2/10/14 class 8

last class $\begin{cases} \text{definition of entropy} \\ \text{encoding : Huffman coding (LZ coding)} \end{cases}$

9.5 Discrete Memoryless channels

$$\mathcal{X} = \begin{matrix} x_0 \\ \vdots \\ x_{J-1} \end{matrix} \qquad X \to \boxed{P(y_k|x_j)} \to Y \qquad \left.\begin{matrix} y_0 \\ y_1 \\ \vdots \\ y_{K-1} \end{matrix}\right\} \mathcal{Y}$$

transition probability

$$0 \le P(y_k|x_j) \le 1 \quad \forall j, k$$

$$P = \begin{bmatrix} P(y_0|x_0) & P(y_1|x_0) & \cdots & P(y_{K-1}|x_0) \\ P(y_0|x_1) & P(y_1|x_1) & \cdots & P(y_{K-1}|x_1) \\ \vdots & \vdots & & \\ P(y_0|x_{J-1}) & P(y_1|x_{J-1}) & \cdots & P(y_{K-1}|x_{J-1}) \end{bmatrix}$$

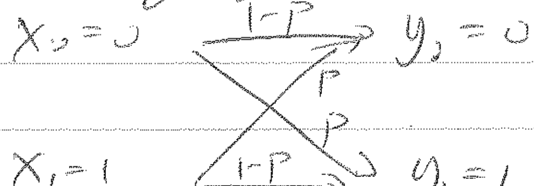
$$\sum_{k=0}^{K-1} P(y_k|x_j) = 1, \quad \forall j$$

Source distribution $P(x_j) = P(X = x_j)$ for $j = 0 \cdots J-1$

$$P(x_j, y_k) = P(y_k|x_j)P(x_j)$$

$$P(y_k) = \sum_{j=0}^{J-1} P(y_k|x_j)P(x_j) \quad k = 0, 1 \cdots K-1$$

EX 9.4 Binary Symmetric Channel



$P_{10} = P_{01} = P$

9.6 Mutual Information

Given channel output Y as a noisy version of channel input X, and $H(\mathcal{X})$ is entropy of X, how to measure the uncertainty about X after observing Y.

$$H(\mathcal{X}|Y=y_k) = \sum_{j=0}^{J-1} P(X_j|y_k) \log\left[\frac{1}{P(X_j|y_k)}\right]$$

suppose $H(\mathcal{X}|Y=y_k)$ with probability $P(y_k)$

we have conditional entropy:

$$H(\mathcal{X}|\mathcal{Y}) = \sum_{k=0}^{K-1} H(\mathcal{X}|Y=y_k)\, P(y_k)$$

The amount of uncertainty remaining about the channel input after the channel output has been observed

$$= \sum_{k=0}^{K-1} \sum_{j=0}^{J-1} P(X_j|y_k)\, P(y_k) \log_2\left[\frac{1}{P(X_j|y_k)}\right]$$

$$= \sum_{k=0}^{K-1} \sum_{j=0}^{J-1} P(X_j, y_k) \log_2\left[\frac{1}{P(X_j|y_k)}\right] \quad \text{— ①}$$

Definition of mutual information

$$I(X;y) = H(\mathcal{X}) - H(\mathcal{X}|y) \quad \text{— ②}$$

the uncertainty about the channel input that is resolved by observing the channel output

Property 1: $I(X;y) = I(y;X)$

proof:

$$H(X) = \sum_{j=0}^{J-1} P(X_j) \log_2\left[\frac{1}{P_2(X_j)}\right]$$

$$= \sum_{j=0}^{J-1} P(X_j) \log\left[\frac{1}{P(X_j)}\right] \sum_{k=0}^{K-1} P(y_k|X_j)$$

$$= \sum_{j=0}^{J-1} \sum_{k=0}^{K-1} P(y_k|X_j)\, P(X_j) \log\left(\frac{1}{P(X_j)}\right)$$

$$= \sum_{j=0}^{J-1} \sum_{k=0}^{K-1} P(X_j, y_k) \log\left[\frac{1}{P(X_j)}\right] \quad \text{— ③}$$

substitute ③, ① to ②

$$④ \text{ — } I(X;y) = \sum_{j=0}^{J-1} \sum_{k=0}^{K-1} P(X_j, y_k) \log\left(\frac{P(X_j|y_k)}{P(X_j)}\right)$$

From Baye's rule $\dfrac{P(X_j|y_k)}{P(X_j)} = \dfrac{P(y_k|X_j)}{P(y_k)}$

$$I(X;y) = \sum_{k=0}^{K-1} \sum_{j=0}^{J-1} P(X_j, y_k) \log\left[\frac{P(y_k|X_j)}{P(y_k)}\right]$$

$$= I(y;X)$$

Property 2 $\quad I(X;Y) \geq 0$

proof: $\quad P(x_j|y_k) = \dfrac{P(x_j, y_k)}{P(y_k)}$

put into ④ $\quad I(X;Y) = \sum_{j=0}^{J-1} \sum_{k=0}^{K-1} P(x_j, y_k) \log\left(\dfrac{P(x_j, y_k)}{P(x_j)P(y_k)}\right)$

≥ 0

equal iff $\quad P(x_j, y_k) = P(x_j)\, P(y_k)$ ← input and output independent

physical meaning, channel will make information more chaotics.

Property 3. $\quad I(X;Y) = H(X) + H(Y) - H(X,Y)$.

where joint entropy $H(X,Y)$ ~~is~~

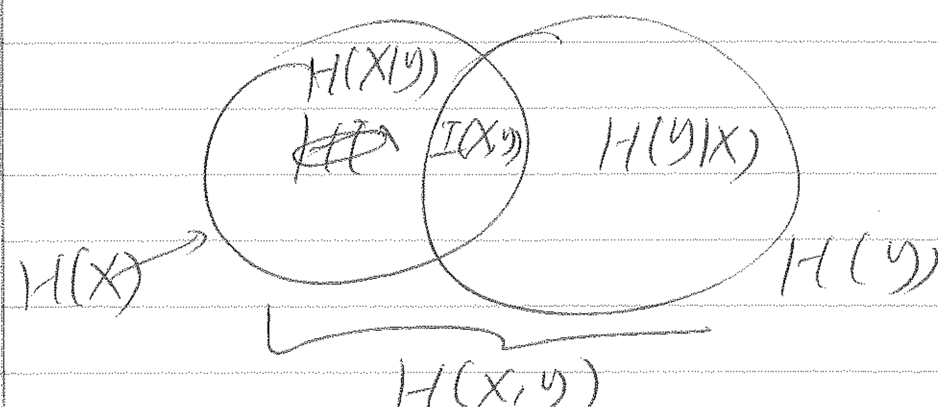
$$H(X,Y) = \sum_{j=0}^{J-1} \sum_{k=0}^{K-1} P(x_j, y_k) \log\left(\frac{1}{P(x_j, y_k)}\right)$$

Proof: $\quad H(X,Y) = \sum_{j=0}^{J-1} \sum_{k=0}^{K-1} P(x_j, y_k) \log_2\left(\dfrac{P(x_j)P(y_k)}{P(x_j, y_k)}\right)$

$+ \sum_{j=0}^{J-1} \sum_{k=0}^{K-1} P(x_j, y_k) \log_2\left(\dfrac{1}{P(x_j)P(y_k)}\right)$

$-I(X;Y) + \sum_{j=0}^{J-1} \log\left(\dfrac{1}{P(x_j)}\right) \underbrace{\sum_{k=0}^{K-1} P(x_j, y_k)}_{1} + \sum_{k=0}^{K-1} \log_2\left[\dfrac{1}{P(y_k)}\right] \underbrace{\sum_{j=0}^{J-1} P(x_j, y_k)}_{1}$

$= H(X) + H(Y) - I(X;Y)$



9.7 channel Capacity

maximize $I(X;Y)$ with respect to $P(X_i)$

$$C = \max_{\{P(X_i)\}} I(X;Y) \quad \text{bit per channel use.}$$

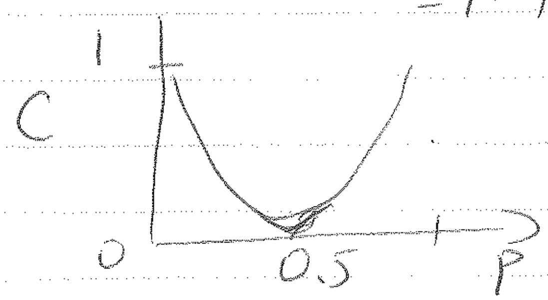
Example Binary symmetric channel

$I(X;Y)$ is maximized when $H(X)$ is maximized with $P(X_0) = P(X_1) = 1/2$

$$C = I(X;Y)\big|_{P(X_0)=P(X_1)=1/2}$$

$$\begin{cases} P(y_0|X_1) = P(y_1|X_0) = P \\ P(y_0|X_0) = P(y_1|X_1) = 1-P \end{cases}$$

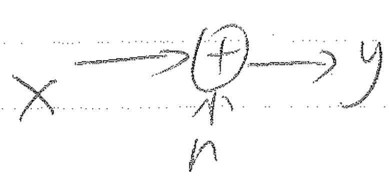
$$C = 1 + P\log_2 P + (1-P)\log(1-P)$$
$$= 1 - H(P)$$



- when channel is noise free $P = 0$ (or $P = 1$)
  C is maximized and equal to information in each channel input $H(X) = 1$, with $P(X_0) = P(X_1) = 1/2$

- when ~~channel~~ $P = 1/2$, $C = 0$. the channel is useless just toss a coin in the receiver.

For AWGN channel



$C = W\log(1+SNR)$

why, next class