EE231A HW4 Shengze Ye 205418909

1. An additive noise channel

$$Y = X + Z$$

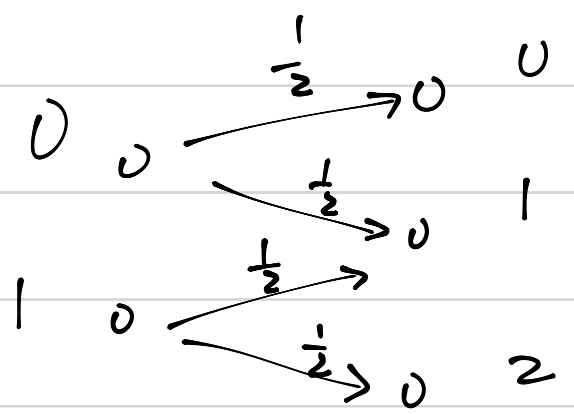
$Pr\{Z=0\} = Pr\{Z=a\} = \frac{1}{2}$. The alphabet for X is $\mathcal{X} = \{0,1\}$. Assume that Z is independent of X.

① if $a=0$ $\quad Y = X + Z = X$

$$C = \max I(X;Y) = \max \left(H(Y) - H(Y|X) \right)$$
$$= \max H(Y) = \max H(X)$$
$$= 1 \text{ bit / transmission}$$

It is achievable with $\begin{cases} p(X=0) = \frac{1}{2} \\ p(X=1) = \frac{1}{2} \end{cases}$

② if $a=1$ $\quad Y = X + Z \quad Pr(Z=0) = Pr(Z=1) = \frac{1}{2}$

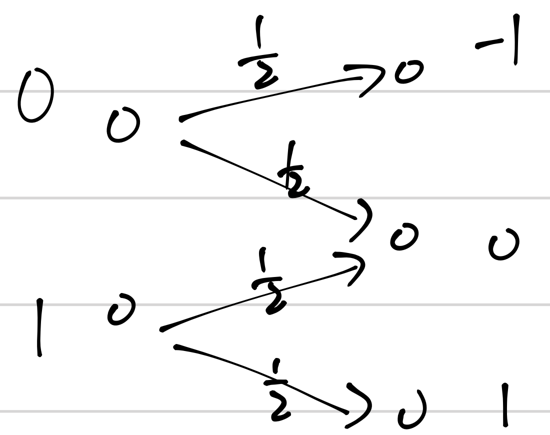


There are 3 outcomes and the channel is binary erasure channel with $\alpha = \frac{1}{2}$

$$C = 1 - \alpha = 1 - \frac{1}{2} = \frac{1}{2} \text{ bit per transmission}$$

The capacity is $\frac{1}{2}$ bit/transmission achieves when $p(X=0) = \frac{1}{2}$ and $p(X=1) = \frac{1}{2}$

③ if $a=-1$ $\quad Y = X + Z \quad Pr(Z=0) = Pr(Z=-1) = \frac{1}{2}$

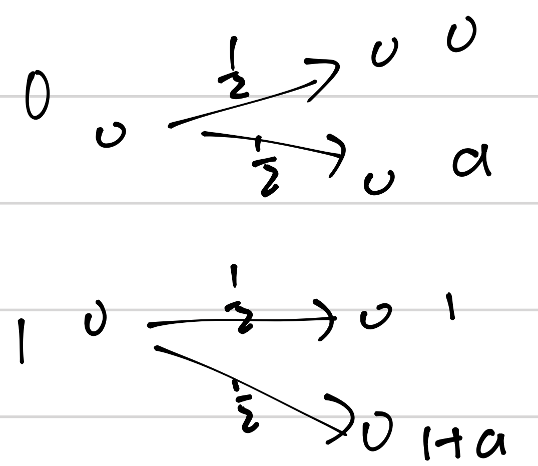


The same with ②, this is a binary erasure channel with $\alpha = \frac{1}{2}$

$\therefore\ C = 1 - \alpha = 1 - \frac{1}{2} = \frac{1}{2}$ bit/transmission

achieves when $p(X=0) = \frac{1}{2}$ and $p(X=1) = \frac{1}{2}$

④ $a \neq 0, \pm 1$ $\quad Y = X + Z$



The results 0, a are from $X=0$

The results 1, 1+a are from $X=1$,

if we know Y, we know X

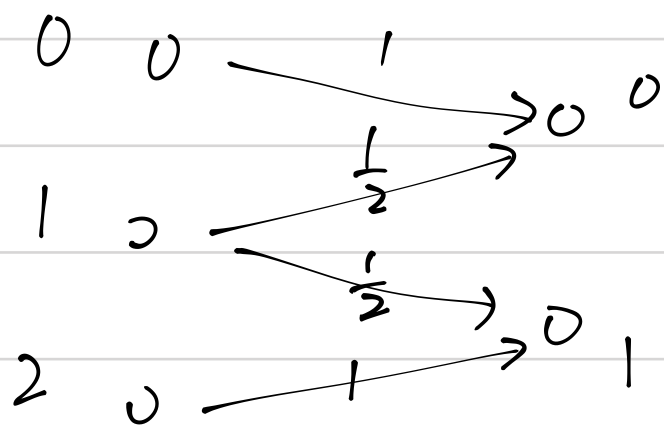
$\therefore\ H(X|Y) = 0$

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

$$= \max H(X) = 1 \text{ bit/trans}$$

achieves when $p(X=0) = \frac{1}{2}$ $\quad p(X=1) = \frac{1}{2}$

2. Inverse Erasure Channel



$$C = \max I(X;Y) = \max\left[H(Y) - H(Y|X)\right]$$
$$= \max\left[H(Y) - \sum_i p(X=x_i)\, H(Y|X=x_i)\right]$$

$$= \max\left[H(Y) - \left[p(X=0)H(Y|X=0) + p(X=1)H(Y|X=1) + p(X=2)H(Y|X=2)\right]\right]$$
$$= \max\left[H(Y) - p(X=1)\cdot H\left(\tfrac{1}{2}\right)\right]$$
$$= \max\ H(Y) - p(X=1)$$

So, we want $H(Y)$ as large as possible and $p(X=1)$ as small as possible.

The largest possible value for $H(Y)$ is 1, when $p(y=0) = p(y=1) = \frac{1}{2}$

This can be achieved as we set $p(X=0) = p(X=2)$, and set $p(X=1)$ to $1 - 2\cdot p(X=0)$

$\therefore$ when $p(x) = \begin{cases} \frac{1}{2} & x=0 \\ 0 & x=1 \\ \frac{1}{2} & x=2 \end{cases}$

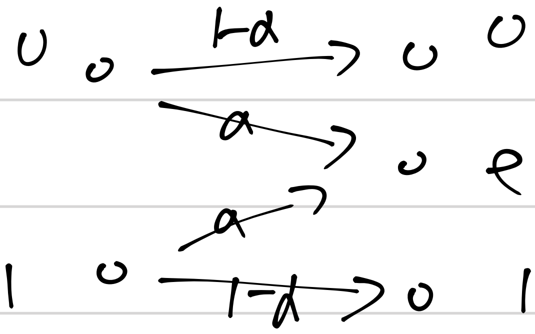
we can have the $\max H(Y) - p(X=1) = 1 - 0 = 1$

$\therefore$ The channel capacity is 1 bit/transmission

3. Cyclic Symmetry

(a) Give P for the binary erasure channel

BEC



$$P = \begin{bmatrix} 1-\alpha & \alpha & 0 \\ 0 & \alpha & 1-\alpha \end{bmatrix}$$

(b) For the binary erasure channel, decompose the columns of P into the subsets S_i described above.

$$P = \begin{array}{c} \begin{array}{ccc} c1 & c2 & c3 \end{array} \\ \begin{bmatrix} 1-\alpha & \alpha & 0 \\ 0 & \alpha & 1-\alpha \end{bmatrix} \end{array}$$ There are 3 colums in P

we can divide them into $S_1 = \begin{bmatrix} \alpha \\ \alpha \end{bmatrix}$ $S_2 = \begin{bmatrix} 1-\alpha & 0 \\ 0 & 1-\alpha \end{bmatrix}$

$\therefore$ $S_1 = \{c_2\}$ $S_2 = \{c_1, c_3\}$

① $\bigcup_i S_i = S_1 \cup S_2 = \{c_1, c_2, c_3\} = C$

② if $S_i \neq S_j$ then $S_i \cap S_j = \phi$

for S_1. The cyclic shift for $\begin{bmatrix} \alpha \\ \alpha \end{bmatrix}$ is itself

for S_2, the cyclic shift for $\begin{bmatrix} 1-\alpha \\ 0 \end{bmatrix}$ is $\begin{bmatrix} 0 \\ 1-\alpha \end{bmatrix}$. they are all in the S_2

$\therefore$ The columns of P can be decomposed into

$S_1 = \{ \begin{bmatrix} \alpha \\ \alpha \end{bmatrix} \}$ $S_2 = \{ \begin{bmatrix} 1-\alpha \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1-\alpha \end{bmatrix} \}$

(c) Prove that the Matrix conditions for Cyclic Symmetry described above are indeed sufficient for a channel to have cyclic symmetry.

The transmission matrix P

$$P = \begin{bmatrix} p(y=y_1|x=x_1) & p(y=y_2|x=x_1) & \cdots & p(y=y_n|x=x_1) \\ p(y=y_1|x=x_2) & p(y=y_2|x=x_2) & \cdots & p(y=y_n|x=x_2) \\ \vdots & & & \\ p(y=y_1|x=x_m) & p(y=y_2|x=x_m) & \cdots & p(y=y_n|x=x_m) \end{bmatrix}$$

I suppose there are m outcomes for X

there are n outcomes for Y

Suppose we have a input distribution for X

X	x_1	x_2	x_3	$\cdots$	x_m
$p(x)$	p_1	p_2	p_3	$\cdots$	p_m

$p_1 + \cdots + p_m = 1$

let $P_X = [p_1, p_2, p_3, \cdots, p_m]$

we can simplify the notation for transmission matrix:

$$\begin{array}{cc} P = & \begin{bmatrix} P_{11} & P_{21} & \cdots & P_{n1} \\ \vdots & \vdots & & \\ P_{1m} & P_{2m} & \cdots & P_{nm} \end{bmatrix} \\ \text{columns} & \begin{matrix} \uparrow & \uparrow & & \uparrow \\ c_1 & c_2 & \cdots & c_n \end{matrix} \end{array}$$

For the input distribution P_X,

the channel capacity $C = \max I(X;Y)$

$= \max H(Y) - H(Y|X)$

Now, if we have a cyclic shift for input distribution

we suppose the shift is i

$\therefore$

X	x_1	x_2	x_3	----	x_m
$p(x)$	$P_{(i+1) \bmod m}$	$P_{(i+2) \bmod m}$	$P_{(i+3) \bmod m}$		$P_{(i+m) \bmod m}$
	$\downarrow$	$\downarrow$	$\downarrow$		$\downarrow$
	p_1'	p_2'	p_3'		p_m'

we can denote this shifted version as $P_X' = [p_1' \cdots p_m']$

For the shifted input distribution P_X'

The channel capacity $C' = \max I(X;Y)'$

$$= \max \; H(Y)' - H(Y|X)'$$

① For $H(Y|X)'$

$$H(Y|X)' = \sum_i p_i' H(Y|X=x_i)$$

according to the first property, all the rows of P are permutations of each other

$$P = \begin{bmatrix} p_{11} & p_{21} & \cdots & p_{n1} \\ \vdots & \vdots & & \vdots \\ p_{1m} & p_{2m} & \cdots & p_{nm} \end{bmatrix}$$

the entropy of each row is the same

$$\therefore H(Y|X=x_1) = H(Y|X=x_2) = \cdots = H(Y|X=x_m)$$

$$\therefore \sum_i p_i' H(Y|X=x_i) = H(Y|X=x_1)$$

$$H(Y|X)' = H(Y|X)$$

The cyclic shift of input distribution does not change $H(Y|X)$

② for $H(Y)'$, we first consider the original $H(Y)$

$$P_X \times P =$$

$$[p_1 \ p_2 \cdots p_m] \times \begin{bmatrix} p_{11} & p_{21} & \cdots & p_{n1} \\ \vdots & \vdots & & \\ p_{1m} & p_{2m} & \cdots & p_{nm} \end{bmatrix} = [q_1 \ q_2 \cdots q_n]$$

distribution for y

$$H(Y) = H(q_1, q_2 \cdots q_n)$$

If we use shifted distribution

$$P_X' \times P =$$

$$[p_1' \ p_2' \cdots p_m'] \times \begin{bmatrix} p_{11} & p_{21} & \cdots & p_{n1} \\ \vdots & \vdots & & \vdots \\ p_{1m} & p_{2m} & \cdots & p_{nm} \end{bmatrix} = [q_1' \ q_2' \cdots q_n']$$

According to the property 2.

The set C of colums of P can be seperated into subset S_i

and for each S_i, each element in S_i has its cyclic shift in the same subset

consider

$$[p_1' \ p_2' \cdots p_m'] \times \begin{bmatrix} | & | & & | \\ S_{i1} & S_{i2} & \cdots & S_{ir} \\ | & | & & | \end{bmatrix} = [q_{i1}' \ q_{i2}' \cdots q_{ir}']$$

$S_{i1}, \cdots S_{ir}$ are r columns in S_i. they are cyclic shift for each other

Since S_i contains one instance with all of its cyclic shift element. $[q_i' \cdots q_{ir}']$ is a permutation of $[q_{i_1}, q_{i_2}, \cdots q_{ir}]$

$\therefore$ for all the subsets

$P_x' \times P =$ stack all the columns together

$$[p_1' \; p_2' \cdots p_m'] \times [S_1, S_2 \cdots S_k] = [q_1' \; q_2' \cdots q_n']$$

$\therefore [q_1' \; q_2' \cdots q_n']$ is a permutation of $[q_1, q_2 \cdots q_n]$

$$\therefore H(Y)' = H(q_1', q_2' \cdots q_n')$$
$$= H(q_1, q_2 \cdots q_n) = H(Y)$$

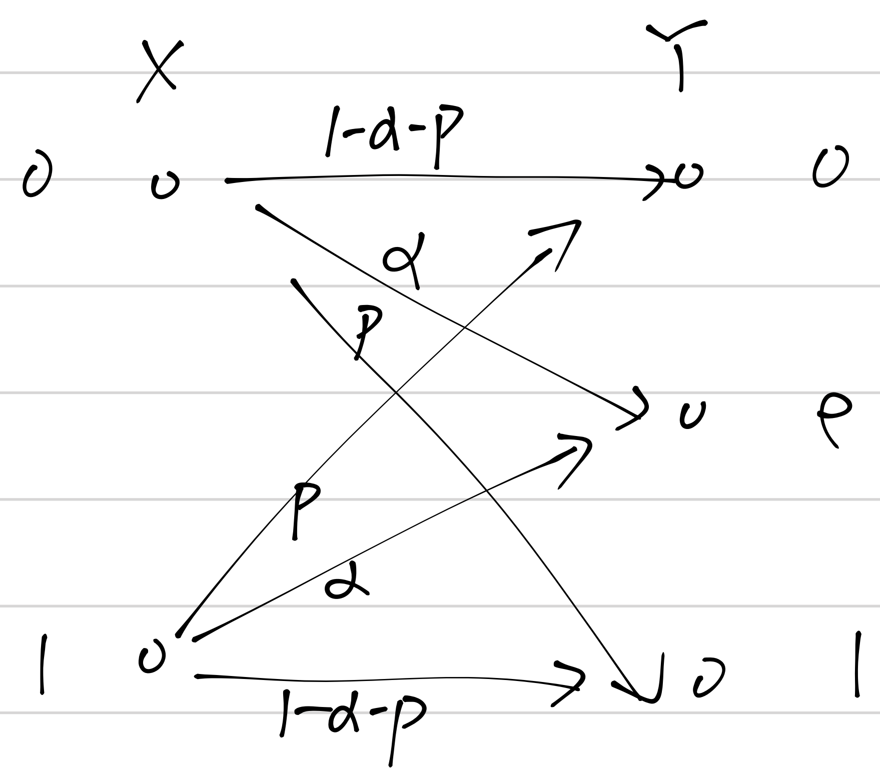
Therefore, we have shown that

$$c' = \max [H(Y)' - H(Y|X)']$$
$$= \max [H(Y) - H(Y|X)]$$
$$= C$$

$\therefore$ when we cyclic shift the input distribution The mutual information remains the same.

$\therefore$ The Matrix conditions for Cyclic Symmetry is sufficient for a channel to have cyclic symmetry.

4. Errors, Erasure, and Symmetry



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

$$P = \begin{bmatrix} 1-\alpha-p & \alpha & p \\ p & \alpha & 1-\alpha-p \end{bmatrix}$$

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

Yes

we can use the results from previous questions.

It satisfies the Matrix conditions for Cyclic Symmetry.

① All the rows of P are permutation of each other.

② the P can be divided into

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

such that $\bigcup_i S_i = C$ and $S_1 \neq S_2$ $S_1 \cap S_2 = \phi$
and each element (columns) in S_1 and S_2 are permutations of each other.
It is cyclic symmetry.

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

No

$$P = \begin{bmatrix} 1-\alpha-p & \alpha & p \\ p & \alpha & 1-\alpha-p \end{bmatrix}$$

The column sum is $1-\alpha$, 2α, $1-\alpha$, if it is weakly symmetry, the columns sums are equal.
if $\alpha = \frac{1}{3}$. it is weakly symmetry, $(1-\alpha = 2\alpha)$
if $\alpha \neq \frac{1}{3}$. it is not weak symmetry.

(d) Find the capacity for this channel.

This is a cyclic symmetry. so the uniform distribution achieves the capacity

$$C = \max I(X;Y)$$
$$= \max H(Y) - H(Y|X)$$
$$= \max H(Y) - \sum_i P(X=x_i) H(Y|X=x_i)$$
$$= \max H(Y) - H(1-\alpha-p, \alpha, p)$$

when input x is uniform distribution

$H(Y) = H(\frac{1-\alpha}{2}, \frac{1-\alpha}{2}, \alpha)$

$C = H(\frac{1-\alpha}{2}, \frac{1-\alpha}{2}, \alpha) - H(1-\alpha-p, \alpha, p)$

$C = H(\alpha) + 1-\alpha - H(1-\alpha-p, \alpha, p)$

achieves when $p(x=0) = p(x=1) = \frac{1}{2}$

(e) Show how your capacity expression simplifies to other channel capacities

if $p=0$, $\alpha \neq 0$

$C = H(\frac{1-\alpha}{2}, \frac{1-\alpha}{2}, \alpha) - H(\alpha)$

$= H(\alpha) + (1-\alpha) H(\frac{1}{2}, \frac{1}{2}) - H(\alpha)$

$= 1-\alpha$

$\therefore$ it degrades to binary erasure channel

if $\alpha=0$, $p \neq 0$

$C = H(\frac{1}{2}, \frac{1}{2}) - H(1-p, p) = 1 - H(p)$

it degrades to Binary Symmetric Channel (BSC)

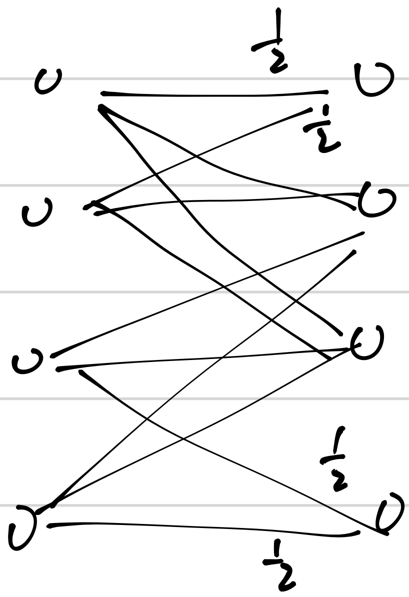
if $\alpha=0$ and $p=0$

$C = H(\frac{1}{2}, \frac{1}{2}) - 0 = 1$

$0\ 0 \xrightarrow{1} 0\ 0$ It is Noiseless Binary channel

$1\ 0 \xrightarrow{1} 0\ 1$

5. Symmetric Channel?



(a) Does this channel meet the conditions to be weakly symmetric?

If a channel is weakly symmetric.

① All the rows are permutations of each other

② column sums are equal.

$$P = \begin{bmatrix} \frac{1}{2} & \frac{1}{4} & \frac{1}{4} & 0 \\ \frac{1}{2} & \frac{1}{4} & \frac{1}{4} & 0 \\ 0 & \frac{1}{4} & \frac{1}{4} & \frac{1}{2} \\ 0 & \frac{1}{4} & \frac{1}{4} & \frac{1}{2} \end{bmatrix}$$

It meets the conditions to be weakly symmetric

(b) Is the mutual information of this channel invariant to shifts in the input probability distribution?

No.

X has 4 outcomes and Y has 4 outcomes

suppose

$$p(x) = \begin{cases} \frac{1}{2} & x=1 \\ \frac{1}{2} & x=2 \\ 0 & x=3 \\ 0 & x=4 \end{cases}$$

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

$$H(Y|X) = \sum_i p(X=x_i) H(Y|X=x_i)$$

$$= H(\tfrac{1}{2}, \tfrac{1}{4}, \tfrac{1}{4}) \cdot \tfrac{1}{2} + H(\tfrac{1}{2}, \tfrac{1}{4}, \tfrac{1}{4}) \cdot \tfrac{1}{2}$$

$$= H(\tfrac{1}{2}, \tfrac{1}{4}, \tfrac{1}{4})$$

$$H(Y) = H(\tfrac{1}{2}, \tfrac{1}{4}, \tfrac{1}{4})$$

$$\therefore \quad C = \max H(Y) - H(Y|X) = 0$$

Then, we make a cyclic shift for input distribution

$$p(x) = \begin{cases} 0 & x=1 \\ \frac{1}{2} & x=2 \\ \frac{1}{2} & x=3 \\ 0 & x=4 \end{cases}$$

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

Again $H(Y|X) = H(\tfrac{1}{2}, \tfrac{1}{4}, \tfrac{1}{4})$

But $H(Y) = H(\tfrac{1}{4}, \tfrac{1}{4}, \tfrac{1}{4}, \tfrac{1}{4}) = 2$

$$C = \max [2 - H(\tfrac{1}{2}, \tfrac{1}{4}, \tfrac{1}{4})]$$

$\therefore$ The mutual information has changed.

It is not invariant to shift.

(c) what is the capacity of this channel?

This is a weakly symmetric channel.

$$C = \log|y| - H(r)$$

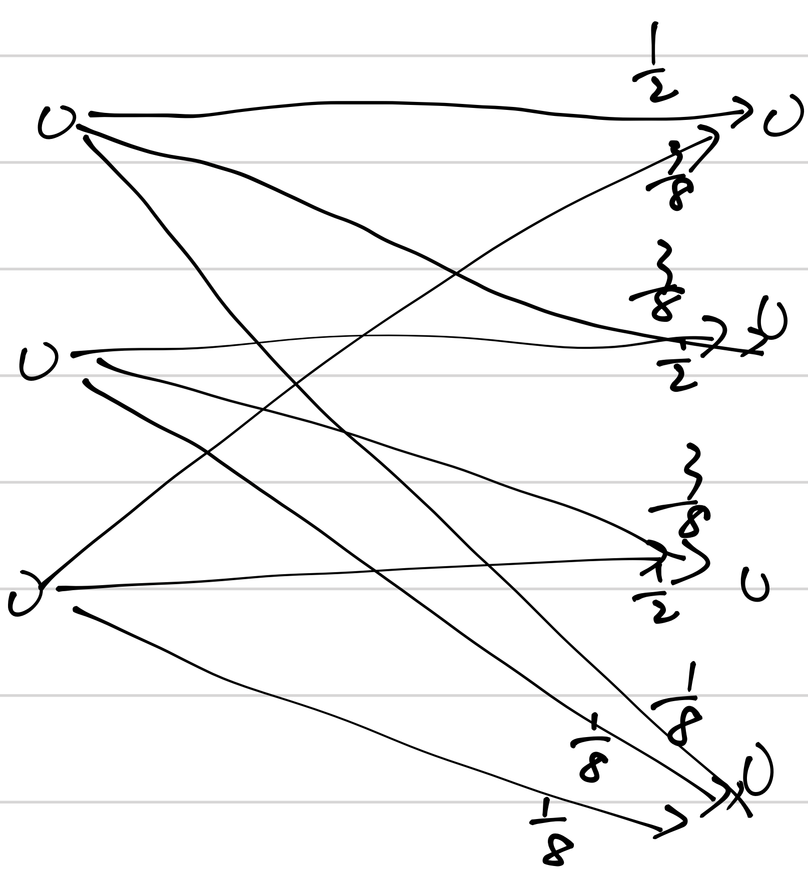
$$\therefore\ C = \log 4 - H\left(\tfrac{1}{2}, \tfrac{1}{4}, \tfrac{1}{4}\right)$$

$$= 2 - \left(-\tfrac{1}{2}\log\tfrac{1}{2} - \tfrac{1}{4}\log\tfrac{1}{4} - \tfrac{1}{4}\log\tfrac{1}{4}\right)$$

$$= 2 - \left[\tfrac{1}{2} + \tfrac{1}{2} + \tfrac{1}{2}\right]$$

$$= \tfrac{1}{2} \text{ bit / transmission}$$

6. Symmetric channel ?



(a) Does this channel meet the conditions to be weakly symmetric?

If a channel is weakly symmetric

① All the rows are permutations of each other
② the columns sums are equal

$$P = \begin{bmatrix} \frac{1}{2} & \frac{3}{8} & 0 & \frac{1}{8} \\ 0 & \frac{1}{2} & \frac{3}{8} & \frac{1}{8} \\ \frac{3}{8} & 0 & \frac{1}{2} & \frac{1}{8} \end{bmatrix}$$

The sum of the 4th column is $\frac{3}{8}$, while others' are $\frac{1}{2}$, columns sums are different.

It is not weakly symmetric.

(b) Is the mutual information of this channel invariant to cyclic shifts?

Yes.

① All the rows are permutations of each other
② According to the previous question we can divide the columns in P into 2 subsets

$$S_1 = \left\{ \begin{bmatrix} \frac{1}{2} \\ 0 \\ \frac{3}{8} \end{bmatrix}, \begin{bmatrix} \frac{3}{8} \\ \frac{1}{2} \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ \frac{3}{8} \\ \frac{1}{2} \end{bmatrix} \right\} \quad S_2 = \left\{ \begin{bmatrix} \frac{1}{8} \\ \frac{1}{8} \\ \frac{1}{8} \end{bmatrix} \right\}$$

s.t. $S_1 \cup S_2 = P$

$S_1 \neq S_2 \quad S_1 \cap S_2 = \phi$

and all the elements in S_1, S_2 are permutations to each other, respectively.

It satisfies the Matrix Condition for Cyclic Symmetry

The capacity is achieved when input distribution is uniform

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

$$p(x) = \begin{cases} \frac{1}{3} & x=1 \\ \frac{1}{3} & x=2 \\ \frac{1}{3} & x=3 \end{cases} \Rightarrow p(y) = \begin{cases} \frac{7}{24} & y=1 \\ \frac{7}{24} & y=2 \\ \frac{7}{24} & y=3 \\ \frac{1}{8} & y=4 \end{cases}$$

$$C = H\left(\frac{7}{24}, \frac{7}{24}, \frac{7}{24}, \frac{1}{8}\right) - H\left(\frac{1}{2}, \frac{3}{8}, \frac{1}{8}\right)$$

(c) Compute the capacity using the technique of finding an upper bound and then achieving it.

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

$$= \max [H(Y) - \sum_i p(X=x_i) H(Y|X=x_i)]$$

$$= \max [H(Y) - H(\tfrac{1}{2}, \tfrac{3}{8}, \tfrac{1}{8})]$$

$$= \max [H(Y)] - H(\tfrac{1}{2}, \tfrac{3}{8}, \tfrac{1}{8})$$

when $p(X=1) = p(X=2) = p(X=3) = \frac{1}{3}$

$H(Y)$ achieves its maximum $H\left(\frac{7}{24}, \frac{7}{24}, \frac{7}{24}, \frac{1}{8}\right)$

$$\therefore C = H\left(\frac{7}{24}, \frac{7}{24}, \frac{7}{24}, \frac{1}{8}\right) - H\left(\frac{1}{2}, \frac{3}{8}, \frac{1}{8}\right)$$

7. Joint Typicality

The sequence pair (X_1^n, Y_1^n) is drawn i.i.d. according to p.m.f

$$p(x^n, y^n) = \prod_{i=1}^{n} p(x_i, y_i)$$

(a) what is the limit as $n \to \infty$ of the probability that (X_1^n, Y_1^n) is in $A_\epsilon^{(n)}$ of joint typical sequences?

$$-\frac{1}{n}\log p(x^n, y^n) = -\frac{1}{n}\log \prod_{i=1}^{n} p(x_i, y_i)$$

$$= -\frac{1}{n}\sum_{i=1}^{n}\log p(x_i, y_i)$$

Since $p(x_1, y_1), p(x_2, y_2) \cdots$ are i.i.d.

According to Weak law of large numbers

$$-\frac{1}{n}\sum_{i=1}^{n}\log p(x_i, y_i) \to -E[\log p(x, y)]$$

$$= H(X, Y)$$

when $n \to \infty$

$\therefore \left| -\frac{1}{n}\log p(x^n, y^n) - H(X, Y)\right| < \epsilon \qquad Pr(A_\epsilon^{(n)}) > 1-\epsilon$

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

$\therefore$ as $n \to \infty$, the limit of probability that (X_1^n, Y_1^n) in $A_\epsilon^{(n)}$ is 1

(b) Give an upper bound on the probability that $(X_1^n, Y_2^n) \in A_\epsilon^{(n)}$. The bound should be a function of n and ϵ.

X_1^n and Y_2^n are independent with the same marginals as $p(X_1^n, Y_1^n)$ or $p(X_2^n, Y_2^n)$, since $p(X_2^n, Y_2^n)$ and $p(X_1^n, Y_1^n)$ are drawn in the same way, but independently.

$$p\left((X_1^n, Y_2^n) \in A_\epsilon^{(n)}\right) = \sum_{(x^n, y^n) \in A_\epsilon^{(n)}} p(x^n)\, p(y_2^n)$$

$$\leq |A_\epsilon^{(n)}|\, 2^{-n(H(X)-\epsilon)}\, 2^{-n(H(Y)-\epsilon)}$$

$$\leq 2^{n(H(X,Y)+\epsilon)}\, 2^{-n(H(X)-\epsilon)}\, 2^{-n(H(Y)-\epsilon)}$$

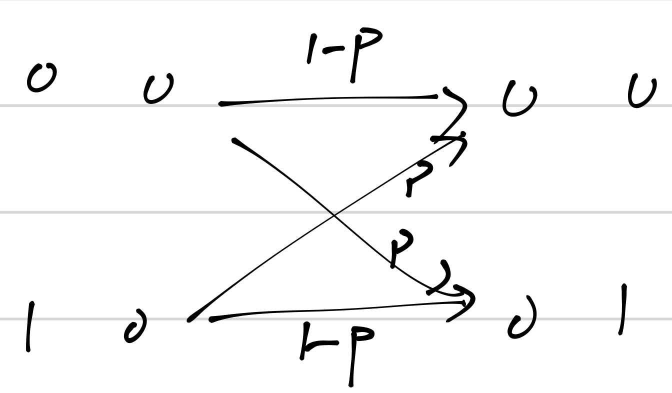
$$= 2^{-n(I(X;Y)-3\epsilon)}$$

$\therefore$ The upper bound is $2^{-n(I(X;Y)-3\epsilon)}$

(c) For X equally likely to be 0 or 1, and for $p(y|x)$ of a BSC with $p = 0.03113$, evaluate this upper bound for $n = 100$ and $\varepsilon = 0.1$

The upper bound is $2^{-n(I(X;Y)-3\varepsilon)}$

For BSC



$p(X=0) = p(X=1) = \frac{1}{2}$

X equally likely to be 0 or 1.

$\therefore I(X;Y) = H(Y) - H(Y|X) = 1 - H(p)$

$p = 0.03113$

$H(p) = -0.03113 \log_2 0.03113 - 0.96889 \cdot \log_2 0.96889$

$= 0.2$

$\therefore I(X;Y) = 1 - 0.2 = 0.8$

upper bound is $2^{-100(0.8 - 3\times 0.1)} = 2^{-50} = 8.8818 \times 10^{-16}$

The upper bound is $2^{-50} = 8.8818 \times 10^{-16}$

(d) What is the limit as $n \to \infty$ of the probability that (x_1^n, y_2^n) is in the set $A_\varepsilon^{(n)}$ of joint typical sequences?

There are about $2^{nH(X)}$ typical X sequences

about $2^{nH(Y)}$ typical Y sequences

about $2^{nH(X,Y)}$ typical X, Y sequenes

x_1^n, y_2^n are chosen randomly.

The probability that (x_1^n, y_2^n) is in $A_\epsilon^{(n)}$ is $2^{-nI(X;Y)}$

when $n \to \infty$ if $I(X;Y) \neq 0$ $2^{-nI(X;Y)} \to 0$

if $I(X;Y) = 0$ $2^{-n \cdot 0} \to 1$

(e) How is this related to the probability of error with typical set decoding ?

The probability of error

$P(\varepsilon | W=1) = P(E_1^c \cup E_2 \cup E_3 \cdots \cup E_{2^{nR}})$

error $\to E_2, E_3 \; -- \; E_i \; \cdots \; E_{2^{nR}}$

E_i is the event that the ith codeword is typical with the recieved sequence, while we transmitted the $W=1$ codeword.

In this problem, (x_1^n, y_2^n) is E_2, we transmitted x_1^n, and recieved y_2^n

It shows that when $n \to \infty$, the prob that (x_1^n, y_2^n) is $A_\epsilon^{(n)} \to 0$. $\therefore n \to \infty$, $Pr(E_2) \to 0$

That is the relationship with typical set decoding.

8. Triple Typicality.

(a) Assuming (x^n, y^n, z^n) are drawn according to $p(x, y, z)$ what does $P_r(A_\varepsilon^{(n)})$ converge to as $n \to \infty$

(x^n, y^n, z^n) are drawn from $p(x, y, z)$

$\therefore p(x^n, y^n, z^n) = \prod_{i=1}^{n} p(x_i, y_i, z_i)$

suppose $x \in \mathcal{X}$, $y \in \mathcal{Y}$ $z \in \mathcal{Z}$

Note that

$$p(X_1 = x_1, X_2 = x_2, Y_1 = y_1, Y_2 = y_2)$$

$$= \sum_{z_1 \in \mathcal{Z}, z_2 \in \mathcal{Z}} p(X_1 = x_1, X_2 = x_2, Y_1 = y_1, Y_2 = y_2, Z_1 = z_1, Z_2 = z_2)$$

$$= \sum_{z_1 \in \mathcal{Z}, z_2 \in \mathcal{Z}} p(X_1 = x_1, Y_1 = y_1, Z_1 = z_1)\, p(X_2 = x_2, Y_2 = y_2, Z_2 = z_2)$$

$$= \sum_{z_1 \in \mathcal{Z}} p(X_1 = x_1, Y_1 = y_1, Z_1 = z_1) \sum_{z_2 \in \mathcal{Z}} p(X_2 = x_2, Y_2 = y_2, Z_2 = z_2)$$

$$= p(X_1 = x_1, Y_1 = y_1)\, p(X_2 = x_2, Y_2 = y_2)$$

$\Downarrow$ in general

$$p(x^n, y^n) = \prod_{i=1}^{n} p(x_i, y_i) \quad ①$$

The same with ①, we can know that

$$p(x^n, z^n) = \prod_{i=1}^{n} p(x_i, z_i) \quad ②$$

$$p(y^n, z^n) = \prod_{i=1}^{n} p(y_i, z_i) \quad ③$$

$p(x^n) = \prod_{i=1}^{n} p(x_i)$ (4) $\quad p(y^n) = \prod_{i=1}^{n} p(y_i)$ (5)

$p(z^n) = \prod_{i=1}^{n} p(z_i)$ (6)

Then, we apply weak law large number

as $n \to \infty$

$-\frac{1}{n}\log p(x^n) \to H(X)$

$-\frac{1}{n}\log p(y^n) \to H(Y)$

$-\frac{1}{n}\log p(z^n) \to H(Z)$

$-\frac{1}{n}\log p(x^n, y^n) \to H(X, Y)$

$-\frac{1}{n}\log p(x^n, z^n) \to H(X, Z)$

$-\frac{1}{n}\log p(y^n, z^n) \to H(Y, Z)$

$-\frac{1}{n}\log p(x^n, y^n, z^n) \to H(X, Y, Z)$

$\therefore$ as $n \to \infty \qquad Pr(A_\epsilon^{(n)}) \to 1$

(b) Suppose that the pair of sequences (x^n, y^n) are drawn according to $p(x, y)$, and z^n is drawn according to $p(z)$ independently of (x^n, y^n). Derive an upper bound on the probability that (x^n, y^n, z^n) is in $A_\epsilon^{(n)}$

$$P((x^n, y^n, z^n) \in A_\epsilon^{(n)}) = \sum_{(x^n, y^n, z^n) \in A_\epsilon^{(n)}} p(x^n, y^n, z^n)$$

$$= \sum_{(x^n, y^n, z^n) \in A_\epsilon^{(n)}} p(x^n, y^n)\, p(z^n)$$

$$\le |A_\epsilon^{(n)}|\, 2^{-n(H(X,Y)-\epsilon)}\, 2^{-n(H(Z)-\epsilon)}$$

$$= 2^{n(H(X,Y,Z)+\epsilon)}\, 2^{-n(H(X,Y)-\epsilon)}\, 2^{-n(H(Z)-\epsilon)}$$

$$= 2^{-n[H(X,Y)+H(Z)-H(X,Y,Z)-3\epsilon]}$$

$$= 2^{-n(I(X,Y;Z)-3\epsilon)}$$

The upper bound is $2^{-n(I(X,Y;Z)-3\epsilon)}$

9. Joint Typicality on the Three-way Channel

(a) Find the upper bound on the probability that X^n, Y^n, Z^n is a member of triple set if X^n and (Y^n, Z^n) are chosen independently.

$$Pr\left((x^n, y^n, z^n) \in A_\epsilon^{(n)}\right) = \sum_{(x^n, y^n, z^n) \in A_\epsilon^{(n)}} p(x^n, y^n, z^n)$$

$$= \sum_{(x^n, y^n, z^n) \in A_\epsilon^{(n)}} p(x^n)\, p(y^n, z^n)$$

$$\le |A_\epsilon^{(n)}|\, 2^{-n(H(X)-\epsilon)}\, 2^{-n(H(Y,Z)-\epsilon)}$$

$$= 2^{n(H(X,Y,Z)+\varepsilon)} 2^{-n(H(X)-\varepsilon)} 2^{-n(H(Y,Z)-\varepsilon)}$$

$$= 2^{-n[H(X)+H(Y,Z)-H(X,Y,Z)-3\varepsilon]}$$

$$= 2^{-n(I(X;Y,Z)-3\varepsilon)}$$

$\therefore$ The upper bound is $2^{-n(I(X;Y,Z)-3\varepsilon)}$

(b) Prove via a random coding argument that with input distribution $P_i(x)$ we can achieve the rate $I(X;Y,Z)$

Just like in lecture

we fix $P_i(x)$, Generate $(2^{nR}, n)$ code C at random according to $P_i(x)$

$$p(x^n) = \prod_{i=1}^{n} p(x_i)$$

$$p(C) = \prod_{w=1}^{2^{nR}} \prod_{i=1}^{n} p(x_i(w))$$

if $(X^n(\hat{w}), Y^n, Z^n) \in A_\varepsilon^{(n)}$, we declares $\hat{w}$

$$\lambda_i = P(\hat{w} \neq i \mid X^n = x^n(i))$$

$$P_e^{(n)}(C) = \frac{1}{2^{nR}} \sum_{w=1}^{2^{nR}} \lambda_w(C)$$

The average prob. of all possible c's is

$$P(\mathcal{E}) = \sum_c p(c) P_e^{(n)}(c)$$

$$\begin{aligned}
P(\mathcal{E}) &= \sum_c p(c) P_e^{(n)}(c) \\
&= \sum_c p(c) \frac{1}{2^{nR}} \sum_{w=1}^{2^{nR}} \lambda_w(c) \\
&= \frac{1}{2^{nR}} \sum_{w=1}^{2^{nR}} \sum_c p_c \cdot \lambda_w(c) \\
&= \sum_c p(c) \lambda_1(c) = P(\mathcal{E} | W=1)
\end{aligned}$$

Define the events $E_i = \{ (X^n(i), Y^n, Z^n) \in A_\varepsilon^{(n)} \}$

$i = 1, 2, \cdots 2^{nR}$

$$\begin{aligned}
P(\mathcal{E}|W=1) &= P(E_1^c \cup E_2 \cup \cdots \cup E_{2^{nR}}) \\
&\le P(E_1^c) + \sum_{i=2}^{2^{nR}} P(E_i) ,
\end{aligned}$$

$P(E_1^c) \to 0$ as $n \to \infty$

if $i \neq 1$ $\quad P(E_i) \le 2^{-n(I(X;Y,Z) - 3\varepsilon)}$

$$P(\mathcal{E}) \le \varepsilon + \sum_{i=2}^{2^{nR}} 2^{-n(I(X;Y,Z) - 3\varepsilon)}$$

$$= \varepsilon + (2^{nR} - 1)\, 2^{-n(I(X;Y,Z) - 3\varepsilon)}$$

$$\le \varepsilon + 2^{nR}\, 2^{-n(I(X;Y,Z) - 3\varepsilon)}$$

$$= \varepsilon + 2^{-n(I(X;Y,Z) - R - 3\varepsilon)}$$

if $R < I(X;Y,Z)$, when $n \to \infty$

the average probability of error converges to zero.

10. All codes are good

Prove Markov's inequality:

X is non-negative $\quad \delta > 0$

$$P(X \geq \delta) \leq \frac{EX}{\delta}$$

Suppose the outcomes of X are $0 \leq x_1 < x_2 < \cdots < x_n$

if $\delta \geq x_n$ (it is larger than the maximum value)

$P(X \geq \delta) = 0$. inequality is true

if not, we suppose that

$$x_j \leq \delta < x_{j+1}$$

$$\therefore x_1 < x_2 < \cdots < x_j \leq \delta < x_{j+1} < \cdots < x_n$$

$$E(X) = \sum_{i=1}^{n} x_i \cdot Pr(X = x_i) \geq \sum_{i=j}^{n} x_i \cdot Pr(X = x_i)$$

$$\geq \sum_{i=j}^{n} \delta \cdot Pr(X = x_i)$$

$$\therefore \frac{E(X)}{\delta} \geq \sum_{i=j}^{n} Pr(X = x_i) \geq \sum_{i=j+1}^{n} Pr(X = x_i)$$

$$= Pr(X \geq \delta)$$

since $x_{j+1} > \delta \geq x_j$

$$\therefore \frac{E(X)}{\delta} \geq Pr[X \geq \delta] \quad \square$$

use it to show that the channel coding theorem implies the following.

Let random variable X be the probability of error for a randomly selected code

$\therefore$ $E[x]$ is the average probability of error for all possible c's. that we discussed in lectures.

we fix a target probability of error ε

According to Markov's inequality

$$P(X \geq \varepsilon) \leq \frac{E[x]}{\varepsilon}$$

The channel coding theorem told us that as block length $n \to \infty$ $P(\varepsilon) \to 0$

$\therefore$ $E[X] \to 0$

$\therefore$ as $n \to \infty$, as the block length goes to infinity.

$P(X \geq \varepsilon) \to 0$

$\hookrightarrow$ the probability that a randomly selected code will exceed that target goes to zero.