{X}^{(n)} \neq X^n \middle| \hat{Y}^{(n)} = Y^n\right] < 4\epsilon$$

for n large enough. Notice that ϵ can be picked to be arbitrarily small, we have shown (R_1, R_2) to be achievable. □

d.) Show that any (R_1, R_2) satisfying the following inequalities are achievable

$$R_1 > H(X|Y), \tag{10}$$
$$R_2 > H(Y|X), \tag{11}$$
$$R_1 + R_2 > H(X,Y). \tag{12}$$

Solution: *We sketch the proof here and omit the technical details.*
By symmetry, we know both $(H(X|Y)+\delta_1, H(Y)+\delta_2)$ and $(H(Y)+\delta_3, H(Y|X)+\delta_4)$ are achievable for all δ_1, δ_2, δ_3, $\delta_4 > 0$. By time multiplexing, all *rational* convex combinations of two achievable rates are achievable. Suppose (R_1, R_2) is a point in the domain described by (10), (11) and (12). These must exists some δ_1, δ_2, δ_3, $\delta_4 > 0$ such that (R_1, R_2) is some convex combination of $(H(X|Y)+\delta_1, H(Y)+\delta_2)$ and $(H(Y)+\delta_3, H(Y|X)+\delta_4)$. If such convex combination is rational, (R_1, R_2) is achievable. Otherwise, we claim that these must exist a point (R_1', R_2') as a rational covvex combination of $(H(X|Y)+\delta_1/2, H(Y)+\delta_2/2)$ and $(H(Y)+\delta_3/2, H(Y|X)+\delta_4/2)$ such that $R_1' < R_1$ and $R_2' < R_2$. Since (R_1', R_2') is achievable, so is (R_1, R_2). □