EE 231A HW6 Shengze Ye 205418959

1. A truncated Gaussian

(a) Prove that the Normal density $\phi(x)$ maximizes differential entropy for a fixed second moment σ^2

$$\phi(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-x^2/2\sigma^2}$$

let X be an any random variable with $E[x]=0$ and the second moment $E[x^2]=\sigma^2$

let $g(x)$ be its density distribution function.

$$0 \le D(g\,\|\,\phi) = \int g(x) \log \frac{g(x)}{\phi(x)}\, dx$$

$$= \int g(x) \log g(x)\, dx - \int g(x) \log \phi(x)\, dx$$

$$= \int g(x) \log g(x)\, dx + \int g(x) \left[\frac{1}{2}\log(2\pi\sigma^2) + \log e \frac{x^2}{2\sigma^2} \right] dx$$

Here, we consider the second term

$$\int g(x) \frac{1}{2}\log(2\pi\sigma^2)\, dx + \int g(x) \log e \frac{x^2}{2\sigma^2}\, dx$$

$$= \frac{1}{2}\log(2\pi\sigma^2)\int g(x)\, dx + \frac{\log e}{2\sigma^2}\int g(x)\, x^2\, dx$$

$\Longrightarrow E[x^2]=\sigma^2$

$$= \frac{1}{2}\log(2\pi\sigma^2)\int \phi(x)\, dx + \frac{\log e}{2\sigma^2}\int \phi(x)\, x^2\, dx$$

$= -\int \phi(x) \log \phi(x) dx$

$\therefore \int g(x) \log g(x)\, dx - \int \phi(x) \log \phi(x) dx = h(\phi) - h(g) \geq 0$

$\therefore h(\phi) \geq h(g)$

$\therefore$ The normal distribution $\phi(x)$. maximizes differential entropy for a fixed second moment σ^2

(b) Use this fact to show that a Gaussian input maximizes the mutual information $I(X;Y)$ on the memoryless Gaussian channel $Y = X + Z$ where Z is a Gaussian independent of X and Y.

$Z \sim (0, \sigma^2)$

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

$= h(Y) - h(Z|X)$

$\because$ Z is independent of X

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

$h(Z)$ is $\frac{1}{2} \log 2\pi e \sigma^2$ is a constant

So, we need to maximize $h(Y)$

from (a) we know

$h(Y) \leq \frac{1}{2} \log 2\pi e \left(E[Y^2] \right)$

$$EY^2 = E(X+Z)^2 = EX^2 + 2EXEZ + EZ^2$$
$$= EX^2 + EZ^2 \qquad (E[X^2] \le P)$$
$$= P + \sigma^2 \qquad (Z \text{ is } 0 \text{ mean})$$

$\therefore$ $h(Y)$ reaches its maximum when $Y \sim (0, P+\sigma^2)$

$\therefore$ X should also be a Gaussian distribution

$\therefore$ The Gaussian input maximize the mutual information

(c) Prove that a truncated Normal density $\tau(x)$ maximize differential entropy for a fixed second moment under a peak limitation constraint.

From the example we can describe the truncated normal density as:

$$\tau(x) = \begin{cases} \frac{1}{k} \frac{1}{\sqrt{2\pi r^2}} e^{-\frac{x^2}{2r^2}} & x \in (-a, a) \\ 0 & \text{else} \end{cases}$$

where $k = \int_{-a}^{a} \frac{1}{\sqrt{2\pi r^2}} e^{-x^2/2r^2} dx$

r^2 is chosen that $\frac{1}{k} \int_{-a}^{a} x^2 \frac{1}{\sqrt{2\pi r^2}} e^{-x^2/2r^2} dx = \sigma^2$

$\therefore$ $E[X^2] = \sigma^2$

Consider a random varible X with a fixed second moment under a peak limitation constrait.

$g(x)$ is its density function, that could be any distribution

$$0 \le D(g \| \tau) = \int_{-a}^{a} g(x) \ln \frac{g(x)}{\tau(x)} dx$$

$$= \int_{-a}^{a} g(x) \ln g(x) dx - \int_{-a}^{a} g(x) \ln \tau(x) dx$$

$$= \int_{-a}^{a} g(x) \ln g(x) dx + \int_{-a}^{a} g(x) \left[-\frac{1}{2} \ln (2\pi\sigma^2 k^2) + \frac{x^2}{2\gamma^2} \right] dx$$

we consider the second term

$$= \int_{-a}^{a} \frac{1}{2} g(x) \cdot \ln (2\pi\sigma^2 k^2) dx + \int_{-a}^{a} g(x) x^2 / 2\gamma^2 \, dx$$

$$= \frac{1}{2} \ln (2\pi\sigma^2 k^2) \int_{-a}^{a} g(x) dx + \frac{1}{2\gamma^2} \int_{-a}^{a} g(x) x^2 dx$$

$$\rightarrow E[x^2] = \sigma^2$$

$$= \frac{1}{2} \ln (2\pi\sigma^2 k^2) \int_{-a}^{a} \tau(x) dx + \frac{1}{2\gamma^2} \int_{-a}^{a} \tau(x) x^2 dx$$

(From the definition of truncated normal density, we know that $\int_{-a}^{a} \tau(x) x^2 dx = \int_{-a}^{a} g(x) x^2 dx = E[x^2] = \sigma^2$

$$\therefore D(p \| q) = \int g(x) \ln g(x) dx - \int \tau(x) \ln (\tau(x)) dx \ge 0$$

$$\therefore h(\tau) - h(g) \ge 0$$

$\therefore$ $\tau(x)$ maximizes differential entropy for a fixed second moment entropy under a peak limitation constraint.

(d) Now consider a channel that is characterized by the peak limitation $X \in (-1, 1)$ on the transmit signal and additive noise that is a truncated Normal with distribution $T(x)$ as described above. Is the mutual information of this channel maximized by a truncated Gaussian input or not?

Not obvious

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

we know that $h(Z)$ is a constant as Z is a normal distribution $Z \sim (0, \sigma^2)$

So, we need to maximize the $h(Y)$

X is a peak limitation transit signal with $E[X^2] \le P$

from (c), we know that normal density maximizes the differential entropy. so $h(Y)$ reaches maximum as Y is a truncated Gaussian.

$Y = X + Z$, however, the sum of truncated Gaussian would not be truncated Gaussian. $\therefore$ The mutual information is not maximized by truncated Gaussian input

2. Shaping Gain.

Consider the additive white Gaussian noise channel with power constrained input X and nois $Z \sim N(0, N)$. The optimal input distribution is a Gaussian $X_g \sim N(0, P)$. For practical reasons, a suboptimal X_s that is not a Gaussian but does have $EX_s^2 = P$ is often used.

maximum shaping gain $= I(X_g; Y_g) - I(X_s; Y_s)$

$Y_g = X_g + Z \qquad Y_s = X_s + Z$

Show that the maximum shaping gain may be expressed succinctly as a relative entropy involving f_s and f_g.

$I(X_g; Y_g) - I(X_s; Y_s)$

$= h(Y_g) - h(Y_g|X_g) - [h(Y_s) - h(Y_s|X_s)]$

$= h(Y_g) - h(Y_g - X_g|X_g) - [h(Y_s) - h(Y_s - X_s|X_s)]$

$= h(Y_g) - h(Z|X_g) - [h(Y_s) - h(Z|X_s)]$

we know that Z is independent of X_g and X_s

$\therefore \quad I(X_g; Y_g) - I(X_s; Y_s)$

$= h(Y_g) - h(Z) - h(Y_s) + h(Z)$

$= h(Y_g) - h(Y_s)$

f_s, f_g is the p.d.f for Y_s and Y_g

$$= -\int f_g \log f_g \, dy + \int f_s \log f_s \, dy$$

since $X_g \sim N(0, P)$ $\quad Z \sim N(0, N)$

$Y_g = X_g + Z$ $\quad Y_g \sim N(0, N+P)$

$$f_g = \frac{1}{\sqrt{2\pi(N+P)}} e^{-\frac{y^2}{2(N+P)}}$$

$$\int f_g \log f_g \, dy = \int f_g \left[\frac{1}{2} \log 2\pi(N+P) + \log e \frac{y^2}{2(N+P)} \right] dy$$

$\because E X_s^2 = P$ $\quad Y_s = X_s + Z$ $\quad E Y_s^2 = P + N$

$\therefore \int f_s y^2 \, dy = N + P = \int f_g y^2 \, dy$

$$\therefore \int f_g \log f_g \, dy = \int f_s \left[\frac{1}{2} \log 2\pi(N+P) + \log e \frac{y^2}{2(N+P)} \right] dy$$
$$= \int f_s \log f_s \, dy$$

$I(X_g; Y_g) - I(X_s; Y_s)$

$$= -\int f_g \log f_g \, dy + \int f_s \log f_s \, dy$$

$$= -\int f_s \log f_g \, dy + \int f_s \log f_s \, dy$$

$$= \int f_s \log \frac{f_s}{f_g} \, dy = D(f_s \| f_g) \geq 0$$

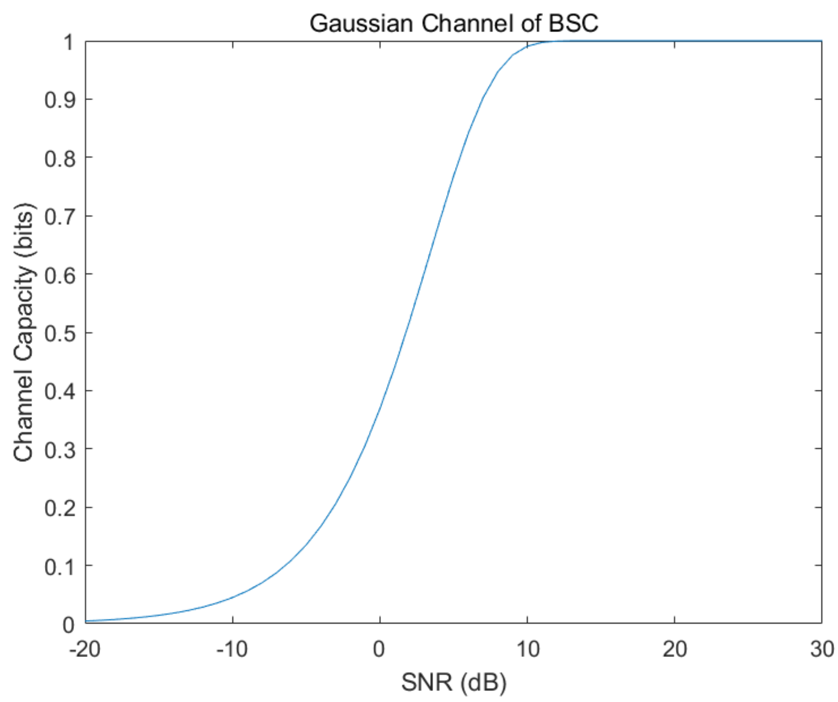
$\therefore$ The maximum shaping gain could be expressed as relative entropy $D(f_s \| f_g) = \int f_s \log \frac{f_s}{f_g} \, dy$

3. BSC. vs Gaussian channel

(a) use Matlab to plot the channel capacity of the Gaussian channel and of the BSC with transition probability

$$p = Q\left(\sqrt{P/N}\right)$$

plot the SNR range between -20 dB and 30 dB.



SNR = -20 : 1 : 30

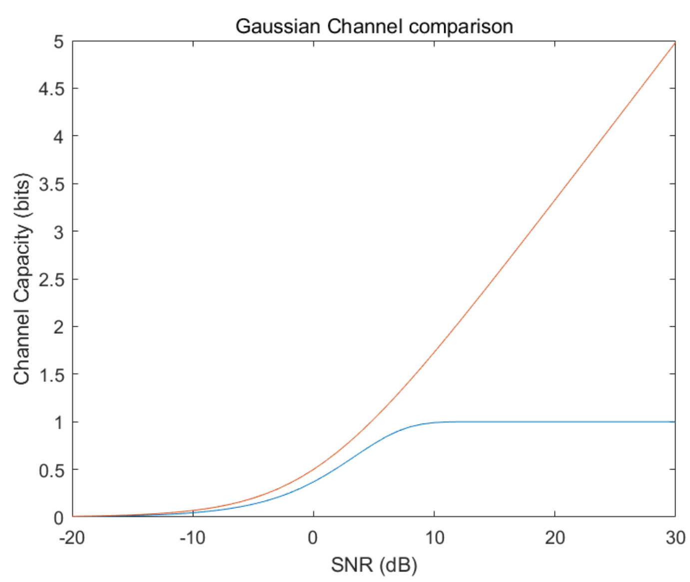
P_N = 10.^(SNR./10)

p = Q(sqrt(P_N))

C = 1 - H(p)

SNR ↑ C ↑

we see that when SNR is over 10 dB, the capacity is approximately reach the maximum capacity 1 bit.



Orange : $C = \frac{1}{2}\log(1+SNR)$

Blue : $C = 1 - H(p)$

Here, I also plot the $C = \frac{1}{2}\log(1+SNR)$ in the same figure. We can see that Gaussian channel with BSC is bounded by 1 bit. But the continuous input does not have this boundary.

(b) Now suppose that the input is quantized as before but the output is not quantized before coding. Prove that the Gaussian channel with binary PAM has a maximum mutual information curve that lies above the capacity curve of the BSC channel, but has the same asymptote.

At the input side, X could only be $\sqrt{P}$ and $-\sqrt{P}$, the mutual information $I(X;Y)$ is a measure of how much information can be transmitted over the channel for each channel use. For 2-PAM, the plots will flatten at the maximum transmitted bits for the number of signal alternatives as the channel becomes good,

$\therefore$ 2-PAM will flatten at 1 bit / channel use. it will aysmptote to 1 just like BSC.

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

For 2-PAM, we only quantize the input, not for output

But for Gaussian BSC, we also quatize on output

$$Y' = \operatorname{sgn}(Y) \cdot (\sqrt{P})$$

denote $Y' = g(Y)$

$$h(Y) > h(Y')$$

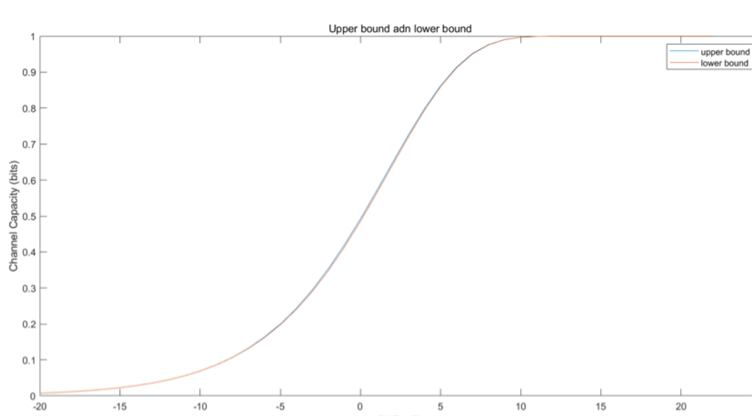
for Both channels, $h(Z)$ is the same

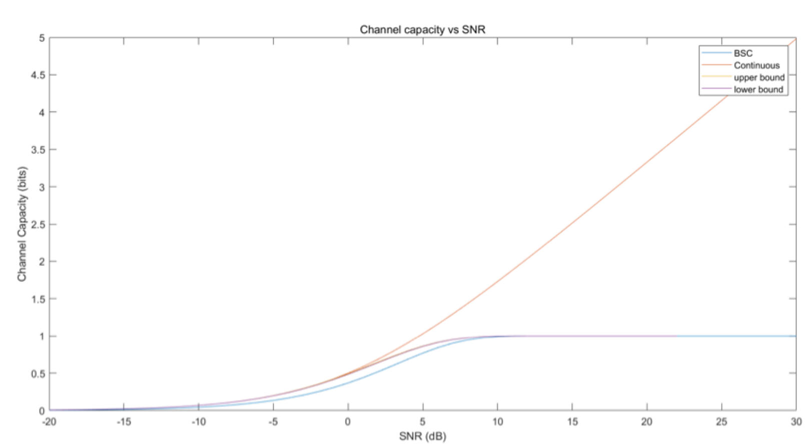
$$h(Y) > h(Y')$$

$\therefore$ 2-PAM will have larger $I(X;Y)$

$\therefore$ The curve of 2-PAM will lie above the BSC curve.

(c) For completeness, plot the upper bound and lower bounds on the mutual information of the AWGN channel when the input is BPSK with a uniform distribution





4. Shannon, Sensors, A/D convertors

(a) write down the capacity of the discrete time additive white Gaussian noise channel $z \sim N(0, N)$ what are the units of capacity?

For discrete inputs with continuous alphabet

$$C = \frac{1}{2}\log(1+SNR) = \frac{1}{2}\log\left(1+\frac{P}{N}\right) \text{ bit/symbol.}$$

The units is bits / symbol

(b) Write down the capacity of the continuous-time bandlimited AWGN channel with bandwidth W Hz. What are the units of capacity?

$$E[z^2] = \int_{-W}^{W} \frac{N_0}{2}\,df = \frac{N_0}{2}\cdot 2W = N_0 W$$

$$E[X_n^2] = P$$

$$C = W\log\left(1+\frac{P}{N_0 W}\right) \text{ bits/ second}$$

The units are bits / second

(c) Find the capacity of the infinite bandwidth AWGN channel with noise psd and transmitter power constraint in (b)

$$\lim_{w \to \infty} w \log\left(1+\frac{P}{N_0 w}\right) = \lim_{w \to \infty} w \log_2 e \ln\left(1+\frac{P}{N_0 w}\right)$$

$$= w \cdot \log_2 e \frac{P}{N_0 w} = \log_2 e \frac{P}{N_0}$$

$\therefore$ For infinite bandwidth.
The capacity is $\log_2 e \frac{P}{N_0}$

(d) The channel between a remote sensor and a base station is modeled by a discrete time AWGN with $N=1$. Give a lower bound on the amount of power/bit.

$$C = \frac{1}{2}\log\left(1+\frac{P}{N}\right) \quad N=1 \quad \therefore C = \frac{1}{2}\log(1+P)$$

energy/bit : $\dfrac{P}{C} = \dfrac{P}{\frac{1}{2}\log(1+P)}$

The minimum value appears when P is very small

$$\lim_{P \to 0} \frac{P}{C} = \lim_{P \to 0} \frac{P}{\frac{1}{2}\log(1+P)} = \frac{P}{\frac{1}{2}(P)} = 2$$

A lower bound is 2

(e) Actually, it is a continuous time AWGN channel. What is the correct choice of bandwidth to minimize energy per bit? How does efficiency compare with the discrete time

AWGN ? What are the practical limitations to achieving this ultimate efficiency?

$$C = W \log\left(1 + \frac{P}{N_0 W}\right) \quad \text{we assume } N_0 = 1$$

$$C = W \log\left(1 + \frac{P}{W}\right)$$

$$\frac{P}{C} = \frac{P}{W \log\left(1 + \frac{P}{W}\right)} = \frac{P/W}{\log(1 + P/W)}$$

to make P/C minimum, we need to set W as large as possible. ∴ The infinite bandwidth will minimize the energy per bit.

For efficiency comparison

for discrete channel $\quad \frac{P}{C} = \frac{2P}{\log(1+P)}$

for continuous channel $\quad \frac{P}{C} = \frac{P/W}{\log(1+P/W)}$

They have same format, but multiplied by different constant. if the P is equal to P/W, discrete channel has high efficiency since it has constant "2"

Practical limitations: ① we may not have infinite band

② we need a very high sampling rat, since we need $F_s \geq 2W$

(f) What is the highest sampling rate that is theoretically possible while still truly achieving one bit of resolution?

1 bit resolution means the system should be able to distinguish at least 1 bit/symbol.

$$C = W\log\left(1+\frac{P}{N_0W}\right)$$

$\therefore \frac{C}{2W} \geq 1$ bit/symbol $\quad \frac{C}{2W} = \frac{1}{2}\log\left(1+\frac{P}{N_0W}\right)$

$$= \frac{1}{2}\log\left(1+\frac{P}{N_0W}\right) \geq 1$$

$\Rightarrow \log\left(1+\frac{P}{N_0W}\right) \geq 2 \qquad 1+\frac{P}{N_0W} \geq 4 \qquad \frac{P}{N_0W} \geq 3$

$$W \leq \frac{P}{3N_0} = \frac{3}{10^{-9}\cdot 3} = 10^9\ Hz$$

$\therefore \quad F_s = 2W \leq 2\times10^9\ Hz$

$\therefore$ The highest sampling rate is $2\times10^9\ Hz$

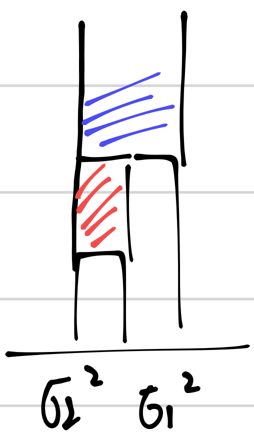
5. Parallel channels and waterfilling.

$$\begin{pmatrix} Y_1 \\ Y_2 \end{pmatrix} = \begin{pmatrix} X_1 \\ X_2 \end{pmatrix} + \begin{pmatrix} Z_1 \\ Z_2 \end{pmatrix}$$

$$\begin{pmatrix} Z_1 \\ Z_2 \end{pmatrix} \sim N\left(0, \begin{bmatrix} \sigma_1^2 & 0 \\ 0 & \sigma_2^2 \end{bmatrix}\right)$$

$E(X_1^2 + X_2^2) \leq 2P$ $\quad \sigma_1^2 > \sigma_2^2$, At what value of $2P$ does the channel stop behaving like a single channel and begin behaving like a pair of channels.

we will put all the powers to channel with less noise until the sum of power and noise in that channel equals to the other one.



when $0 < 2P < \sigma_1^2 - \sigma_2^2$

it is a single channel

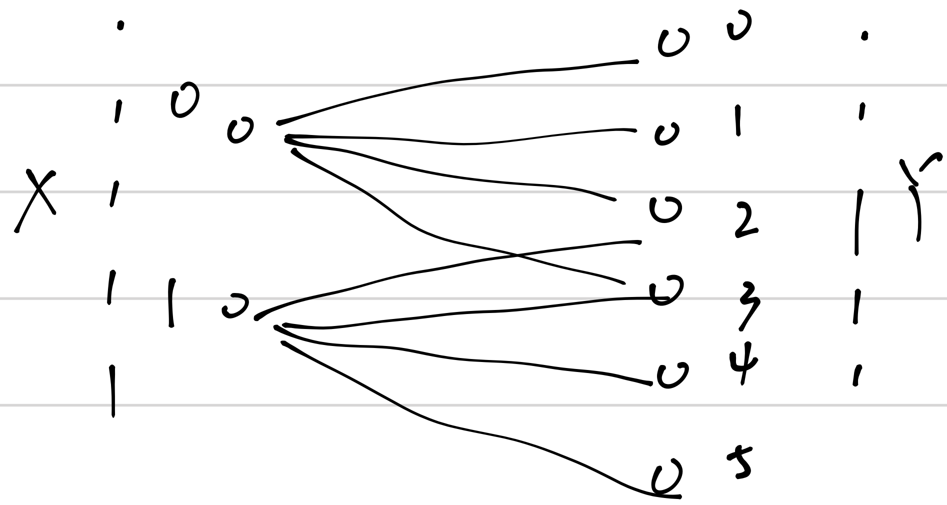
$\therefore$ when $2P = \sigma_1^2 - \sigma_2^2$, the channel

begins to behave like a pair of channels

$$2P = \sigma_1^2 - \sigma_2^2$$

6. Split Ends

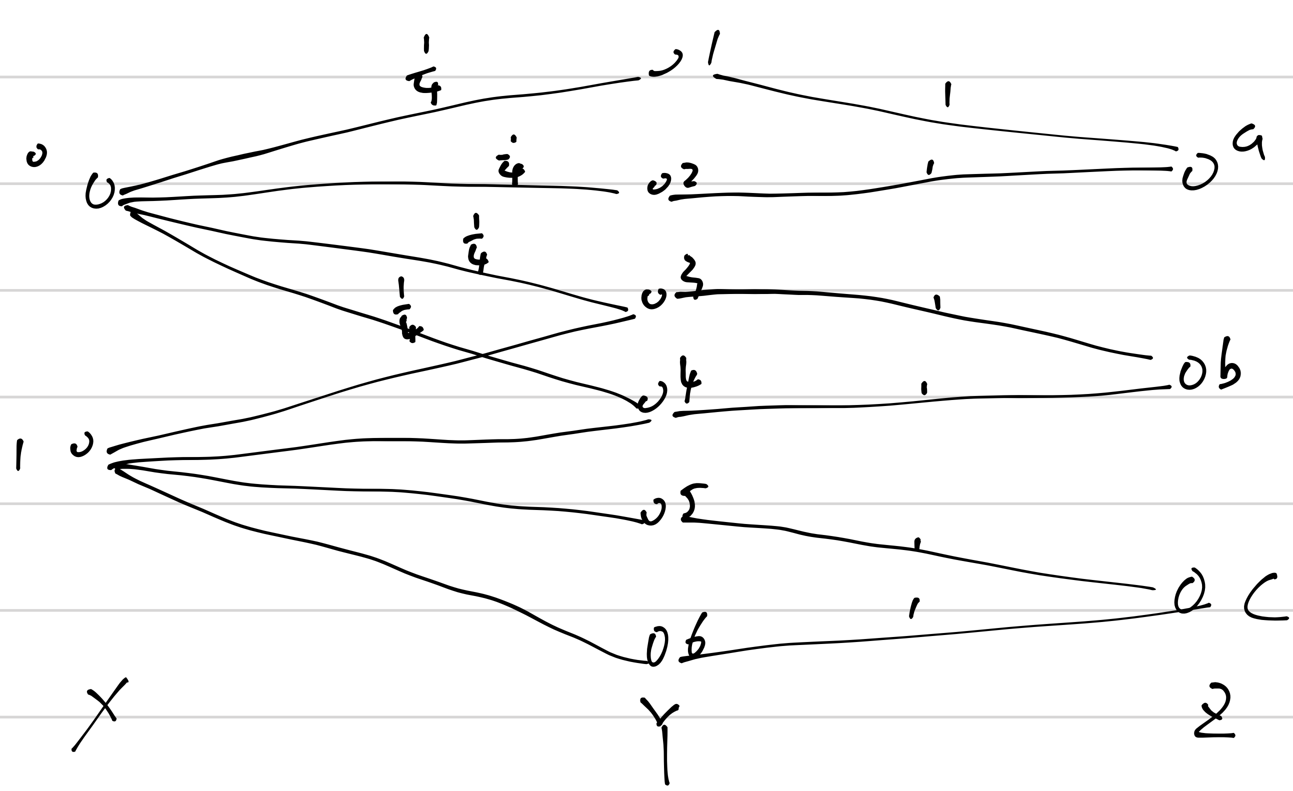
Consider the channel shown below where all the transition probabilities are $\frac{1}{4}$



(a) Define Z as the following function of Y

$$Z = \begin{cases} a & \text{if } y \in \{0, 1\} \\ b & \text{if } y \in \{2, 3\} \\ c & \text{if } y \in \{4, 5\} \end{cases}$$

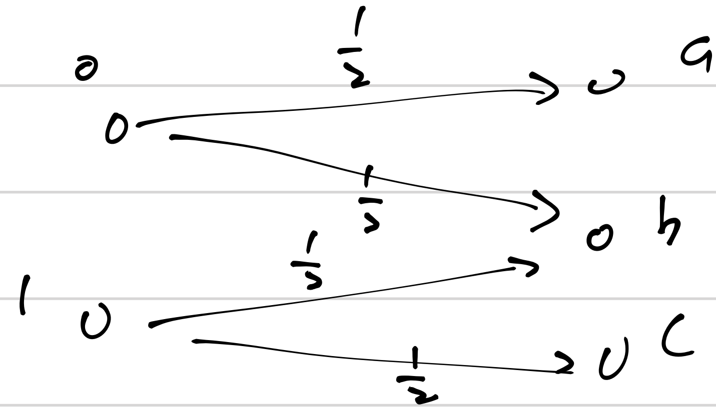
Find the capacity of channel of X, Y



Z is a sufficient statistic of Y for X

$\therefore \; C = \max I(X;Y) = \max I(X;Z)$

for $X \to Y \to Z$, the channel is



it is a binary Erasure channel

$I(X;Z) = 1-\alpha = 1-\frac{1}{2} = \frac{1}{2}$

$\therefore$ The channel capacity is $C = 1-\frac{1}{2} = \frac{1}{2}$ bit/symbol

with maximizing $p(x)$ distribution

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

(b) prove the hint. Z is a sufficient statistic of Y for X.

$X \to Y \to Z$

Z is a sufficient statistic (need to prove)

need to prove $X \to Z \to Y$

need to prove that $p(y|x,z) = p(y|z)$

$p(y|z)$

$Z \backslash Y$	0	1	2	3	4	5
a	$\frac{1}{2}$	$\frac{1}{2}$	0	0	0	0
b	0	0	$\frac{1}{2}$	$\frac{1}{2}$	0	0
c	0	0	0	0	$\frac{1}{2}$	$\frac{1}{2}$

$p(y|x,z)$

$x,z \backslash Y$		0	1	2	3	4	5
$x=0$	$z=a$	$\frac{1}{2}$	$\frac{1}{2}$	0	0	0	0
$x=0$	$z=b$	0	0	$\frac{1}{2}$	$\frac{1}{2}$	0	0
$x=0$	$z=c$	0	0	0	0	0	0
$x=1$	$z=a$	0	0	0	0	0	0
$x=1$	$z=b$	0	0	$\frac{1}{2}$	$\frac{1}{2}$	0	0
$x=1$	$z=c$	0	0	0	0	$\frac{1}{2}$	$\frac{1}{2}$

$$p(y|x,z) = p(x,y,z)/p(x,z)$$

$$p(y|z) = p(y,z)/p(z)$$

$$p(y|z) = \sum_x p(y|z, X=x)\, P_X(x)$$

$$p(y|z=a) = p(y|z=a, x=0)\, P_X(x=0) + p(y|z=a, x=1)\, p(x=1)$$

$$= [\tfrac{1}{2} \ \tfrac{1}{2} \ 0 \ 0 \ 0 \ 0]$$

$$+ [0 \ 0 \ 0 \ 0 \ 0 \ 0] = p(y|z=a)$$

$$p(y|z=b) = p(y|z=b, x=0)\, P_X(x=0) + p(y|z=b, x=1)\, P_X(x=1)$$

$$= P_X(x=0)\, [0 \ 0 \ \tfrac{1}{2} \ \tfrac{1}{2} \ 0 \ 0]$$

$$+ P_X(x=1)\, [0 \ 0 \ \tfrac{1}{2} \ \tfrac{1}{2} \ 0 \ 0]$$

$$= [0 \ 0 \ \tfrac{1}{2} \ \tfrac{1}{2} \ 0 \ 0] = p(y|z=b)$$

same for $p(y|z=c)$

$\therefore\ p(y|z) = p(y|z,x)$

$\therefore$ z is a sufficient statistic of Y for X

7. A channel with two independent looks.

(a) show that $I(X;Y_1,Y_2) = 2I(X;Y_1) - I(Y_1;Y_2)$

$I(X;Y_1,Y_2) = H(Y_1,Y_2) - H(Y_1,Y_2|X)$

Since Y_1 and Y_2 are conditionally independent given X

$\therefore\ H(Y_1,Y_2) = H(Y_1) + H(Y_2) - I(Y_1,Y_2)$

$H(Y_1,Y_2|X) = H(Y_1|X) + H(Y_2|X)$

$I(X;Y_1,Y_2) = H(Y_1) + H(Y_2) - I(Y_1,Y_2) - H(Y_1|X) + H(Y_2|X)$

$= I(X;Y_1) + I(X;Y_2) - I(Y_1;Y_2)$

Since, Y_1, Y_2 are conditionally identically distributed given X

$I(X;Y_1,Y_2) = 2I(X;Y_1) - I(Y_1;Y_2)$ ㅁ

(b) Conclude that the capacity of the channel

①

is less than or equal to twice the capacity of the channel

$X \to \boxed{\quad} \to Y_1$ ②

For channel ②

$$C_2 = \max I(X;Y_1)$$

For channel ①

$$C_1 = \max I(X;Y_1, Y_2)$$

$$= \max 2I(X;Y_1) - I(Y_1;Y_2)$$

$$\le \max 2I(X;Y_1) \qquad (I(Y_1;Y_2) \ge 0)$$

$$= 2C_2$$

$\therefore$ The capacity of channel $X \to \boxed{\quad} \begin{smallmatrix} \to Y_1 \\ \to Y_2 \end{smallmatrix}$ is less than or equal to the twice the capacity of $X \to \boxed{\quad} \to Y_1$

8. Mutual information for correlated normals.
   Find the mutual information $I(X;Y)$
   where $\begin{pmatrix} X \\ Y \end{pmatrix} \sim N\left(0, \begin{bmatrix} \sigma^2 & \rho\sigma^2 \\ \rho\sigma^2 & \sigma^2 \end{bmatrix}\right)$

   for $\rho = 1$, $\rho = 0$, $\rho = -1$

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

$$K = \begin{bmatrix} \sigma^2 & \rho\sigma^2 \\ \rho\sigma^2 & \sigma^2 \end{bmatrix} \quad |K| = \sigma^4(1-\rho^2)$$

$$h(X,Y) = \frac{1}{2}\log(2\pi e)^2 |K| = \frac{1}{2}\log(2\pi e)^2\sigma^4(1-\rho^2)$$

entropy of a multivariate normal Gaussian

$$h(X) = h(Y) = \frac{1}{2}\log 2\pi e\sigma^2$$

$$I(X;Y) = h(X) + h(Y) - h(X;Y)$$
$$= \frac{1}{2}\log 2\pi e\sigma^2 + \frac{1}{2}\log 2\pi e\sigma^2 - \frac{1}{2}\log(2\pi e)^2\sigma^4(1-\rho^2)$$
$$= \frac{1}{2}\log\frac{(2\pi e)^2\sigma^4}{(2\pi e)^2\sigma^4(1-\rho^2)} = -\frac{1}{2}\log(1-\rho^2)$$

$$\therefore I(X;Y) = \frac{1}{2}\log(1-\rho^2)$$

① $\rho = 1$ In this case $X = Y$, if we know X, we know Y, the mutual information $I(X;Y)$ goes to infinity as $\rho = 1$ $\left(\frac{1}{2}\log(1-1) \to \infty\right)$

② $\rho = 0$ $I(X;Y) = \frac{1}{2}\log(1-0) = 0$

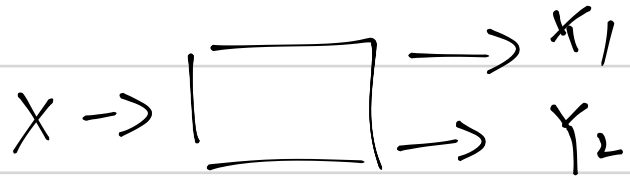
In this case, X, Y are independent, the mutual information is 0.

③ $p=-1$, In this case $X=-Y$, again, if we know X, we know Y

$$I(X;Y) = -\frac{1}{2}\log(1-p^2) \Rightarrow \infty \quad \text{when } p=-1$$

The mutual information is infinity.

9. The two-look Gaussian Channel.



$Y=(Y_1, Y_2)$

$$Y_1 = X + Z_1$$
$$Y_2 = X + Z_2$$

with power constraint P on X and $(Z_1, Z_2) \sim N_2(0, K)$

$$K = \begin{bmatrix} N & Np \\ Np & N \end{bmatrix}$$

find the capacity C for

① $p=1$ ② $p=0$ ③ $p=-1$

$$C = \max I(X; Y_1, Y_2)$$

The capacity is achieved when X is Gaussian input

$$X \sim N(0, P)$$

$$C = h(Y_1, Y_2) - h(Y_1, Y_2 | X)$$
$$= h(Y_1, Y_2) - h(Z_1, Z_2 | X)$$

$= h(Y_1, Y_2) - h(Z_1, Z_2)$ $\quad Z$ is independent of X

$(Z_1, Z_2) = \mathcal{N}\left(0, \begin{bmatrix} N & \rho N \\ \rho N & N \end{bmatrix}\right)$

From the previous problem, we know that

$h(Z_1, Z_2) = \frac{1}{2}\log(2\pi e)^2 |K| = \frac{1}{2}\log(2\pi e)^2 N^2 (1-\rho^2)$

$Y_1 = X + Z_1 \qquad Y_2 = X + Z_2$

$Y_1 \sim (0, P+N) \quad Y_2 \sim (0, P+N) \qquad \therefore Y_1, Y_2 \sim \mathcal{N}\left(0, \begin{bmatrix} P+N & P+\rho N \\ P+\rho N & P+N \end{bmatrix}\right)$

$E(Y_1, Y_2) : E[(X+Z_1)(X+Z_2)]$

$= E[X^2 + X(Z_1+Z_2) + Z_1 Z_2] = P + E[Z_1 Z_2] = P + \rho N$

$h(Y_1, Y_2) = \frac{1}{2}\log(2\pi e)^2 |K|$

$= \frac{1}{2}\log(2\pi e)^2 [(P+N)^2 - (P+\rho N)^2]$

$= \frac{1}{2}\log(2\pi e)^2 [N^2(1-\rho^2) + 2PN(1-\rho)]$

$\therefore C = h(Y_1, Y_2) - h(Z_1, Z_2)$

$= \frac{1}{2}\log \frac{(2\pi e)^2 [N^2(1-\rho^2) + 2PN(1-\rho)]}{(2\pi e)^2 [N^2(1-\rho^2)]}$

$= \frac{1}{2}\log\left[1 + \frac{2P}{N(1+\rho)}\right]$

(a) $\rho = 1 \qquad C = \frac{1}{2}\log\left[1 + \frac{P}{N}\right] \qquad Y_1 = Y_2,$

therefore, the capacity of this channel is a single look channel

(b) $\rho=0$ $\quad C=\frac{1}{2}\log\left[1+\frac{2P}{N}\right]$

this corresponds to using twice the power in a single look.

(c) $\rho=-1$, $\quad C=\frac{1}{2}\log\left[1+\frac{2P}{N\cdot 0}\right]=\infty$

In this case, the capacity is infinite, if we add Y_1 and Y_2, we can recover X exactly.