# DC Sheet I & Highlights

⇒ We consider the output of a communication source as a random variable $S$ taking symbols from a finite alphabet.

- $S : \{S_0, S_1, \dots, S_{K-1}\}$ with Probabilities

$$P(S = S_i) = P_i \quad \text{for } 0 \leq i < K \qquad \left(\sum_{i=0}^{K-1} P_i = 1\right)$$

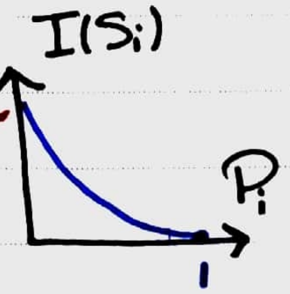
* The information content in a message $S_i$ occuring with Probability $P_i$

$$I(S_i) = \log_2\left(\frac{1}{P(S_i)}\right) = -\log_2 P(S_i)$$

$$= -\log_2 P_i$$

← The lower the Probability the more information

too much Surprise (new info)




* The entropy of the Source

$$H(S) = E(I(S)) = \sum_{0}^{K-1} \log_2\left(\frac{1}{P_i}\right) P_i$$

average amount of info assoc. with Source

$$= -\sum_{i=0}^{K-1} P_i \log_2 P_i$$

$$A \rightarrow \boxed{P(b_j|a_i)} \rightarrow B$$

$A$ takes $\{a_1, a_2, \dots, a_r\}$

$B$ takes $\{b_1, b_2, \dots, b_s\}$

Channel maps each input symbol $a_i$ to an output symbol $b_j$

* Thus it takes a $r \times s$ transition matrix to characterize channel behaviour

$$T_{i,j} = P(b_j|a_i)$$

$$\begin{array}{cc} b_1 \; b_2 \dots & \\ \begin{pmatrix} \quad & \quad \end{pmatrix} & \begin{matrix} a_1 \\ a_2 \\ \vdots \end{matrix} \end{array}$$

→ Given the probability distribution of $A$ as well we can compute $rs + rs + r$ more quantities

$rs \rightarrow P(a_i, b_j)$, $rs \rightarrow P(a_i|b_j)$, $r \rightarrow P(b_j)$

- $P(b_j) = \sum_{i=1}^{r} P(b_j, a_i) = \sum_{i=1}^{r} P(b_j|a_i) P(a_i)$
  - Sum the columns in $P(A,B)$

- $P(a_i|b_j) = \dfrac{P(b_j|a_i) P(a_i)}{P(b_j)}$
  - Divide each column by $b_j$ in $P(A,B)$
  - Send $a_i$ given that $b_j$ was received.

- $P(a_i, b_j) = P(b_j|a_i) P(a_i) = P(a_i|b_j) P(b_j)$
  1. Multiply each row in $T$ by $a_i \rightarrow P(A,B)$ matrix

Given a binary Channel

$$T = \begin{array}{cc} b_1 & b_2 \end{array}$$
$$T = \begin{pmatrix} 2/3 & 1/3 \\ 1/10 & 9/10 \end{pmatrix} \begin{array}{c} a_1 \\ a_2 \end{array} , \quad P(A) = \begin{pmatrix} 1/3 \\ 2/3 \end{pmatrix} \begin{array}{c} a_1 \\ a_2 \end{array}$$

Then

$$\Rightarrow \begin{pmatrix} P(b_1,a_1) & P(b_2,a_1) \\ P(b_1,a_2) & P(b_2,a_2) \end{pmatrix} = \begin{pmatrix} 2/3 \cdot \frac{1}{3} & 1/3 \cdot \frac{1}{3} \\ 1/10 \cdot \frac{2}{3} & 9/10 \cdot \frac{2}{3} \end{pmatrix}$$

$$= \begin{pmatrix} 2/9 & 1/9 \\ 1/15 & 3/5 \end{pmatrix}$$

. as $P(a_i, b_j) = P(b_j | a_i) P(a_i)$

$$\Rightarrow \begin{pmatrix} P(b_1) \\ P(b_2) \end{pmatrix} = \begin{pmatrix} 2/9 + 1/15 \\ 1/9 + 3/5 \end{pmatrix} = \begin{pmatrix} 13/45 \\ 32/45 \end{pmatrix}$$

. as $P(b_j) = \sum_{i=1}^{r} P(a_i, b_j)$

$$\Rightarrow \begin{pmatrix} P(a_1|b_1) & P(a_1|b_2) \\ P(a_2|b_1) & P(a_2|b_2) \end{pmatrix} = \begin{pmatrix} \frac{2/9}{13/45} & \frac{1/9}{32/45} \\ \frac{1/15}{13/45} & \frac{3/5}{32/45} \end{pmatrix}$$

. as $P(a_i|b_j) = P(a_i,b_j) / P(b_j)$

Scanned with CamScanner

$$= \begin{pmatrix} 10/13 & 5/32 \\ 3/13 & 27/32 \end{pmatrix}$$

and now we've computed everything pertaining to probabilities.

## * Entropies

### → Source Entropy

$$H(A) = -\sum_{i=1}^{r} P(a_i)\log_2 P(a_i)$$

$$= -\left(\frac{1}{3} \times \log_2 \frac{1}{3} + \frac{2}{3} \times \log_2 \frac{2}{3}\right)$$

$$= 0.9183 \text{ bits}$$ # Apriori Entropy

### → Destination Entropy

$$H(B) = -\sum_{i=1}^{s} P(b_i)\log_2 P(b_i)$$

$$= -\left(\frac{13}{45} \cdot \log_2 \frac{13}{45} + \frac{32}{45} \cdot \log_2 \frac{32}{45}\right)$$

$$= 0.8673$$

# Joint Entropy

$$H(A,B) = -\sum_{i,j} P(a_i, b_j) \log_2 P(a_i, b_j)$$

$$P(A,B) = \begin{array}{cc} \begin{pmatrix} 2/9 & 1/9 \\ 1/15 & 3/5 \end{pmatrix} & \begin{array}{c} a_2 \\ a_1 \end{array} \\ \begin{array}{cc} b_1 & \quad b_2 \end{array} & \end{array}$$

$$= -\left( \frac{2}{9} \times \log_2 \frac{2}{9} + \frac{1}{9} \times \log_2 \frac{1}{9} + \frac{1}{15} \times \log_2 \frac{1}{15} + \frac{3}{5} \times \log_2 \frac{3}{5} \right)$$

$$= 1.537 \text{ bits}$$

# Conditional Entropy

$P(b_j, a_i) P(a_i)$

- $$H(B|A) = -\sum_{i,j} P(b_j, a_i) \log_2 P(b_j | a_i)$$

$$= H(A,B) - H(A)$$

$$= 1.537 - 0.9183 = 0.6187 \text{ bits}$$

$P(b_j, a_i) P(b_j)$

- $$H(A|B) = -\sum_{i,j} P(b_j, a_i) \log_2 P(a_i | b_j)$$

$$= H(A,B) - H(B)$$

$$= 1.537 - 0.8673 = 0.6697$$

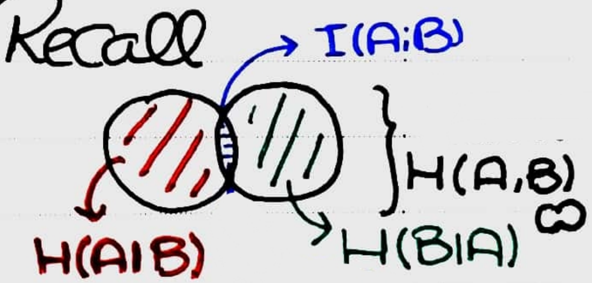
# Mutual Information

$$
\begin{aligned}
I(A;B) &= H(A) - H(A|B) \\
&= H(B) - H(B|A) \\
&= H(A) + H(B) - H(A,B) \\
&= \sum_{i,j} P(b_j, a_i) \log_2 \frac{P(b_j|a_i)}{P(b_j)}
\end{aligned}
$$

$$
= 0.9183 + 0.8673 - 1.537
$$

$$
= 0.2486
$$

Recall



# Interpreting Quantities

$I(S_i)$ • the amount of new information associated with the message

Meh. • As $I\uparrow$, we'll need more bits to measure information
• It's also the minimum no. of bits needed to encode the msg

$H(S)$ • The average amount of information associated with the source
• Average amount of uncertainty per message (randomness in S)
• Minimum no. of bits needed to encode a message from the source such that its uniquely decodable.

$P(b_j|a_i)$ • The probability of receiving $b_j$ given that $a_i$ was sent
→ Characterizes channel behavior

$P(a_i)$ • The probability distribution of source's symbols "how the channel is used"

$P(b_j)$ • Output symbols probability distribution

$P(a_i|b_j)$ • $a_i$ sent given $b_j$ was received

$P(a_i, b_j)$ • Probability of sending symbol $a_i$ but getting $b_j$

* Notice that in the past problem we started by getting

$$
\begin{tikzcd}
P(a_i) \arrow[r] \arrow[d] & P(a_i, b_j) \arrow[rrr] \arrow[ddr] \arrow[ddd] & & & H(A,B) \arrow[r, "-H(B)"] \arrow[dr, "-H(A)"] \arrow[dd, "H(A)"', "H(B)"] & H(A|B) \\
H(A) & & & & & H(B|A) \\
P(b_j|a_i) \arrow[uur] & & P(b_j) \arrow[r] & H(B) & I(A;B) & \\
& P(a_i|b_j) & & & &
\end{tikzcd}
$$

$H(A)$ • Average amount of uncertainty at source

$H(B)$ • Average amount of uncertainty at destination

$H(A,B)$ • Average amount of information due to simultaneously observing A, B

$H(A|B)$ • Average amount of uncertainty remaining in A after knowing B was received.

* It's also a measure of information lost in the channel due to its characteristics.

$H(B|A)$ • Average amount of uncertainty in B given that A was transmitted

$I(A;B)$ • The average amount of information due to observing a single output symbol (info - info lost)

• Reduction of uncertainty in A after knowing B was received.

$\max_{\forall P(A)} \{I(A;B)\}$ • The maximum amount of mutual information over all possible input probability distributions for a given discrete memoryless channel

• Channel Capacity (C)

• maximum rate at which info can be reliably transmitted

Ranges)

* $0 \leqslant I(S=S_i) < \infty$

$P_i = 1$ • no surprise

$P_i = 0$ • Big surprise

* $0 \leqslant H(S) \leqslant \log_2 K$

$P=1$ S

$P=\frac{1}{K}$ (K symbols) • most random

* $H(A) \leqslant H(A,B) \leqslant H(A) + H(B)$

A=B

A,B are independent

* $0 \leqslant H(A|B) \leqslant H(A)$

No info lost

All info lost

* $H(A) \geqslant I(A;B) \geqslant 0$

• Derive an expression for maximum source entropy $H(S)$

→ Since entropy quantifies uncertainty. It will be max when all source symbols $S_1, S_2, \dots, S_K$ occur at equal probabilities $\frac{1}{K}$ $\left(\sum_{i}^{K} \frac{1}{K} = 1\right)$

$$H(S) = -\sum_{i=1}^{K} P_i \log P_i = -\sum_{i=1}^{K} \frac{1}{K} \log_2 \frac{1}{K}$$

as $\sum_{i}^{K} \frac{1}{K} = 1$

$$= -\log \frac{1}{K} = \log_2 K \text{ bits}$$

• Derive an expression for channel capacity and mutual information of a binary symmetric channel where

→ $P_0 = P(X = x_0)$ // Sending $x_0$

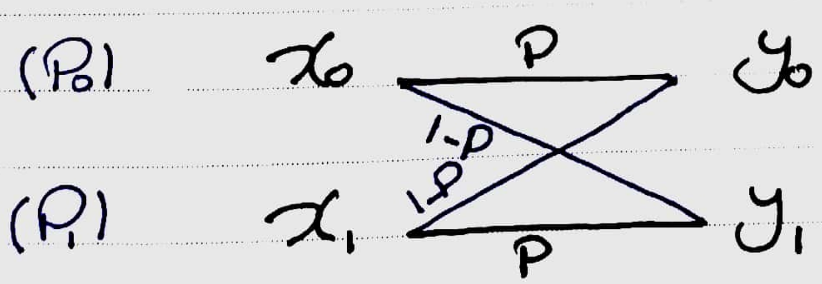
$P_1 = P(X = x_1)$ // Sending $x_1$

$= 1 - P_0$

→ $P$ is the transition probability of the channel $(P(b_1|a_1) = P(b_2|a_2) = P)$ . Here $a \leftrightarrow x$

Sol.

This corresponds to the channel




$$\longleftrightarrow \quad T = \begin{array}{cc} y_0 & y_1 \end{array}$$

$$T = \begin{pmatrix} P & 1-P \\ 1-P & P \end{pmatrix} \begin{array}{c} x_0 \\ x_1 \end{array}$$

$$P(X) = \begin{pmatrix} P_0 \\ P_1 \end{pmatrix}$$

# An expression for mutual information

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

$$H(Y) = -\sum_{i=0}^{1} P(y_i)\log_2 P(y_i)$$

$$H(Y|X) = -\sum_{i,j} P(x_i, y_j)\log_2 P(y_j|x_i)$$

$$\overset{①}{\Rightarrow} \begin{pmatrix} P(x_0, y_0) & P(x_0, y_1) \\ P(x_1, y_0) & P(x_1, y_1) \end{pmatrix} = \begin{array}{cc} y_1 & y_0 & \\ \left(\begin{array}{cc} PP_0 & (1-P)P_0 \\ (1-P)P_1 & PP_1 \end{array}\right) & \begin{array}{c} x_0 \\ x_1 \end{array} \end{array}$$

$$\overset{②}{\Rightarrow} \begin{pmatrix} Z \\ Z_1 \end{pmatrix} = \begin{pmatrix} PP_0 + P_1(1-P) \\ PP_1 + P_0(1-P) \end{pmatrix}$$

where $Z = P(Y = y_0)$, $Z_1 = P(Y = y_1) = 1 - Z$

- we used $P(x_i, y_j) = P(y_j|x_i)P(x_i)$ ①

$$P(y_j) = \sum_{i=0}^{1} P(x_i, y_j) \quad ②$$

Now

$$H(Y) = -(Z\log_2 Z + Z_1\log_2 Z_1)$$
$$= -(Z\log_2 Z + (1-Z)\log_2(1-Z))$$

$$\begin{aligned} H(Y|X) = -(\, & PP_0 \,.\, \log_2 P \\ & + (1-P)P_0 \,.\, \log_2(1-P) \\ & + (1-P)P_1 \,.\, \log_2(1-P) \\ & + PP_1 \,.\, \log_2 P) \end{aligned}$$

$$H(Y|X) = -\Big( \log_2 P \cdot P \cdot (P_0 + P_1) + \log_2 (1-P) \cdot (1-P) \cdot (P_0 + P_1) \Big)$$

Since $P_0 + P_1 = 1$

$$H(Y|X) = -(P \log_2 P + (1-P) \log_2 (1-P))$$

and finally if we define a function

$$h(x) = -(x \log_2 x + (1-x) \log_2 (1-x))$$

then

$$I(X;Y) = \underbrace{h(Z)}_{H(Y)} - \underbrace{h(P)}_{H(Y|X)}$$

.where

$Z = P P_0 + P_1 (1-P)$

Observe that its purely a function in the input's distribution $(P_0, P_1)$ and the channel cics $(P)$

. To obtain an expression of channel capacity

$$C = \max_{\forall P(x)} \{ I(X;Y) \}$$

Start by finding $P(x)$ (i.e. $P_0, P_1$) such that $I$ is maximized

- $I(X;Y) = h(Z) - h(P)$ where
  $Z = PP_0 + P_1(1-P)$
  $h(x) = -(x\log_2 x + (1-x)\log_2(1-x))$

Set $\frac{\partial I}{\partial P_0} = 0$

func. in $P_0$ — not func. in $P_0$

$$\frac{\partial}{\partial P_0}(h(Z) - h(P)) = \frac{\partial h(Z)}{\partial P_0} = 0$$

$$\frac{\partial}{\partial P_0}\left(-(Z\log_2 Z + (1-Z)\log_2(1-Z))\right) =$$

$$-\left((Z'\log_2 Z + Z \cdot \frac{Z'}{Z} \cdot \frac{1}{\ln 2}\right)$$

$$+ (-Z')\log_2(1-Z) + (1-Z) \cdot \frac{-Z'}{1-Z} \cdot \frac{1}{\ln 2}))$$

$$= -Z'\left(\log_2 Z + \frac{1}{\ln 2} - \log_2(1-Z) - \frac{1}{\ln 2}\right)$$

$$= -Z'(\log_2 Z - \log_2(1-Z))$$

By equating to 0.

$$\log_2 Z = \log_2(1-Z) \rightarrow Z = 1 - Z \rightarrow Z = \frac{1}{2}$$

$$PP_0 + P_1(1-P) = \frac{1}{2}$$

$$PP_0 + (1-P_0)(1-P) = \frac{1}{2}$$

$$PP_0 + 1 - P - P_0 + PP_0 = \frac{1}{2}$$

$$P_0(2P-1) - P = \frac{1}{2} - 1$$

$$P_0(2P-1) = P - \frac{1}{2}$$

$$P_0 = \frac{P-1/2}{2P-1} = \frac{P-1/2}{2(P-1/2)} = \frac{1}{2}$$

- Thus, $P_0 = P_1 = \frac{1}{2}$ is the Probability distribution that maximizes $I$

$$C = \max_{\forall P(x)} \{ I(X;Y) \} = I(X;Y)\Big|_{P_0=P_1=\frac{1}{2}}$$

$$= I(X;Y)\Big|_{z=\frac{1}{2}} \quad \text{(1st line)}$$

$$\overbrace{I(X;Y)\Big|_{z=\frac{1}{2}}}^{C} = h\left(\frac{1}{2}\right) - h(P)$$

- $h\left(\frac{1}{2}\right) = -\left(\frac{1}{2}\log_2\frac{1}{2} + \frac{1}{2}\log_2\frac{1}{2}\right) = 1$

Thus, $C = 1 - h(P)$ # An expression for channel capacity

## Mutual Information

→ By Shannon's 1st theorem, need at least $H(A)$ bits on average to convey one symbol of $A$

→ On average $H(A|B)$ bits of info are lost in the channel.

→ Thus, on average the observation of a single output gives us

$$H(A) - H(A|B)$$

bits of information.

- Define such quantity as the mutual information of A & B (of the channel also)

$$I(A;B) = H(A) - H(A|B)$$

$$0 \leq I(A;B) \leq H(A)$$

Worst case (no mutual information)

Best case → All uncertainity is gone. (no information is lost) → Channel adds no randomness

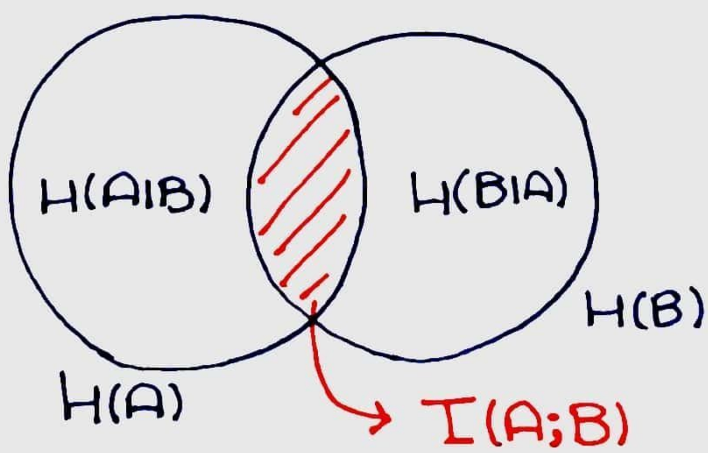
Properties) $I(A;B) \geq 0$

Can be easily shown. ← $I(A;B) = I(B;A)$

$I(A;B) = H(A) - H(A|B)$

- It follows by substitution that

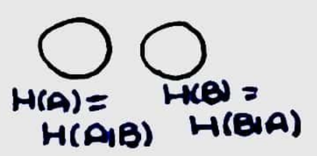
$$
\begin{aligned}
I(A;B) &= H(A) - H(A|B) \\
&= H(B) - H(B|A) \\
&= H(A) + H(B) - H(A,B)
\end{aligned}
$$





- If A & B are independent (worst case)

$I(A;B) = 0$




- else if $A = B$

$I(A;B) = H(A) = H(B)$