ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

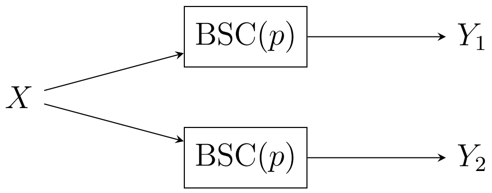
School of Computer and Communication Sciences

Handout 35 — Information Theory and Coding

Solutions to Final exam — Jan. 25, 2024

PROBLEM 1. (10 points)

Consider the following binary input channel: the input X is passed through two identical independent BSC's with crossover probability p to produce binary outputs Y_1 and Y_2, as shown in the figure below. The channel's output is $Y = (Y_1, Y_2)$.



(a) (2 points) What is the capacity achieving input distribution?

Solution: The capacity of the channel is given by $\max_{p_X} I(X; Y_1, Y_2)$. Since $I(X; Y_1, Y_2)$ is concave in p_X, the capacity is achieved by $p_X(0) = p_X(1) = 1/2$.

(b) (4 points) What is the capacity C_1 of this channel (from X to (Y_1, Y_2))?

Solution: We simply compute $I(X; Y_1, Y_2)$ (with $p_X(0) = p_X(1) = 1/2$ and $p_{Y_1,Y_2|X} = p_{Y_1|X} p_{Y_2|X}$) as

$$I(X; Y_1, Y_2) = H(Y_1, Y_2) - H(Y_1, Y_2 \mid X).$$

Since Y_1, Y_2 are conditionally independent given X, $H(Y_1, Y_2 \mid X) = H(Y_1 \mid X) + H(Y_2 \mid X) = 2h_2(p)$. The distribution of (Y_1, Y_2) when X is uniform is given by

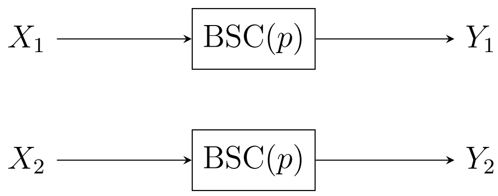
$$\begin{aligned}
p_{Y_1,Y_2}(y_1, y_2) &= \begin{cases} \frac{1}{2}[(1-p)^2 + p^2] & (y_1, y_2) = (0,0) \\ \frac{1}{2}[(1-p)p + p(1-p)] & (y_1, y_2) = (0,1) \\ \frac{1}{2}[p(1-p) + (1-p)p] & (y_1, y_2) = (1,0) \\ \frac{1}{2}[p^2 + (1-p)^2] & (y_1, y_2) = (1,1) \end{cases} \\
&= \begin{cases} \frac{1}{2} - p + p^2 & (y_1, y_2) = (0,0) \\ p(1-p) & (y_1, y_2) = (0,1) \\ p(1-p) & (y_1, y_2) = (1,0) \\ \frac{1}{2} - p + p^2 & (y_1, y_2) = (1,1). \end{cases}
\end{aligned}$$

Hence, we have

$$\begin{aligned}
H(Y_1, Y_2) &= -2\left(\frac{1}{2} - p + p^2\right)\log\left(\frac{1}{2} - p + p^2\right) - 2p(1-p)\log p(1-p) \\
&= -2\left(\frac{1}{2} - p + p^2\right)\log\left(\frac{1 - 2p + 2p^2}{2}\right) - 2p(1-p)\log\frac{2p(1-p)}{2} \\
&= (-1 + 2p - 2p^2)\log(1 - 2p + p^2) - (-1 + 2p - 2p^2) \\
&\qquad - 2p(1-p)\log[2p(1-p)] + 2p(1-p) \\
&= 1 + h_2\left(2p(1-p)\right).
\end{aligned}$$

Putting them together, we have $C_1 = 1 + h_2\left(2p(1-p)\right) - h_2(p)$.

Consider another channel whose input is (X_1, X_2) with binary X_1 and X_2. X_1 is passed through a BSC(p) to produce Y_1; X_2 is passed through an independent BSC(p) to produce Y_2, as shown in the figure. The channel's output is again $Y = (Y_1, Y_2)$.



(c) (2 points) What is the capacity C_2 of this channel (from (X_1, X_2) to (Y_1, Y_2))?

Solution: The capacity C_2 is again achieved by (X_1, X_2) having a uniform distribution on $\{0, 1\}^2$ (i.e, X_1 and X_2 are both uniform on $\{0, 1\}$ and they are independent). Then we have

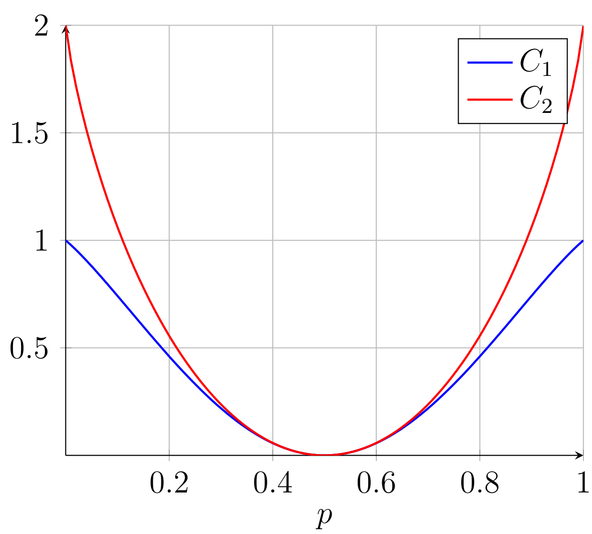
$$\begin{aligned} C_2 = I(X_1, X_2; Y_1, Y_2) &= H(Y_1, Y_2) - H(Y_1, Y_2 \mid X_1, X_2) \\ &= H(Y_1) + H(Y_2) - H(Y_1 \mid X_1) - H(Y_2 \mid X_2) \\ &= I(X_1; Y_1) + I(X_2; Y_2) \\ &= 2 - 2h_2(p), \end{aligned}$$

where we use the facts that Y_1 and Y_2 are independent and that Y_1 and X_2 are independent (and so are Y_2 and X_1) in the second line.

(d) (2 points) How do C_1 and C_2 compare?

Solution: Clearly, $C_2 \geq C_1$, since $h_2(2p(1-p)) \leq 1$, and the inequality is strict unless $p = \frac{1}{2}$ (in which case $C_1 = C_2 = 0$ — otherwise, we have $C_2 > C_1$).

Remark: Even before making any computations, it is clear that $C_2 \geq C_1$ since C_1 can be thought of as a special case of C_2 except that X_1 and X_2 are forced to be equal. We might even expect that $C_2 > C_1$ in general because of the following reason: $I(X; Y_1, Y_2) \leq H(X) \leq 1$ in the first case, while C_2 is twice the capacity of the BSC(p), can be more than 1 when p is such that $h_2(p) < \frac{1}{2}$ (i.e., $p < 0.11$ or $p > 0.89$). By computing the capacities explicitly in (b) and (c), we see that this intuition is true — in fact, $C_2 > C_1$ for all values of p except 0. The curves are plotted below.



PROBLEM 2. (12 points)

Consider a sequence of binary block codes $C_1, C_2, \ldots$ constructed as follows (with $[x, y]$ denoting the concatenation of x and y and $\mathbf{1}$ denoting the all-1 codeword $1 \ldots 1$ of appropriate length):

- $C_1 = \{0, 1\}^2 = \{00, 01, 10, 11\}$,
- $C_{k+1} = \{[u, u] : u \in C_k\} \cup \{[u + \mathbf{1}, u] : u \in C_k\}$, for $k \geq 2$.

Note that C_k is a code of blocklength 2^k.

(a) (2 points) Show that C_k contains the all-1 codeword.
Hint: Use induction.

Solution: First note that $C_1 = \{0, 1\}^2$ contains the all-1 codeword 11. Now suppose that C_{k-1} contains the all-1 codeword for some $k \geq 2$. Then, since C_k contains the element $[u, u]$ for all $u \in C_{k-1}$, by choosing $u = \mathbf{1}$, we have that C_k contains $[\mathbf{1}, \mathbf{1}] = \mathbf{1}$, and we are done.

(b) (2 points) Show that C_k is linear.
Hint: Use induction.

Solution: First note that $C_1 = \{0, 1\}^2$ is linear (this is the set of all possible binary codewords of length 2, the sum of any two length-2 binary codewords is still of length 2). Now suppose that C_{k-1} is linear for some $k \geq 2$, i.e., for any $x, y \in C_{k-1}$, $x + y$ is also in C. Note that each codeword in C_k is either of the form $[u, u]$ or $[u + \mathbf{1}, u]$ for some $u \in C_{k-1}$. We show that C_k is linear by showing that for all possible combinations of $x, y \in C_k$, we still have $x + y \in C_k$, as follows.

- $x = [u, u]$, $y = [u', u']$ for some $u, u' \in C_{k-1}$: then $x + y = [u + u', u + u'] = [u'', u'']$, where $u'' = u + u' \in C_{k-1}$, by linearity of C_{k-1};
- $x = [u, u]$, $y = [u' + \mathbf{1}, u']$ for some $u, u' \in C_{k-1}$: then $x + y = [u + u' + \mathbf{1}, u + u'] = [u'' + \mathbf{1}, u'']$, where $u'' = u + u' \in C_{k-1}$, by linearity of C_{k-1}; and
- $x = [u + \mathbf{1}, u]$, $y = [u' + \mathbf{1}, u']$ for some $u, u' \in C_{k-1}$: then $x + y = [u + u' + \mathbf{1} + \mathbf{1}, u + u'] = [u'', u'']$, where $u'' = u + u' \in C_{k-1}$, by linearity of C_{k-1}.

(c) (2 points) Show that C_k has 2^{k+1} codewords, or equivalently, $|C_k| = 2^{k+1}$.
Hint: Use induction.

Solution: First note that $|C_1| = 4 = 2^{1+1}$. Now assume that $|C_{k-1}| = 2^k$, for some $k \geq 2$. From each $u \in C_{k-1}$ we get two distinct, unique codewords in C_k. To see that the codewords are distinct, note that $[u, u] + [u + \mathbf{1}, u] = [\mathbf{1}, 0] \neq 0$. It is also clear that $[u, u]$ or $[u + \mathbf{1}, u]$ cannot be equal to $[u', u']$ or $[u' + \mathbf{1}, u']$ unless $u = u'$. Hence, $|C_k| = 2|C_{k-1}|$, and since $|C_1| = 2^2$, we have $|C_k| = 2^{k+1}$.

The Plotkin bound says the blocklength n, number of codewords M and minimum distance d of a binary code C satisfy

$$d \leq \left\lfloor \frac{nM}{2(M-1)} \right\rfloor .$$

(d) (2 points) Show that the minimum distance of C_k, $d_{\min}(C_k)$, satisfies $d_{\min}(C_k) \leq 2^{k-1}$.
Hint: Use the Plotkin bound on C_k.

Solution: For C_k, we have $n = 2^k$ and $M = |C_k| = 2^{k+1}$. Simply computing the above expression with these values, we have

$$\begin{aligned} d_{\min} &\leq \left\lfloor \frac{2^k 2^{k+1}}{2(2^{k+1}-1)} \right\rfloor = \left\lfloor \frac{2^{2k}}{2^{k+1}-1} \right\rfloor \\ &= \left\lfloor \frac{2^{k-1}(2^{k+1}-1+1)}{2^{k+1}-1} \right\rfloor = \left\lfloor 2^{k-1} + \frac{1}{2^{k+1}-1} \right\rfloor \\ &= 2^{k-1}. \end{aligned}$$

(e) (4 points) Show that $d_{\min}(C_k) \geq 2^{k-1}$.

Hint: Show that $w_{\min}(C_k) \geq 2w_{\min}(C_{k-1})$, with $w_{\min}(C_k)$ denoting the minimum weight of the nonzero codewords in C_k.

Solution: The minimum distance of a linear code is simply given by the minimum weight of its nonzero codewords. Since C_k is linear (as shown in (b)), $d_{\min}(C_k) = w_{\min}(C_k)$. Since each codeword in C_k is of the form $[u, u]$ or $[u+\mathbf{1}, u]$ where $u, u+\mathbf{1} \in C_{k-1}$, we have that the weight of any codeword is at least

$$\min\left\{ \min_{u \in C_{k-1}} \text{weight}([u, u]), \min_{u \in C_{k-1}} \text{weight}([u+\mathbf{1}, u]) \right\} \geq 2w_{\min}(C_{k-1}).$$

Hence, we have $w_{\min}(C_k) \geq 2w_{\min}(C_{k-1})$ and (since C_k is linear) $d_{\min}(C_k) \geq 2^{k-1}$.

Remark: The Plotkin bound says that any code with a given blocklength and number of codewords cannot have too large a minimum distance — the family of codes $\{C_k\}_{k\geq 1}$ (called first-order Reed-Muller codes) exactly achieves this maximum value of minimum distance, as we show in (d) and (e). Note that the rate of the code is $\frac{k+1}{2^k}$, which becomes small very quickly as k increases. This is what allows us to have a large minimum distance.

PROBLEM 3. (16 points)

Let $\mathcal{A}_1, \ldots, \mathcal{A}_n$ be disjoint subsets of $\mathbb{Z}$ with $a_i = |\mathcal{A}_i|$, $i = 1, \ldots, n$, and let $\mathcal{B}_1, \ldots, \mathcal{B}_n$ be disjoint subsets of $\mathbb{Z}$ with $b_i = |\mathcal{B}_i|$, $i = 1, \ldots, n$. Let $\mathcal{C}_i = \mathcal{A}_i \times \mathcal{B}_i \subseteq \mathbb{Z}^2$ for $i = 1, \ldots, n$ (observe that $|\mathcal{C}_i| = a_i b_i$). Pick an index $C \in \{1, \ldots, n\}$ according to the probability distribution

$$\Pr(C = c) = \frac{a_c b_c}{\sum_{k=1}^n a_k b_k}.$$

Now, given $C = c$, pick two points (X_1, Y_1) and (X_2, Y_2) uniformly and independently from $\mathcal{C}_c = \mathcal{A}_c \times \mathcal{B}_c$, i.e.,

$$\Pr\Big((X_1, Y_1) = (x_1, y_1), (X_2, Y_2) = (x_2, y_2) \mid C = c\Big) = \frac{1}{(a_c b_c)^2} \mathbb{1}\Big\{(x_1, y_1) \in \mathcal{C}_c, (x_2, y_2) \in \mathcal{C}_c\Big\}.$$

Observe that this means that X_1, Y_1, X_2, Y_2 are pairwise conditionally independent given $C = c$.

(a) (2 points) Compute $\Pr\big((X_1, Y_1) = (x_1, y_1)\big)$ and $\Pr\big((X_2, Y_2) = (x_2, y_2)\big)$.

Solution: First, we compute the conditional distribution of (X_1, Y_1) given C.

$$\begin{aligned}
\Pr\big((X_1, Y_1) = (x_1, y_1) \mid C = c\big) &= \sum_{(x_2, y_2) \in \mathbb{Z}^2} \Pr\big((X_1, Y_1) = (x_1, y_1), (X_2, Y_2) = (x_2, y_2) \mid C = c\big) \\
&= \sum_{(x_2, y_2) \in \mathbb{Z}^2} \frac{1}{(a_c b_c)^2} \mathbb{1}\Big\{(x_1, y_1) \in \mathcal{C}_c, (x_2, y_2) \in \mathcal{C}_c\Big\} \\
&= \frac{1}{a_c b_c} \mathbb{1}\Big\{(x_1, y_1) \in \mathcal{C}_c\Big\}.
\end{aligned}$$

Hence, we have

$$\begin{aligned}
\Pr\big((X_1, Y_1) = (x_1, y_1)\big) &= \sum_{c=1}^n \Pr\big((X_1, Y_1) = (x_1, y_1) \mid C = c\big) \Pr(C = c) \\
&= \sum_{c=1}^n \frac{1}{a_c b_c} \frac{a_c b_c}{\sum_{k=1}^n a_k b_k} \mathbb{1}\Big\{(x_1, y_1) \in \mathcal{C}_c\Big\} \\
&= \frac{1}{\sum_{k=1}^n a_k b_k} \mathbb{1}\Big\{(x_1, y_1) \in \cup_{c=1}^n \mathcal{C}_c\Big\}.
\end{aligned}$$

The same computation also gives $\Pr\big((X_2, Y_2) = (x_2, y_2)\big) = \frac{1}{\sum_{k=1}^n a_k b_k} \mathbb{1}\left\{(x_2, y_2) \in \cup_{c=1}^n \mathcal{C}_c\right\}$.

(b) (2 points) Show that $H(X_1, Y_1) + H(X_2, Y_2) = 2 \log \sum_{k=1}^n a_k b_k$.

Hint: Conclude from (a) that (X_1, Y_1) and (X_2, Y_2) are uniformly distributed on $\cup_{k=1}^n \mathcal{C}_k$.

Solution: The computation in (a) showed that (X_1, Y_1) are uniformly distributed on the $\sum_{k=1}^n a_k b_k$ points in $\cup_{k=1}^n \mathcal{C}_k$. Hence, $H(X_1, Y_1) = H(X_2, Y_2) = \log \sum_{k=1}^n a_k b_k$, and the result follows.

(c) (4 points) Show that $H(X_1, Y_1) = H(X_1, Y_1, C)$ and $H(X_1, X_2) = H(X_1, X_2, C)$.

Hint: Use the chain rule.

Solution: By the chain rule, we have

$$H(X_1, Y_1, C) = H(X_1, Y_1) + H(C \mid X_1, Y_1).$$

Since the sets $\mathcal{C}_i$ are disjoint, knowing the value of (X_1, Y_1) completely determines the value of C, hence, $H(C \mid X_1, Y_1) = 0$ implying $H(X_1, Y_1, C) = H(X_1, Y_1)$. Similarly, we also have $H(C \mid X_1, X_2) = 0$, and $H(X_1, X_2, C) = H(X_1, X_2)$.

(d) (4 points) Show that $H(X_1, Y_1) + H(X_2, Y_2) = H(X_1, X_2) + H(Y_1, Y_2)$.
Hint: Use the chain rule with (c) and the conditional independence of X_1, Y_1, X_2, Y_2 given C.

Solution: By the same arguments as in (c), we also have $H(X_2, Y_2) = H(X_2, Y_2, C)$ and $H(Y_1, Y_2) = H(Y_1, Y_2, C)$. We now use the chain rule to write

$$\begin{aligned}
H(X_1, Y_1) + H(X_2, Y_2) &= H(X_1, Y_1, C) + H(X_2, Y_2, C) \\
&= H(C) + H(X_1, Y_1 \mid C) + H(C) + H(X_2, Y_2 \mid C) \\
&\overset{(*)}{=} H(C) + H(X_1 \mid C) + H(Y_1 \mid C) + H(C) + H(X_2 \mid C) + H(Y_2 \mid C) \\
&= H(C) + H(X_1 \mid C) + H(X_2 \mid C) + H(C) + H(Y_1 \mid C) + H(Y_2 \mid C) \\
&\overset{(*)}{=} H(C) + H(X_1, X_2 \mid C) + H(C) + H(X_1, Y_2 \mid C) \\
&= H(X_1, X_2, C) + H(Y_1, Y_2, C) \\
&= H(X_1, X_2) + H(Y_1, Y_2),
\end{aligned}$$

where we use the conditional independence of X_1, Y_1, X_2, Y_2 given C in the steps marked with $(*)$.

(e) (4 points) Show that $H(X_1, X_2) \le \log\left(\sum_{k=1}^n a_k^2\right)$, and then show that

$$\left(\sum_{k=1}^n a_k b_k\right)^2 \le \left(\sum_{k=1}^n a_k^2\right)\left(\sum_{k=1}^n b_k^2\right).$$

Solution: The pair (X_1, Y_1) takes values on the set $\cup_{k=1}^n \mathcal{A}_k \times \mathcal{A}_k$ (they must both belong to the same $\mathcal{A}_k$, since (x_1, y_1) and (x_2, y_2) are picked from the same $\mathcal{C}_k$), which has cardinality $\sum_{k=1}^n a_k^2$. Hence, $H(X_1, X_2) \le \log \sum_{k=1}^n a_k^2$, and similarly, $H(Y_1, Y_2) \le \log \sum_{k=1}^n b_k^2$.
Finally, from (b), (d) and the above results, we have

$$\begin{aligned}
&2\log\left(\sum_{k=1}^n a_k b_k\right) \le \log\left(\sum_{k=1}^n a_k^2\right) + \log\left(\sum_{k=1}^n b_k^2\right) \\
\implies &\left(\sum_{k=1}^n a_k b_k\right)^2 \le \left(\sum_{k=1}^n a_k^2\right)\left(\sum_{k=1}^n b_k^2\right),
\end{aligned}$$

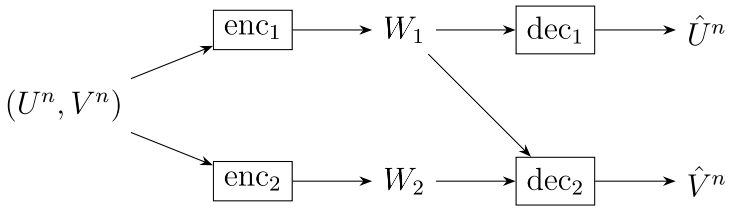
which, of course, is the well-known Cauchy-Schwarz inequality.

Remark: We have managed to derive the Cauchy-Schwarz inequality for the special case where the vectors have distinct, nonnegative, integer elements (recall that the sets $\mathcal{A}_i, \mathcal{B}_i$ are disjoint sets and a_i, b_i are the cardinalities of the sets). An extension to distinct, negative integers is immediate: the left-hand side stays the same, while the right-hand side can only decrease by making some terms negative. We could further extend this proof to the case where the vectors have distinct, rational elements by simply dividing both sides by appropriately large numbers. Continuity arguments can be used to conclude that the inequality holds for all real vectors.

One point to note is that the entropy inequalities that we use throughout are equivalent to the concavity of the logarithm, which is equivalent to the arithmetic mean–geometric mean inequality, which can in turn be derived from the Cauchy-Schwarz inequality, so this should not be considered a "proof" of the Cauchy-Schwarz inequality.

Problem 4. (16 points)

Suppose $(U_1, V_1), (U_2, V_2), \ldots$ are i.i.d. pairs with distribution p_{UV} on $\mathcal{U} \times \mathcal{V}$. An multiple descriptions system is a pair of encoder functions $\mathrm{enc}_1 : (U^n, V^n) \mapsto W_1 \in \{1, \ldots, M_1\}$ and $\mathrm{enc}_2 : (U^n, V^n) \mapsto W_2 \in \{1, \ldots, M_2\}$ with a pair of decoding functions $\mathrm{dec}_1 : W_1 \mapsto \hat{U}^n$ and $\mathrm{dec}_2 : (W_1, W_2) \mapsto \hat{V}^n$. In other words the encoder gives two descriptions W_1 and W_2; from W_1 we recover U^n, and from the pair (W_1, W_2) we recover the V^n, as shown in the figure. We define $p_e = \Pr\big((\hat{U}^n, \hat{V}^n) \neq (U^n, V^n)\big)$ as the error probability and, $R_1 = \frac{1}{n} \log M_1$, $R_2 = \frac{1}{n} \log M_2$ as the rates of the two descriptions.



(a) (2 points) Show that $R_1 \geq H(U) - p_e \log|\mathcal{U}| - \frac{1}{n} h_2(p_e)$, where $h_2(x) = -x \log x - (1-x)\log(1-x)$ is the binary entropy function.
Hint: $nR_1 \geq H(W_1) \geq I(U^n; W_1)$; and Fano's inequality upper bounds $H(U^n|W_1)$.

Solution: Let $p_{e,U} = \Pr(\hat{U}^n \neq U^n)$, clearly $p_{e,U} \leq p_e$. Following the hint, we have

$$\begin{aligned} R_1 &\geq \frac{1}{n} H(W_1) \geq \frac{1}{n} I(U^n; W_1) \geq \frac{1}{n} I(U^n; \hat{U}^n) \\ &= \frac{1}{n} H(U^n) - \frac{1}{n} H(U^n \mid \hat{U}^n) = H(U) - \frac{1}{n} H(U^n \mid \hat{U}^n) \end{aligned}$$

where the last inequality follows from the data processing inequality. By Fano's inequality, we have

$$\begin{aligned} H(U^n \mid \hat{U}^n) &\leq h_2(p_{e,U}) + p_{e,U} \log|\mathcal{U}^n| \\ \implies \frac{1}{n} H(U^n \mid \hat{U}^n) &\leq \frac{1}{n} h_2(p_{e,U}) + p_{e,U} \log|\mathcal{U}|. \end{aligned}$$

Consider the function $x \mapsto \frac{1}{n} h_2(x) + x \log|\mathcal{U}|$. By taking the derivative, we see that this is increasing for all $x \leq 1 - \frac{1}{|\mathcal{U}|^n + 1}$. Hence, for $p_e \leq 1 - \frac{1}{|\mathcal{U}|^n+1}$, we have that the right-hand side above is upper bounded by $\frac{1}{n} h_2(p_e) + p_e \log|\mathcal{U}|$, and this completes the proof (for $p_e \leq 1 - \frac{1}{|\mathcal{U}|^n+1}$, which goes to 1 very quickly as n and $|\mathcal{U}|$ increase — it is possible that the statement may not be true for $p_e > 1 - \frac{1}{|\mathcal{U}|^n+1}$, but in the later parts, we are interested in regions which have a small error probability, so this does not matter).

(b) (2 points) Show that $R_1 + R_2 \geq H(UV) - p_e \log|\mathcal{U}||\mathcal{V}| - \frac{1}{n} h_2(p_e)$.
Hint: Similar to (a).

Solution: Similar to (a), we have

$$\begin{aligned} R_1 + R_2 &\geq \frac{1}{n} H(W_1, W_2) \geq \frac{1}{n} I(U^n, V^n; W_1, W_2) \geq \frac{1}{n} I(U^n, V^n; \hat{U}^n, \hat{V}^n) \\ &= \frac{1}{n} H(U^n, V^n) - \frac{1}{n} H(U^n, V^n \mid \hat{U}^n, \hat{V}^n) = H(U, V) - \frac{1}{n} H(U^n, V^n \mid \hat{U}^n, \hat{V}^n) \end{aligned}$$

where the last inequality follows from the data processing inequality. By Fano's inequality, we have

$$\begin{aligned} H(U^n, V^n \mid \hat{U}^n, \hat{V}^n) &\le h_2(p_e) + p_e \log |\mathcal{U}^n||\mathcal{V}^n| \\ \implies \frac{1}{n} H(U^n, V^n \mid \hat{U}^n, \hat{V}^n) &\le \frac{1}{n} h_2(p_e) + p_e \log |\mathcal{U}||\mathcal{V}|, \end{aligned}$$

and this, with the above, completes the proof (for completeness, note that unlike (a), this is true for all values of p_e).

Let $\mathcal{C}$ be the set of (r_1, r_2) pairs for which for any $\epsilon > 0$ there is $\mathrm{enc}_1, \mathrm{enc}_2, \mathrm{dec}_1, \mathrm{dec}_2$ with $p_e < \epsilon$, $R_1 < r_1 + \epsilon$, and $R_2 < r_2 + \epsilon$.

(c) (2 points) Show that $\mathcal{C}$ is included in the region

$$\mathcal{R} = \{(r_1, r_2) : r_1 \ge H(U), r_2 \ge 0, r_1 + r_2 \ge H(UV)\}.$$

Solution: From parts (a) and (b), if $p_e < \epsilon'$ (for $\epsilon' \le 1/2$) $R_1 \ge H(U) - \epsilon' \log |\mathcal{U}| - \frac{1}{n} h_2(\epsilon')$ and $R_1 + R_2 \ge H(UV) - \epsilon' \log |\mathcal{U}||\mathcal{V}| - \frac{1}{n} h_2(\epsilon')$. Hence, given any $\epsilon > 0$, pick ϵ' such that

$$\epsilon > \max \left\{ \epsilon', \epsilon' \log |\mathcal{U}| + \frac{1}{n} h_2(\epsilon'), \epsilon' \log |\mathcal{U}||\mathcal{V}| + \frac{1}{n} h_2(\epsilon') \right\}$$

(such a choice is possible for any $\epsilon > 0$ because all these quantities go to zero as $\epsilon' \to 0$). Then, we have $p_e < \epsilon$, $R_1 \ge H(U) - \epsilon$, and $R_1 + R_2 \ge H(UV) - \epsilon$. For $r_1 + \epsilon > R_1$, we must have $r_1 > H(U) - 2\epsilon$ and for $r_2 + \epsilon > R_2 \ge 0$, we must have $r_1 + r_2 > R_1 + R_2 - 2\epsilon \ge H(UV) - 3\epsilon$. Since this must hold for any $\epsilon > 0$, taking $\epsilon \to 0$, we have that for any $(r_1, r_2) \in \mathcal{C}$, we must have $r_1 \ge H(U)$, $r_2 \ge 0$ and $r_1 + r_2 \ge H(UV)$, i.e., $\mathcal{C} \subseteq \mathcal{R}$.

[For parts (d) and (e), assume that the Huffman code described has expected length equal to the entropy exactly.]

(d) (2 points) Suppose we design a Huffman code c for the pair (U, V). Let W be the concatenation of $c(U_1, V_1), \ldots, c(U_n, V_n)$. Show that for $r > H(UV)$, $\lim_{n\to\infty} \Pr(\mathrm{length}(W) > nr) = 0$, and conclude that $(r_1 = H(UV), r_2 = 0)$ belongs to $\mathcal{C}$.

Hint: Use the law of large numbers with $\frac{1}{n} \mathrm{length}(W) = \frac{1}{n} \sum_{i=1}^n \mathrm{length}(c(U_i, V_i))$.

Solution: We compute the probability as follows:

$$\begin{aligned} \Pr(\mathrm{length}(W) > nr) &= \Pr\left(\frac{1}{n} \mathrm{length}(W) > r\right) \\ &= \Pr\left(\frac{1}{n} \mathrm{length}(W) - H(UV) > r - H(UV)\right) \to 0, \end{aligned}$$

by the law of large numbers, since $\frac{1}{n} \mathrm{length}(W) = \frac{1}{n} \sum_{i=1}^n \mathrm{length}(c(U_i, V_i)) \to \mathbb{E}[\mathrm{length}(c(U, V))] = H(UV)$, as given. The decoder dec_1 can thus losslessly recover U^n from $W_1 = W$, while dec_2 does nothing and $W_2 = 0$ always. Hence, the rate pair $(H(UV), 0)$ is achieved.

(e) (4 points) Show that $(r_1 = H(U), r_2 = H(V|U))$ belongs to $\mathcal{C}$.

Hint: Follow a similar logic to (d).

Solution: As in (d), we design Huffman codes (again assuming that they achieve entropy as the expected length). First, we design a Huffman code c_1 for the distribution p_U. Then, for each $u \in \mathcal{U}$, we design Huffman codes $c_{2,u}$ for $p_{V|U=u}$. We then form W_1 and W_2 by concatenating $c_1(U_1), c_1(U_2), \ldots$ and $c_{2,U_1}(V_1), c_{2,U_2}(V_2), \ldots$ respectively. By the same reasoning as in (d), for $r_1 > H(U)$ and $r_2 > H(V \mid U)$, we have $\Pr(\text{length}(W_1) > nr_1)$ and $\Pr(\text{length}(W_2) > nr_2)$ both go to zero as $n \to \infty$. The decoders dec_1 and dec_2 can losslessly recover U^n from W_1 and dec_2 can losslessly recover V^n from W_2 (by using the value of U^n that it recovered to identify the correct codebook to be used for decoding). Hence, the rate pair $(H(U), H(V \mid U))$ can be achieved.

(f) (4 points) Show that the region $\mathcal{R}$ is included in $\mathcal{C}$ (and thus, because of (c), $\mathcal{R} = \mathcal{C}$).

Hint: In (d) and (e) you have shown that the extreme points of $\mathcal{R}$ are in $\mathcal{C}$, now use time-sharing. You may find it useful to make a sketch of $\mathcal{R}$ and the two points in (d) and (e).

Solution: Clearly, if any point (a, b) is achievable, so is (a', b') with $a' \geq a$ and $b' \geq b$. Hence, to show that $\mathcal{R}$ is included in $\mathcal{C}$, it is sufficient to show that the line joining the points $(H(UV), 0)$ and $(H(U), H(V \mid U))$ is achievable. Since we can use time-sharing, it is further sufficient to show that the endpoints of the line are achievable, which we have already done in (c) and (d).

Remark: This is a version of a "multiple-descriptions" problem, where we can choose to encode V^n either in W_1 or W_2, since dec_2 sees (W_1, W_2) before attempting to recover V^n. This gives us a trade-off between R_1 and R_2, which is captured in the region $\mathcal{C} = \mathcal{R}$.

The assumption that the Huffman code achieves as its expected length exactly the entropy rate was a simplification made to make the solution easy. The more precise way to show (d) and (e) would be to construct the Huffman codes for the pairs (U^k, V^k), let $n = k^2$ and encode k k-length blocks at a time to get (U^n, V^n). This way, we can get $H(UV) \leq \frac{1}{k}\mathbb{E}[\text{length}(c(U^k, V^k))] < H(UV) + \frac{1}{k}$ (since Huffman codes are known to achieve expected lengths within one of the entropy), and as $k \to \infty$, we have the same result as in (d).

The proof technique to show the achievability in (d), (e) and (f) is different from the usual techniques that we have seen in the course (such as random coding and typicality decoding). In particular, we see that it is enough to show that it is achieve the extreme points (via Huffman coding) and then use time-sharing to show the achievability of the other points (which implies that the region must be convex).