Information Theory and Coding (Lec-01)

Eftekhar Hossain
Lecturer
Dept. of ETE, CUET

Outline

- ➡ Uncertainty and Information
- ➡ Self Information
- ➡ Mutual Information
- ➡ Average Mutual Information

Uncertainty and Information

Consider the following sentences:

- ➡ Tomorrow the sun will rise from the East. (most probable, least information)
- ➡ The phone will ring in the next one hour.
- ➡ It will snow in Dhaka this winter. (least probable, most information)

The three sentences carry different amounts of information.

Occurence of a less probable event conveys more information

Since a lower probability implies a higher degree of uncertainty (and vice versa). a random variable with a higher degree of uncertainty contains more information.

Uncertainty and Information

Consider a discrete random variable X with possible outcomes x_i $, i = 1, 2, 3....n$

The self information of the event $X = x_i$ is defined as

$$\boxed{I(x_i) = log(\frac{1}{P(x_i)}) = -logP(x_i)}$$

- When the base of the logarithm is 2 the units of $I(x)$ are in bits
- When the base is e. the units are in nats (natural units).

Example

- Consider a binary source which tosses a fair coin
- It produces an output equal to 1 if a head appears and a 0 if a tail appears.
- For this source $\boxed{P(1) = P(0) = 0.5}$ The information content of each output from the source

$$\boxed{\begin{aligned} I(x_i) &= -log_2 P(x_i) \\ &= -log_2(0.5) = 1bit \end{aligned}}$$

- Indeed. we have to use only one bit to represent the output from this binary source
- We use a 1 to represent H and a 0 to represent T.

Example(cont.)

- Suppose the successive outputs from this binary source are statistically independent. i.e.. the source is memoryless.
- Consider a block of m binary digits.
- There are 2^m possible m-bit blocks. each of which is equally probable with probability 2^{-m} .
- The self information of an m-bit block is

$$\boxed{\begin{aligned} I(x_i) &= -log_2 P(x_i) \\ &= -log_2 2^{-m} = mbit \end{aligned}}$$

- Again, we observe that we indeed need m bits to represent the possible m-bit blocks

Example

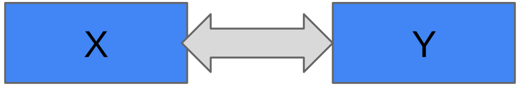
- Consider a discrete memoryless source (source C) that generates two bits at a time.
- This source comprises of two binary sources (sources A and B) and each source contributing one bit.
- The two binary sources within the source C are independent.
- Intuitively, the information content of the aggregate source (source C) should be the sum of the information contained in the outputs of the two independent sources that constitute this source C.

$$\boxed{P(C) = P(A)P(B) = (0.5)(0.5) = 0.25}$$

$$\boxed{I(C) = -log_2 P(x_i) = -log_2 (0.25) = 2bits}$$

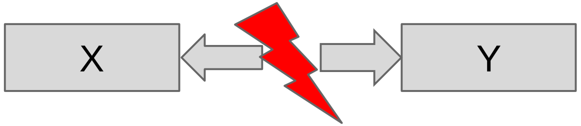
Mutual Information

- Consider two discrete random variables X and Y with possible outcomes $x_i, i = 1, 2, 3...n$ and $y_j, j = 1, 2, 3...m$ in respectively.
- Suppose we observe some outcome $\boxed{Y = y_j}$ and we want to determine the amount of information this event provides about the event $\boxed{X = x_i, i = 1, 2, 3..n}$
- That is. we want to mathematically represent the mutual information.

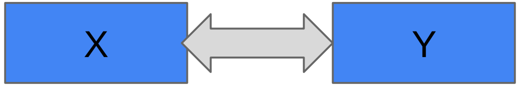


Mutual Information

- Note the following:
- If X and Y are independent, in which case the occurrence of $\boxed{Y = y_j}$ provides no information about $\boxed{X = x_i}$



- If X and Y are fully dependent events, in which case the occurrence of $\boxed{Y = y_j}$ determines the occurrence of the event of $\boxed{X = x_i}$



Mutual Information(Definition)

➡ The mutual information $I(x_i;y_j)$ between x_i and y_j is defined as

$$I(x_i;y_j) = log(\frac{P(x_i|y_j)}{y_j})$$

➡ As before. the units of $I(x)$ are determined by the base of the logarithm. which is usually selected as 2 or e

➡ When the base is 2 the units are in bits.

Mutual Information(Definition)

- The mutual information

$$I(x_i ; y_j) = log(\frac{P(x_i|y_j)}{x_i})$$

- Observe that

$$\frac{P(x_i|y_j)}{P(x_i)} = \frac{P(x_i|y_j)P(y_j)}{P(x_i)P(y_j)} = \frac{P(x_i,y_j)}{P(x_i)P(y_j)} = \frac{P(x_i,y_j)P(y_j|x_i)}{P(x_i,y_j)P(y_j)} = \frac{P(y_j|x_i)}{P(y_j)}$$

- Therefore,

$$I(x_i;y_j) = log(\frac{P(x_i|y_j)}{P(x_i)}) = log(\frac{P(y_j|x_i)}{P(y_j)}) = I(y_j;x_i)$$

Physical Interpretation

- When the random variables X and Y are statistically independent $\boxed{P(x_i|y_j) = P(x_i)}$ which leads to $\boxed{I(x_i; y_j) = 0}$
- When the occurrence of $Y = y_j$ uniquely determines the occurrence of the event $X = x_i$, $\boxed{P(x_i|y_j) = 1}$ the mutual information becomes ,

$$\boxed{I(x_i; y_j) = log\frac{1}{P(x_i)} = -logP(x_i)}$$

- This is the self information of the event $X = x_i$

Mutual Information (Example)

$$P(Y=0) = P(X=0)P(Y=0|X=0)+ \\ P(X=1)P(Y=0|X=1)$$

$$P(Y=1) = P(X=0)P(Y=1|X=0)+ \\ P(X=1)P(Y=1|X=1)$$

$$I(x_0;y_0) = I(0,0) = log_2(\frac{P(Y=0|X=0)}{P(Y=0)}) \\ = log_2(\frac{1-p}{0.5}) = log_2 2(1-P)$$

$$I(x_1;y_0) = I(1,0) = log_2(\frac{P(Y=0|X=1)}{P(Y=0)}) \\ = log_2(\frac{p}{0.5}) = log_2 p$$

Mutual Information (Example)

➡ Suppose $p = 0$, i.e, it is an ideal channel(noiseless), In that case

$$I(x_0; y_0) = I(0; 0) = log_2 2(1 - p) = 1bit$$

➡ Hence having observed with certainty the output can determine what was transmitted. Recall that the self information about the event was 1 bit.

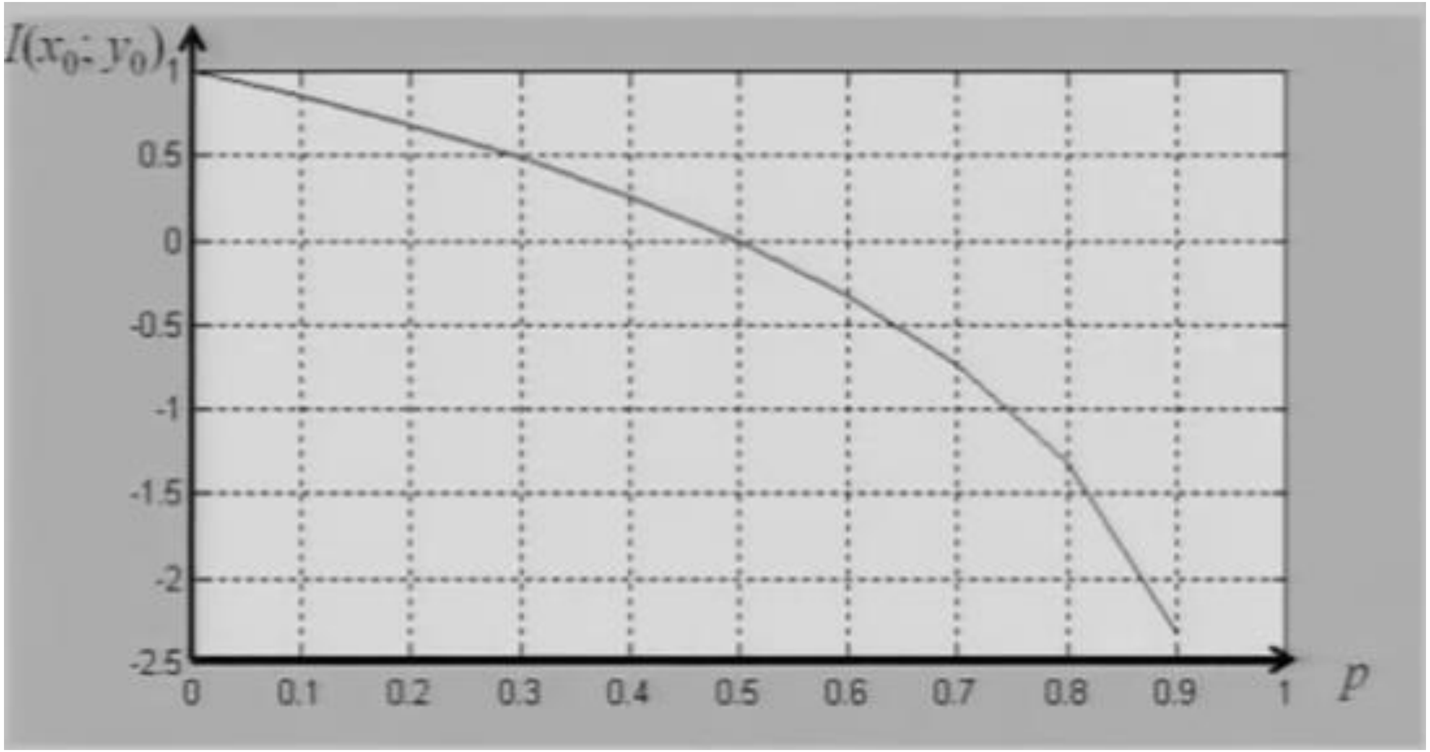
Mutual Information (Example)

- However if $p = 0.5$, we obtain

$$I(x_0; y_0) = I(0;0) = log_2 2(1-p)$$
$$= log_2 2(0.5) = 0$$

- This implies that having observed the output. we have no information about what was transmitted. Thus. it is a useless channel.
- For such a channel, there is no point in observing the received symbol and trying to make a guess as to what was sent.
- Instead we can as well toss a fair coin at the receiver in order to estimate what was sent!

Mutual Information (Significance)



Mutual Information (Significance)

- At $p = 0.5$, $I(x_i ; y_j) = 0$ but when $p > 0.5$ mutual information becomes negative
- $I(x_0 ; y_0) < 0$ when $p > 0.5$

- If $I(x_0 ; y_0)$ large then it supposed to say that by seeing at $Y = 0$ it can be strongly tell that the transmitted signal most probably was $X = 0$
- If $I(x_0 ; y_0) = 0$ then, nothing can be said
- If $I(x_0 ; y_0) < 0$ and if we get $Y = 0$ at the receiver end then it could be said that the transmitted bit was $X = 1$ and channel made a mistake. This is the significance of negative MI

Binary Channel

$$
\begin{aligned}
P(Y=0) = P(X=0)P(Y=0|X=0)+ \\
P(X=1)P(Y=0|X=1) \\
= 0.5(1-p_0) + 0.5(p_1) \\
= 0.5(1-p_0+p_1)
\end{aligned}
$$

$$
\begin{aligned}
P(Y=1) = P(X=0)P(Y=1|X=0)+ \\
P(X=1)P(Y=1|X=1) \\
= 0.5(p_0) + 0.5(1-p_1) \\
= 0.5(1-p_1+p_0)
\end{aligned}
$$

Binary Channel

- The mutual information about the occurrence of the event $X = 0$ given that $Y = 0$ is

$$I(x_0; y_0) = I(0;0) = log_2(\frac{P(Y{=}0|X{=}0)}{P(Y{=}0)})$$

$$= log_2(\frac{1-p_0}{1-p_0+p_1})$$

$$I(x_1; y_0) = I(1;0) = log_2(\frac{P(Y{=}0|X{=}1)}{P(Y{=}0)})$$

$$= log_2(\frac{2p_1}{1-p_0+p_1})$$

CUET

Mutual Information (Significance)

- Mutual information tells the goodness of the channel. It is a measure of the goodness of a channel.
- Channel is supposed to communicate information. How much ? That is exactly what mutual information tells us.

Average Mutual Information(Definition)

- The average mutual information between two random variables X and Y is given by:

$$I(X;Y) = \sum_{i=1}^{n} \sum_{j=1}^{m} P(x_i, y_j) I(x_i; y_j)$$

$$= \sum_{i=1}^{n} \sum_{j=1}^{m} P(x_i, y_j) log\left(\frac{P(x_i, y_j)}{P(x_i) P(y_j)}\right)$$

- Average mutual information tells the overall capability of the channel.
- $I(X;Y) \geq 0$, With equality if and only if X & Y are statistically independent
- The amount of information being conveyed from X to Y is the same as amount of being communicated from Y to X.

Thank You