<u>Def</u> discrete channel : $\mathcal{X} \to \mathcal{Y}$ (input alphabet to output alphabet)

$p(y|x)$: output symbol is y when input is x

memoryless : if the probability distribution of the output depends only on the current input. Conditionally independent of previous channel inputs or outputs.

<u>Def</u> 'Information' channel capacity of a discrete memoryless channel

$$C = \max_{p(x)} I(X:Y)$$

for all posible input distribution $p(x)$.

e.g.

Noiseless Binary Channel

$$C = \max_{p(x)} I(X:Y)$$

$$X \quad \begin{matrix} 0 \xrightarrow{1} 0 \\ 1 \xrightarrow{1} 1 \end{matrix} \quad Y$$

$$I(X:Y) = H(Y) - H(Y|X)$$

$$= \mathbb{E}_{Y\sim p}[-\log p(Y)]$$

$$- \mathbb{E}_{(X,Y)\sim p}[-\log (Y|X)]$$

$$p(y|x) = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \qquad // \; 0$$

$$= H(Y)$$

$$C = \max_{p(x)} H(Y) = 1$$

uniform distribution

Noisy Channel with nonoverlapping Output

$$X \quad \begin{matrix} 0 \xrightarrow{\frac{1}{2}} 1 \\ 0 \xrightarrow{\frac{1}{2}} 2 \\ 1 \xrightarrow{\frac{1}{3}} 3 \\ 1 \xrightarrow{\frac{2}{3}} 4 \end{matrix} \quad Y$$

$$I(X:Y) = H(X) - H(X|Y)$$

$$= \quad ''$$

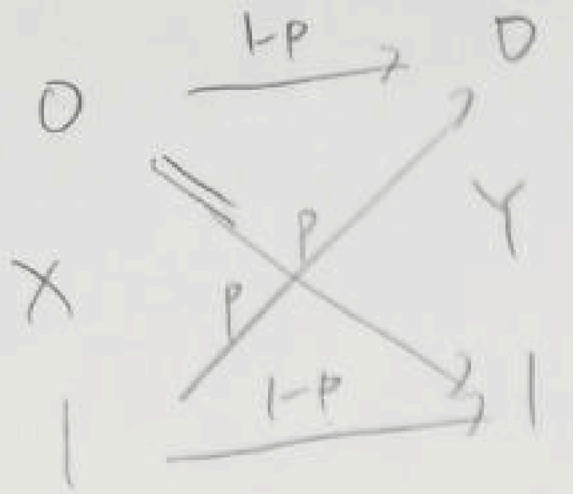
$$- \mathbb{E}_{(X,Y)\sim p}[-\log (X|Y)]$$

$$p(y|x) = \begin{bmatrix} \frac{1}{2} & \frac{1}{2} & 0 & 0 \\ 0 & 0 & \frac{1}{3} & \frac{2}{3} \end{bmatrix} \qquad // \; 0$$

$$= H(X)$$

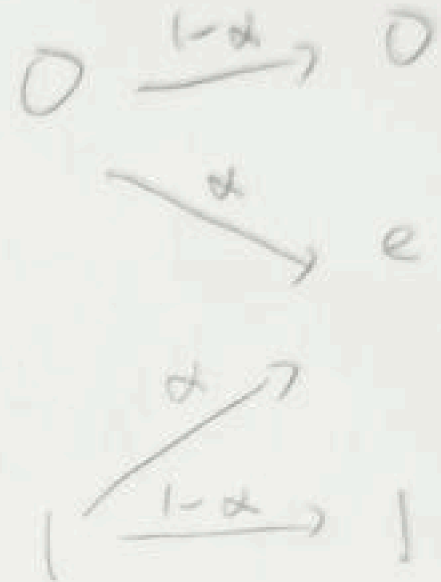
$$C = \max_{p(x)} H(X) = 1$$

Binary Symmetric Channel



$$
\begin{aligned}
I(X;Y) &= H(Y) - H(Y|X) \\
&= H(Y) - \sum p(x) H(Y|X=x) \\
&= H(Y) - \sum p(x) H(p) \\
&= H(Y) - H(p) \\
&\leq 1 - H(p)
\end{aligned}
$$

Binary erasure Channel



$$
\max_{p(x)} I(X;Y) = \max_{p(x)} H(Y) - H(Y|X)
$$

$$
= \underline{\max_{p(x)} H(Y)} - H(\alpha)
$$

cannot be uniform by any $p(x)$

Recall Chain Rule

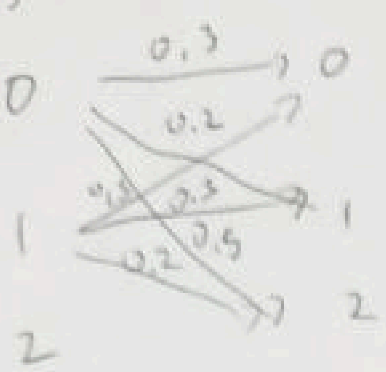
$$
H(X,Y) = H(X) + H(Y|X)
$$

Let $P_r(X=1) = \pi$

$$
\begin{aligned}
H(Y) &= H(Y=0, Y=1, Y=e) \\
&= H\big((1-\pi)(1-\alpha), \pi(1-\alpha), \alpha\big) \\
&= H(\alpha) + (1-\alpha) H(\pi)
\end{aligned}
$$

$$
\begin{aligned}
\max_{p(x)} I(X;Y) &= \max_{p(x)} H(Y) - H(\alpha) \\
&= \max_{p(x)} (1-\alpha) H(\pi) \\
&= 1 - \alpha
\end{aligned}
$$

Symmetric Channels



$$p(y|x) = \begin{bmatrix} 0.3 & 0.2 & 0.5 \\ 0.5 & 0.3 & 0.2 \\ 0.2 & 0.5 & 0.3 \end{bmatrix}$$

Def symmetric: all the row and columns of the probability transition matrix are permutations of each other

$$\begin{aligned} I(X:Y) &= H(Y) - H(Y|X) \\ &= H(Y) - H(r) \\ &\le \log|\mathcal{Y}| - H(r) \end{aligned}$$

equality at uniform (We can achive this by uniform $P(x)$!)

$\log 3 - H(0.5, 0.3, 0.2)$

Def weak symmetric: row is permutation but column is only have the same sum.

Similary, $\log|\mathcal{Y}| - H(r)$

Properties

1. $C \geq 0$ since $I(X;Y) \geq 0$
2. $C \leq \log|\mathcal{X}|$ since $C = \max I(X;Y) \leq \max H(X) = \log|\mathcal{X}|$
3. $C \leq \log|\mathcal{Y}|$
4. $I(X;Y)$ is a continuous funtion of $p(x)$
5. $I(X;Y)$ is a concave function of $p(x)$

Channel Coding

For input n sequence there is $2^{nH(Y|X)}$ possible Y sequences.

What to have disjoint sets.

$$2^{n(H(Y)-H(Y|X))} = 2^{nI(X;Y)}$$

at most $\approx 2^{nI(X;Y)}$ distinquishable sequences

$$\underset{}{W} \xrightarrow[\text{encode}]{} X_n \xrightarrow[\text{channel}]{} Y_n \xrightarrow[\text{decode}]{} \hat{W}$$

Def nth extension of the discrete memoryless channel is

$(\mathcal{X}^n, p(y^n, x^n), \mathcal{Y}^n)$, where

$$p(y_k | x^k, y^{k-1}) = p(y_k | x_k)$$

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

Def An (M, n) code for $(\mathcal{X}, p(y|x), \mathcal{Y})$ is consists of the following

1. An index set $\{1, 2, \cdots, M\}$
2. An encoding function $X^n : \{1, 2, \cdots, M\} \to \mathcal{X}^n$

   code words as $x^n(M)$

   set of code words is code-book

3. Decoding function

   $$g : \mathcal{Y}^n \to \{1, 2, \cdots, M\}$$

i index conditional error

$$\lambda_i = P_r(g(Y^n) \neq i \mid X^n = x^n(i)) = \sum_{y^n} p(y^n | x^n(i)) I(g(y^n) \neq i)$$

maximum error $\lambda^{(n)} = \max \lambda_i$

$$P_e^{(n)} \leq \lambda^{(n)}$$

average error $P_e^{(n)} = \frac{1}{M} \sum_{i=1}^{M} \lambda_i$

rate of transmission $R = \frac{\log(M)}{n}$

rate R of (M, n) code

rate R is achievable if there exists a sequence of $(\lceil 2^{nR} \rceil, n)$ codes such that $\lambda^{(n)}$ tends to 0 as $n \to \infty$

capacity of a channel is the supremum of all achievable rates

Def jointly typical sequences $\{x^n, y^n\}$

with distribution $p(x,y)$

$$A_\epsilon^{(n)} = \{ (x^n, y^n) \in \mathcal{X}^n \times \mathcal{Y}^n :$$
$$\left| -\frac{1}{n} \log p(x^n) - H(X) \right| < \epsilon,$$
$$\left| -\frac{1}{n} \log p(y^n) - H(Y) \right| < \epsilon,$$
$$\left| -\frac{1}{n} \log p(x^n, y^n) - H(X,Y) \right| < \epsilon \}$$

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

Theorem 7.6.1

1. $P_r((X^n, Y^n) \in A_\epsilon^{(n)}) \to 1$ as $n \to \infty$
2. $|A_\epsilon^{(n)}| \leq 2^{n(H(X,Y)+\epsilon)}$
3. If $(\tilde{X}^n, \tilde{Y}^n) \sim p(x^n)p(y^n)$ (X, Y independent)

$$P_r((\tilde{X}^n, \tilde{Y}^n) \in A_\epsilon^{(n)}) \leq 2^{-n(I(X;Y)-3\epsilon)}$$

for large n

$$P_r((\tilde{X}^n, \tilde{Y}^n) \in A_\epsilon^{(n)}) \geq (1-\epsilon)2^{-n(I(X;Y)+3\epsilon)}$$

Channel Coding

- Allow an arbitrarily small but nonzero probability of error
- Using the channel many times in succession, so that the law of large numbers take place
- Calculating the average of the probability of error over a random choice of codebooks and which then can be used to show the existence of at least one good code

Theo 7.7.1 Channel coding theorem

For a channel all rates below C are achievable

For ever $R<C$, there exists a sequence of $(2^{nR}, n)$ codes with maximum prob of error $\lambda^{(n)} \to 0$

$\longleftrightarrow$ any $(2^{nR}, n)$ codes with $\lambda^{(n)} \to 0$ must have $R \le C$

$$
\begin{aligned}
P_r(\mathcal{E}) &= \sum_{C} P_r(C) P_e^{(n)}(C) \\
&= \sum_{C} P_r(C) \frac{1}{2^{nR}} \sum_{w=1}^{2^{nR}} \lambda_w(C) \\
&= \frac{1}{2^{nR}} \sum_{w=1}^{2^{nR}} \sum_{C} P_r(C) \lambda_w(C) \\
&= \sum_{C} P_r(C) \lambda_1(C) \\
&= P_r(\mathcal{E} \mid W=1)
\end{aligned}
$$

$$E_i = \{ X^n(i), Y^n) \text{ is in } A_\epsilon^{(n)} \}$$

$$P_r(\mathcal{E} / W=1) = P(E_1^c \cup \cdots \cup E_{2^{nR}} \mid W=1) \le P(E_1^c \mid W=1) + \sum_{i=2}^{2^{nR}} P(E_i \mid W=1)$$