ECE 231A Assessment 3: Capacity, Lossy Compression, Multi-User Information Theory

Information Theory
Instructor: Rick Wesel

Posted Wednesday May 28, 2020
Due Friday June 5, 2020

This assessment is open book and open note, but you may not perform an internet search to seek a worked solution to a problem. Do not post this assessment on the internet or post questions about this assessment to an internet site that provides assistance. Do not ask for help from your classmates or anyone else besides Prof. Wesel and Hengjie.

Your Name: Shengze Ye

Your ID Number: 20541895[9]

Are you using this final as a comprehensive exam question?

Circle one: **YES** **NO** ✓

Problem	Score	Possible
1		10
2		12
3		14
4		12
5		10
6		10
7		12
Total		80

1. (10 pts) (warm-up) *The Generalized Grouping Axiom for Three Coin Tosses.*

This exercise applies the generalized grouping axiom to three coin tosses where p_1, p_2, p_3 are the probabilities of heads for the three coins respectively. Define $\bar{p}_i = 1 - p_i$. Use the grouping axiom (perhaps multiple times) to show that

$$H(p_1)+H(p_2)+H(p_3) = H\left(p_1p_2p_3, p_1p_2\bar{p}_3, p_1\bar{p}_2p_3, p_1\bar{p}_2\bar{p}_3, \bar{p}_1p_2p_3, \bar{p}_1p_2\bar{p}_3, \bar{p}_1\bar{p}_2p_3, \bar{p}_1\bar{p}_2\bar{p}_3\right).$$

Recall the generalized grouping axiom:

$$\begin{aligned} H(p_1, p_2, \ldots, p_m) &= H(p_1 + p_2 + \ldots + p_k, p_{k+1}, \ldots, p_m) \\ &\quad + (p_1 + p_2 + \ldots + p_k) H\left(\frac{p_1}{p_1 + p_2 + \ldots + p_k}, \ldots, \frac{p_k}{p_1 + p_2 + \ldots + p_k}\right). \end{aligned}$$

$$H(p_1p_2p_3 \cdot p_1p_2\bar{p}_3, p_1\bar{p}_2p_3, p_1\bar{p}_2\bar{p}_3, \bar{p}_1p_2p_3, \bar{p}_1p_2\bar{p}_3, \bar{p}_1\bar{p}_2p_3, \bar{p}_1\bar{p}_2\bar{p}_3)$$

$$= H(p_1p_2(p_3+\bar{p}_3). p_1\bar{p}_2(p_3+\bar{p}_3), \bar{p}_1p_2(p_3+\bar{p}_3), \bar{p}_1\bar{p}_2(p_3+\bar{p}_3))$$

$$+ p_1p_2 \cdot H(p_3. \bar{p}_3) + p_1\bar{p}_2 H(p_3, \bar{p}_3) + \bar{p}_1p_2 H(p_3.\bar{p}_3) + \bar{p}_1\bar{p}_2 H(p_3.\bar{p}_3)$$

$$= H(p_1p_2, p_1\bar{p}_2, \bar{p}_1p_2, \bar{p}_1\bar{p}_2)$$

$$+ p_1p_2 H(p_3) + p_1\bar{p}_2 H(p_3) + \bar{p}_1p_2 H(p_3) + \bar{p}_1\bar{p}_2 H(p_3)$$

$$= H(p_1p_2, p_1\bar{p}_2, \bar{p}_1p_2, \bar{p}_1\bar{p}_2) + H(p_3)$$

$$= H(p_1(p_2+\bar{p}_2). \bar{p}_1(p_2+\bar{p}_2)) + H(p_3)$$

$$+ p_1 H(p_2, \bar{p}_2) + \bar{p}_1 H(p_2, \bar{p}_2)$$

$$= H(p_1, \bar{p}_1) + H(p_3) + H(p_2, \bar{p}_2)$$

$$= H(p_1) + H(p_2) + H(p_3) \quad \square$$

2. (12 pts) *Two Nearest Neighbors Channel with Erasures.*

Consider the Two Nearest Neighbors Channel with Erasures, specifically with four inputs. This channel is illustrated below.

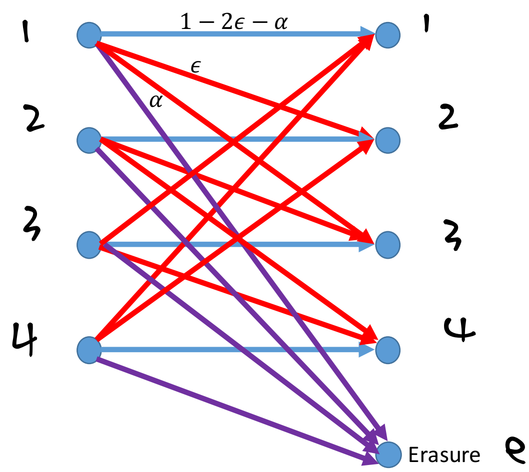


Figure 1: The Two Nearest Neighbors Channel. For a given input, the output is the same as the input with probability $1 - 2\epsilon - \alpha$ (blue arrows). For a given input, the output is the erasure symbol with probability α (purple arrows). For a given input, the output is one of two neighboring non-erasure symbols (usually the two symbols just below), each with probability ϵ (red arrows).

(a) (3 pts) Write down the transition matrix for this channel.

(b) (3 pts) Does this channel satisfy the matrix conditions for cyclic symmetry? Explain your answer.

(c) (3 pts) Does this channel satisfy the matrix conditions for weak symmetry? Explain your answer.

(d) (3 pts) Find the capacity of the Two Nearest Neighbors Channel with Erasures with four inputs. To get full credit, you must use the Grouping Axiom (See problem 1.) to simplify your answer to be a product of $1 - \alpha$ and another term.

For convenience, I note the input as 1, 2, 3, 4,
and note the output as 1, 2, 3, 4, e

(a) $$p(y|x) = \begin{bmatrix} 1-2\epsilon-\alpha & \epsilon & \epsilon & 0 & \alpha \\ 0 & 1-2\epsilon-\alpha & \epsilon & \epsilon & \alpha \\ \epsilon & 0 & 1-2\epsilon-\alpha & \epsilon & \alpha \\ \epsilon & \epsilon & 0 & 1-2\epsilon-\alpha & \alpha \end{bmatrix}$$

(b) Yes.

① The 4 rows in the transition matrix are permutations of each other.

② The set of columns can be seperated into two subsets that are collectively exhaustic, mutually exclusive, and such that each subset is exactly all cyclic shifts of any one element in the subset.

we can divide it into two subsets

$$S_1 = \left\{ \begin{bmatrix} 1-2\varepsilon-\alpha \\ 0 \\ \varepsilon \\ \varepsilon \end{bmatrix}, \begin{bmatrix} \varepsilon \\ 1-2\varepsilon-\alpha \\ 0 \\ \varepsilon \end{bmatrix}, \begin{bmatrix} \varepsilon \\ \varepsilon \\ 1-2\varepsilon-\alpha \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ \varepsilon \\ \varepsilon \\ 1-2\varepsilon-\alpha \end{bmatrix} \right\} \quad S_2 = \left\{ \begin{bmatrix} \alpha \\ \alpha \\ \alpha \\ \alpha \end{bmatrix} \right\}$$

(c) No.

A channel is weakly symmetric, then all the rows are permutations of each other and the column sums are equal.

for first 4 columns, the sum is $1-\alpha$.

for the last column, the sum is 4α $\qquad 4\alpha \neq 1-\alpha$

only when $\alpha = \frac{1}{5}$, the column sums are equal,

otherwise, it is not weakly symmetric.

(d) Since the channel has cyclic symmetry, the uniform distribution achieves capacity

$$C = \max I(X;Y) = \max [H(Y) - H(Y|X)]$$

$$= H(Y) - H(Y|X) \quad (\text{for } X \text{ is uniform})$$

$$p(Y=1) = \frac{1}{4}[1-2\varepsilon-\alpha+\varepsilon+\varepsilon] = \frac{1}{4}(1-\alpha)$$

$$p(Y=2) = p(Y=3) = p(Y=4) = p(Y=1) = \frac{1}{4}(1-\alpha)$$

$$p(Y=e) = \frac{1}{4}\cdot[\alpha+\alpha+\alpha+\alpha] = \alpha$$

$$H(Y) = H\left(\frac{1}{4}(1-\alpha), \frac{1}{4}(1-\alpha), \frac{1}{4}(1-\alpha), \frac{1}{4}(1-\alpha), \alpha\right)$$

$$H(Y|X) = \sum_{1} H(Y|X=x)\, p(X=x)$$

$$= \frac{1}{4}\cdot H(Y|X=1) + \frac{1}{4} H(Y|X=2) + \frac{1}{4} H(Y|X=3) + \frac{1}{4} H(Y|X=4)$$

$$= H(1-2\varepsilon-\alpha, \varepsilon, \varepsilon, \alpha)$$

$$\therefore C = H(Y) - H(Y|X)$$

$$= H\left(\frac{1}{4}(1-\alpha), \frac{1}{4}(1-\alpha), \frac{1}{4}(1-\alpha), \frac{1}{4}(1-\alpha), \alpha\right) - H(1-2\varepsilon-\alpha, \varepsilon, \varepsilon, \alpha)$$

$$= H(\alpha) + (1-\alpha)\cdot H\left(\frac{1}{4}, \frac{1}{4}, \frac{1}{4}, \frac{1}{4}\right) - \left[H(\alpha) + (1-\alpha) H\left(\frac{1-2\varepsilon-\alpha}{1-\alpha}, \frac{\varepsilon}{1-\alpha}, \frac{\varepsilon}{1-\alpha}\right)\right]$$

$$= H(\alpha) + 2\cdot(1-\alpha) - H(\alpha) - (1-\alpha) H\left(\frac{1-2\varepsilon-\alpha}{1-\alpha}, \frac{\varepsilon}{1-\alpha}, \frac{\varepsilon}{1-\alpha}\right)$$

$$= (1-\alpha)\left[2 - H\left(\frac{1-2\varepsilon-\alpha}{1-\alpha}, \frac{\varepsilon}{1-\alpha}, \frac{\varepsilon}{1-\alpha}\right)\right]$$

$\therefore$ The capacity is

$$C = (1-\alpha)\left[2 - H\left(\frac{1-2\varepsilon-\alpha}{1-\alpha}, \frac{\varepsilon}{1-\alpha}, \frac{\varepsilon}{1-\alpha}\right)\right]$$

3. (14 pts) *Discrete and continuous mutual information.* Consider a channel that has discrete inputs but continuous outputs. Specifically, the input X is binary with probability p that $X = 1$ and probability $1 - p$ that $X = 0$. The channel output is $Y = X + Z$ where Z is uniform over the interval $[0, a]$, $a > 1$. Z is independent of X.

(a) (5 pts) Compute the mutual information as the difference of two discrete entropies:

$$I(X;Y) = H(X) - H(X|Y)\,. \tag{1}$$

(b) (5 pts) Compute the mutual information as the difference of two differential entropies:

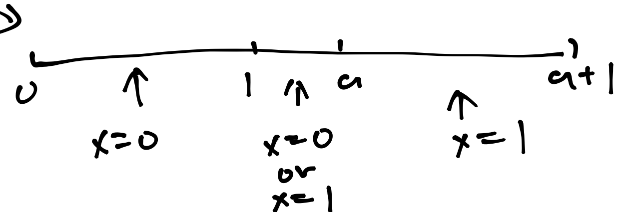
$$I(X;Y) = h(Y) - h(Y|X)\,. \tag{2}$$

(c) (4 pts) Compute the capacity of the channel by maximizing over p.

(a) $p(x=0) = 1-p \quad p(x=1) = p \quad f(z) = \frac{1}{a}$ for $0 \le z \le a$

① $\boxed{H(X) = H(p)}$

$Y = X + Z$



$$\therefore p(X=1) = \begin{cases} 0 & 0 \le y \le 1 \\ p & 1 \le y \le a \\ 1 & a \le y \le a+1 \end{cases} \qquad \therefore H(X|Y=y) = \begin{cases} 0 & 0 \le y \le 1 \\ H(p) & 1 \le y \le a \\ 0 & a \le y \le a+1 \end{cases}$$

$$f(y) = \begin{cases} \frac{1}{a}(1-p) & 0 \le y \le 1 \\ \frac{1}{a} & 1 \le y \le a \\ \frac{1}{a} \cdot p & a \le y \le a+1 \end{cases}$$

$$H(X|Y) = \int_{y=0}^{\infty} f(y) H(X|Y=y)\,dy$$

$$= \int_0^1 \frac{1}{a}(1-p) \cdot 0\,dy + \int_1^a \frac{1}{a} \cdot H(p)\,dy + \int_a^{a+1} \frac{1}{a} \cdot p \cdot 0\,dy$$

$$= 0 + \frac{a-1}{a} \cdot H(p)$$

② $\boxed{H(X|Y) = \frac{a-1}{a} \cdot H(p)}$

$$\therefore I(X;Y) = H(X) - H(X|Y)$$

$$= H(p) - \frac{a-1}{a} H(p) = \frac{1}{a} \cdot H(p)$$

$$I(X;Y) = \frac{1}{a} H(p)$$

(b) $h(Y) = -\int_0^\infty f(y)\log f(y)\,dy$

$$= -\int_0^1 \frac{1}{a}(1-p)\log \frac{1}{a}(1-p)\,dy - \int_1^a \frac{1}{a}\log\frac{1}{a}\,dy$$

$$-\int_a^{a+1} \frac{p}{a}\log\frac{p}{a}\,dy$$

$$= \frac{1-p}{a}\log\frac{a}{1-p} + \frac{a-1}{a}\log a + \frac{p}{a}\log\frac{a}{p}$$

$$= \frac{1-p}{a}[\log a - \log(1-p)] + \frac{a-1}{a}\log a + \frac{p}{a}[\log a - \log p]$$

$$= \frac{1}{a}\log a - \frac{1-p}{a}\log(1-p) + \frac{a-1}{a}\log a - \frac{p}{a}\log p$$

$$= -\frac{1-p}{a}\log(1-p) - \frac{p}{a}\log p + \log a$$

$$= \frac{1}{a}[-(1-p)\log(1-p) - p\log p] + \log a$$

$$= \frac{1}{a}H(p) + \log a$$

$$f(y|x) = \begin{cases} \frac{1}{a} & 0 \le y \le a \quad \text{if } x=0 \\ \frac{1}{a} & 1 \le y \le 1+a \quad \text{if } x=1 \end{cases}$$

$\therefore h(Y|X) = \sum_x p_X(x)\, h(Y|X=x)$

$$= p(X=0)\,h(Y|X=0) + p(X=1)\,h(Y|X=1)$$

$$= (1-p)\cdot\log a + p\cdot\log a = \log a$$

$I(X;Y) = h(Y) - h(Y|X)$

$$= \frac{1}{a}H(p) + \log a - \log a = \frac{1}{a}H(p)$$

$\therefore I(X;Y) = \frac{1}{a}H(p)$

(c) $C = \max I(X;Y) = \max_p \frac{1}{a}H(p)$

$H(p) \le 1$ when $p = \frac{1}{2}$ $H(p) = 1$

$\therefore C = \max_p \frac{1}{a}H(p) = \frac{1}{a}$ bits

$\therefore$ The capacity of channel is $\frac{1}{a}$ bits

4. (12 pts) *More Parallel Gaussian Channels.*
A transmitter has a total amount of energy P to distribute over five parallel Gaussian channels. That is, $P_1 + P_2 + P_3 + P_4 + P_5 = P$. These five Gaussian channels have, respectively, the noise variances $N_1 = 2$, $N_2 = 2$, $N_3 = 2$, $N_4 = 2$, and $N_5 = 4$.

(a) (4 pts) Assuming an optimal distribution of power, find the value of P where the number of active parallel channels transitions from 4 to 5. Compute the channel capacity at this point.

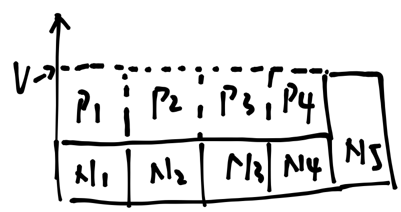
(b) (4 pts) For values of P larger than the transition point, (i.e. when all five of the parallel sub-channels are active), find an expression for P as a function of the water level v.

(c) (4 pts) Assuming that P is large enough that all five channels are active, and using the variable v (rather than P) in the expression, show that $C(v)$, the capacity, which is the mutual information of the optimal power distribution at water level v has the form

$$C(v) = \frac{1}{2}\log\left(\frac{v^{\alpha}}{2^{\beta}}\right), \tag{3}$$

and find α and β.

(a)



At this point

$P = P_1 + P_2 + P_3 + P_4 = 2 \times 4 = 8$

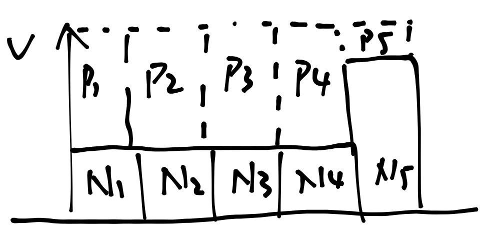
$C = \sum_{i=1}^{4} \frac{1}{2}\log\left(1 + \frac{P_i}{N_i}\right)$

$= 4 \cdot \frac{1}{2}\log\left(1 + \frac{2}{2}\right) = 2$

The fifth channel is not activate

$\therefore$ At this point, $P = 8$ $C = 2$

(b)



when P is larger than the transition point, all the sub-channels are active

$V = P_i + N_i$ $P_i = V - N_i$

$\therefore P = P_1 + P_2 + P_3 + P_4 + P_5$

$P = 5V - N_1 - N_2 - N_3 - N_4 - N_5$

$= 5V - 12$

$\therefore P = 5V - 12$

(c) We first consider a single sub-channel with P_i. N_i

$$V = P_i + N_i \quad \therefore P_i = V - N_i$$

$$C_i = \frac{1}{2}\log\left(1+\frac{P_i}{N_i}\right) = \frac{1}{2}\log\left(\frac{N_i}{N_i} + \frac{V-N_i}{N_i}\right)$$

$$= \frac{1}{2}\log\left(\frac{V}{N_i}\right)$$

The total capacity is the sum of C_i, and as the $N_1 = N_2 = N_3 = N_4 = 2^1$ $N_5 = 2^2$,

$$\therefore C = \sum_{i=1}^{5} \frac{1}{2}\log\left(\frac{V}{N_i}\right) = \frac{1}{2}\log\left(\frac{V}{2}\right)\times 4 + \frac{1}{2}\log\left(\frac{V}{4}\right)$$

$$= \frac{1}{2}\log\left(\frac{V^4}{2^4}\right) + \frac{1}{2}\log\left(\frac{V^1}{2^2}\right)$$

$$= \frac{1}{2}\log\left(\frac{V^5}{2^6}\right)$$

Therefore,

$$C(V) = \frac{1}{2}\log\left(\frac{V^5}{2^6}\right)$$

And $\alpha = 5$, $\beta = 6$

5. (10 pts) *Lossy Compression with a Troublesome Reconstruction Alphabet*
Consider a lossy compression problem in which the source X is binary and equally likely to be zero or one. The reconstruction symbols $\hat{X}$ belong to the set $\{0, 1, 2\}$. The distortion matrix is given below:

$$d(x, \hat{x}) = \begin{bmatrix} 0 & 1 & 2 \\ 1 & 0 & 2 \end{bmatrix}. \tag{4}$$

For example, going along the first row we have $d(x = 0, \hat{x} = 0) = 0$, $d(x = 0, \hat{x} = 1) = 1$, and $d(x = 0, \hat{x} = 2) = 2$.

Find the equation describing the optimal rate vs. distortion curve for this situation.

for $d(x,\hat{x}) = \begin{bmatrix} 0 & 1 & 2 \\ 1 & 0 & 2 \end{bmatrix}$

$E[d] = \sum p(x) p(\hat{x}|x) \cdot d(x,\hat{x}) \le D$

$\therefore p(x=0,\hat{x}=1) + p(x=1,\hat{x}=0) + 2p(\hat{x}=2) \le D$

we want the distortion as small as possible and the rate as small as possible. so we should set $p(\hat{x}=2, x=0) = p(\hat{x}=2, x=1) = 0$

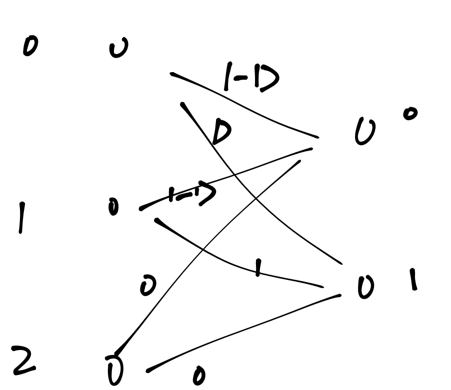
$R(D) = \min_{p(\hat{x}|x),\, E[d] \le D} I(X;\hat{X})$

$I(X;\hat{X}) = H(X) - H(X|\hat{X})$ X is binary, equally likely to be 0 and 1

$\therefore H(X) = H(\frac{1}{2}, \frac{1}{2}) = 1$

Fano → $H(X|\hat{X}) \le H(D) + D\log(2-1) = H(D)$

The test channel should be



we can find the input distribution $p(\hat{x})$ to make it work

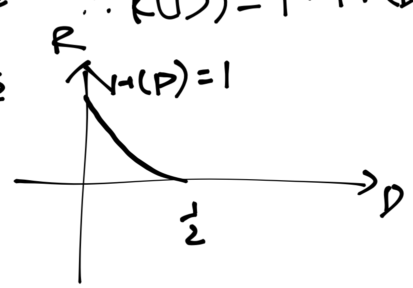
$p(x=0) = (1-D)p(\hat{x}=0) + D\,p(\hat{x}=1) = \frac{1}{2}$

$p(x=1) = (1-D)p(\hat{x}=1) + D\,p(\hat{x}=0) = \frac{1}{2}$

$p(\hat{x}=0) = \dfrac{\frac{1}{2}-D}{1-2D} = \frac{1}{2}$

$p(\hat{x}=1) = \dfrac{\frac{1}{2}-D}{1-2D} = \frac{1}{2}$

$\therefore R(D) = 1 - H(D)$



6. (10 pts) *Multiple Access on an Adder Channel.*

Consider the two-user noiseless modulo addition channel

$$Y = X_1 + X_2 \mod 3$$

with the following alphabets for X_1, X_2, and Y:

$$\mathcal{X}_1 = \{0,1\},$$
$$\mathcal{X}_2 = \{1,2,3\},$$
$$\mathcal{Y} = \{0,1,2\}.$$

Note that all three alphabets above are different. Define $p = P(X_1 = 1)$. Note also that the addition is modulo-3 so that, for example, $3+0=0$, $1+2=0$, and $3+1=1$.

Find AND DRAW the achievable rate region for this multiple access channel.

To find the (R_1, R_2) pair, we need that

$R_1 < I(X_1; Y|X_2) = H(Y|X_2) - H(Y|X_1, X_2)$

Since if we know $X_1, X_2 \rightarrow$ we know Y $\therefore H(Y|X_1,X_2) = 0$

$R_1 < I(X; Y|X_2) = H(Y|X_2) = H(X_1)$

$R_2 < I(X_2; Y|X_1) = H(Y|X_1) - H(Y|X_1, X_2)$

$= H(Y|X_1)$

$= H(X_2)$

$R_1 + R_2 < I(X_1, X_2; Y) = H(Y) - H(Y|X_1, X_2)$

$= H(Y)$

$\therefore R_1 < H(X_1) \leq 1$ (when X_1, X_2 are uniform distribution,

$R_2 < H(X_2) \leq \log 3$ $H(X_1), H(X_2)$ reach its maximum)

$P(Y=0) = P(X_1=0) \cdot P(X_2=3) + P(X_1=1) \cdot P(X_2=2) = \frac{1}{3}$

$P(Y=1) = P(X_1=0) \cdot P(X_2=1) + P(X_1=1) \cdot P(X_2=3) = \frac{1}{3}$

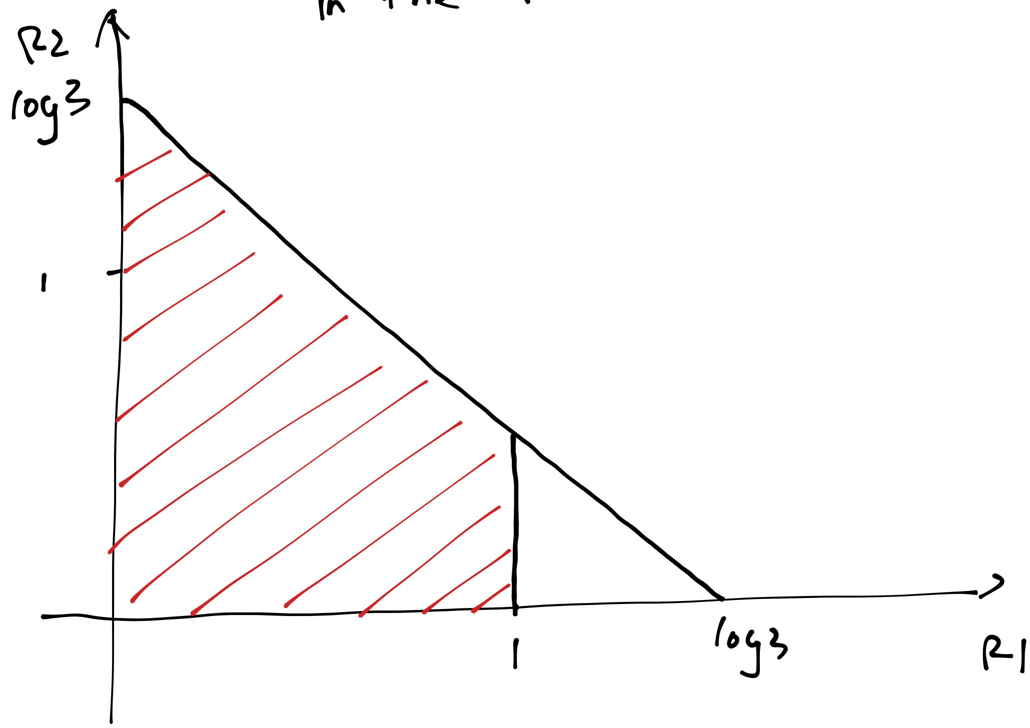
$P(Y=2) = P(X_1=0) P(X_2=2) + P(X_1=1) P(X_2=1) = \frac{1}{3}$

$\therefore$ when X_1, X_2 are uniform, Y is uniform

$R_1+R_2 < H(X) \leq \log 3$

$\therefore \; R_1 \leq 1 \quad R_2 \leq \log 3 \quad R_3 \leq \log 3$

The achievable region is shown below in the red color



7. (12 pts) *Find a Slepian-Wolf Region.*

Two sensor nodes each observe a noisy version of the data bit $D \in \{0, 1\}$. D is Bernoulli with $P(D = 1) = 0.5$.

The observation S_i of the i^{th} sensor node is described by $S_i = D \oplus N_i$ where N_i is a bit of binary noise that has a Bernoulli PMF with $P(N_i = 1) = p$. The operation $\oplus$ is exclusive-or; if $N_i = 0$ then $S_i = D$, and if $N_i = 1$ then S_i is $\bar{D}$, the complement of D. Furthermore, N_1 and N_2 are independent. Essentially, the data bit D is transmitted through two independent BSC's, each with probability of error p.

The two sensors need to communicate their noisy observations to a central decision-making node. The goal of this problem is to find the Slepian-Wolf region of rates sufficient for the two nodes to communicate their information to the central decision-making node.

(a) (4 pts) Find the probabilities $P(D = 0|S_1 = 0)$ and $P(D = 1|S_1 = 0)$.

(b) (4 pts) Show that the probability $P(S_2 = 0|S_1 = 0) = 1 - 2p(1 - p)$. To do this, note that

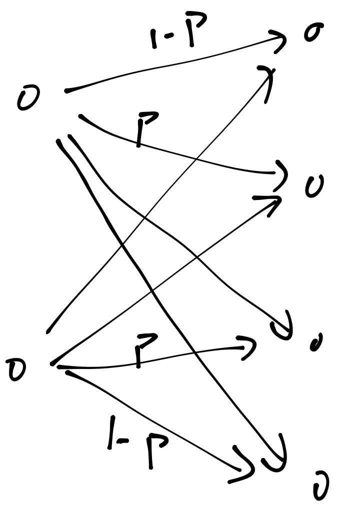
$$P(S_2 = 0|S_1 = 0) = \sum_{d=0}^{1} P(S_2 = 0, D = d|S_1 = 0)$$

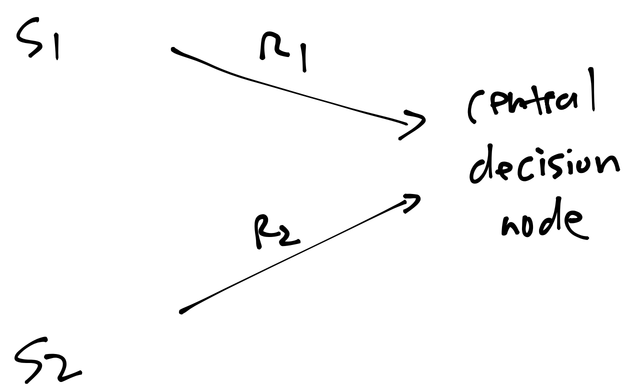
and then decompose each term in the summation as

$$P(S_2 = 0, D = d|S_1 = 0) = P(D = d|S_1 = 0)P(S_2 = 0|D = d, S_1 = 0)\,.$$

(c) (4 pts) Find the Slepian Wolf region of achievable rates for the two sensor nodes communicating their information to the central decision-making node.

(a)





$D \in \{0,1\}$ $\qquad P(D=1) = 0.5 \qquad \therefore P(D=0) = 0.5$

$S_i = D \oplus N_i \qquad \therefore P(S_1=0) = P(D=0) \cdot P(N_1=0) + P(D=1) \cdot P(N_1=1)$

$= 0.5 \cdot (1-p) + 0.5 \cdot (p) = 0.5$

$$P(D=0 \mid S_1=0) = \frac{P(D=0, S_1=0)}{P(S_1=0)} = \frac{P(S_1=0 \mid D=0)\, P(D=0)}{P(S_1=0)}$$

$$= \frac{(1-p)\cdot 0.5}{(1-p)\cdot 0.5 + p\cdot 0.5} = 1-p$$

$$P(D=1 \mid S_1=0) = \frac{P(D=1, S_1=0)}{P(S_1=0)} = \frac{P(S_1=0 \mid D=1)\, P(D=1)}{P(S_1=0)}$$

$$= \frac{p\cdot 0.5}{(1-p)\cdot 0.5 + p\cdot 0.5} = p$$

$\therefore$ $P(D=0 \mid S_1=0) = 1-p$ $\quad P(D=1 \mid S_1=0) = p$

(b) $$P(S_2=0 \mid S_1=0) = \sum_{d=0}^{1} P(S_2=0, D=d \mid S_1=0)$$

$$= \sum_{d=0}^{1} P(D=d \mid S_1=0)\, P(S_2=0 \mid D=d, S_1=0)$$

$$= P(D=0 \mid S_1=0)\, P(S_2=0 \mid D=0, S_1=0) +$$
$$P(D=1 \mid S_1=0)\, P(S_2=0 \mid D=1, S_1=0)$$

$$= (1-p)\cdot(1-p) + p\cdot p = 1-2p+p^2+p^2 = 1-2p(1-p)$$

$\therefore$ $P(S_2=0 \mid S_1=0) = 1-2p(1-p)$

(c) $P(S_2=0 \mid S_1=0) = 1-2p(1-p)$ $\quad\therefore$ $P(S_2=1 \mid S_1=0) = 2p(1-p)$

For $P(S_2=0 \mid S_1=0)$

$$= \sum_{d=0}^{1} P(S_2=0, D=d \mid S_1=1)$$

$$= P(D=0 \mid S_1=1)\cdot P(S_2=0 \mid D=0, S_1=1) +$$
$$P(D=1 \mid S_1=1)\cdot P(S_2=0 \mid D=1, S_1=1)$$

$$= p\cdot(1-p) + p\cdot(1-p) = 2p(1-p)$$

$$H(S_2 \mid S_1) = H(S_2 \mid S_1=0)\cdot P(S_1=0) + H(S_2 \mid S_1=1)\cdot P(S_1=1)$$

$$= H(2p(1-p))\cdot 0.5 + H(2p(1-p))\cdot 0.5$$

$$= H(2p(1-p))$$

With symmetry

$$p(S_1=0 \mid S_2=0) = 1-2p(1-p)$$
$$p(S_1=1 \mid S_2=0) = 2p(1-p)$$
$$p(S_1=0 \mid S_2=1) = 2p(1-p)$$
$$p(S_1=1 \mid S_2=1) = 1-2p(1-p)$$

$$H(S_1|S_2) = H(S_1|S_2=0)\,p(S_2=0) + H(S_1|S_2=1)\,p(S_2=1)$$
$$= H(2p(1-p))\cdot 0.5 + H(2p(1-p))\cdot 0.5$$
$$= H(2p(1-p))$$

$$p(S_1=0) = P(D=0)P(N_1=0) + P(D=1)\cdot P(N_1=1)$$
$$= 0.5$$

$$\therefore P(S_1=1)=0.5 \qquad H(S_1) = H(\tfrac{1}{2}, \tfrac{1}{2}) = 1$$

$$\therefore H(S_2|S_1) + H(S_1) = H(S_1, S_2) = 1 + H(2p(1-p))$$

$$R_1 > H(S_1|S_2) = H(2p(1-p))$$
$$R_2 > H(S_2|S_1) = H(2p(1-p))$$
$$R_1 + R_2 > H(S_1, S_2) = 1 + H(2p(1-p))$$