# Entropy and mutual information

## Communication Theory

Manuel A. Vázquez

April 8, 2024

# Joint entropy

If $X$, $Y$ are **independent**

$$
\begin{aligned}
H(X,Y) &= \sum_{i=1}^{M}\sum_{j=1}^{L} p(x_i,y_j)\log_2\frac{1}{p(x_i,y_j)} = \sum_{i=1}^{M}\sum_{j=1}^{L} p(x_i)p(y_j)\log_2\frac{1}{p(x_i)p(y_j)} \\
&= \sum_{i=1}^{M}\sum_{j=1}^{L} p(x_i)p(y_j)\left(\log_2\frac{1}{p(x_i)} + \log_2\frac{1}{p(y_j)}\right) \\
&= \sum_{i=1}^{M}\sum_{j=1}^{L} \left(p(x_i)p(y_j)\log_2\frac{1}{p(x_i)} + p(x_i)p(y_j)\log_2\frac{1}{p(y_j)}\right) \\
&= \sum_{i=1}^{M}\sum_{j=1}^{L} p(x_i)p(y_j)\log_2\frac{1}{p(x_i)} + \sum_{i=1}^{M}\sum_{j=1}^{L} p(x_i)p(y_j)\log_2\frac{1}{p(y_j)} \\
&= \sum_{i=1}^{M} p(x_i)\log_2\frac{1}{p(x_i)}\cancelto{1}{\sum_{j=1}^{L} p(y_j)} + \cancelto{1}{\sum_{i=1}^{M} p(x_i)}\sum_{j=1}^{L} p(y_j)\log_2\frac{1}{p(y_j)} \\
&= \underbrace{\sum_{i=1}^{M} p(x_i)\log_2\frac{1}{p(x_i)}}_{H(X)} + \underbrace{\sum_{j=1}^{L} p(y_j)\log_2\frac{1}{p(y_j)}}_{H(Y)} = H(X) + H(Y)
\end{aligned}
$$

# Joint and conditional entropy

$$H(X,Y) = \sum_{i=1}^{M}\sum_{j=1}^{L} p(x_i,y_j)\log_2\frac{1}{p(x_i,y_j)} = \sum_{i=1}^{M}\sum_{j=1}^{L} p(x_i,y_j)\log_2\frac{1}{p(y_j|x_i)p(x_i)}$$

$$= \sum_{i=1}^{M}\sum_{j=1}^{L} p(x_i,y_j)\left(\log_2\frac{1}{p(y_j|x_i)} + \log_2\frac{1}{p(x_i)}\right)$$

$$= \underbrace{\sum_{i=1}^{M}\sum_{j=1}^{L} p(x_i,y_j)\log_2\frac{1}{p(y_j|x_i)}}_{H(Y|X)} + \sum_{i=1}^{M}\sum_{j=1}^{L} p(x_i,y_j)\log_2\frac{1}{p(x_i)}$$

$$= H(Y|X) + \sum_{i=1}^{M}\log_2\frac{1}{p(x_i)}\underbrace{\sum_{j=1}^{L} p(x_i,y_j)}_{p(x_i)} = H(Y|X) + \underbrace{\sum_{i=1}^{M}\log_2\frac{1}{p(x_i)}p(x_i)}_{H(X)}$$

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

Likewise one can show that

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

# Conditional entropies

Since we have both

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

and

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

the two right-hand sides above should be equal and we have

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

and

$$H(X|Y) = H(Y|X) + H(X) - H(Y).$$

These are not so common/useful.

# Mutual information and conditional entropy

$$I(X,Y) = \sum_{i=1}^{M}\sum_{j=1}^{L} p(x_i,y_j)\log_2\frac{p(x_i,y_j)}{p(x_i)p(y_j)} = \sum_{i=1}^{M}\sum_{j=1}^{L} p(x_i,y_j)\log_2\frac{p(x_i|y_j)\cancel{p(y_j)}}{p(x_i)\cancel{p(y_j)}}$$

$$= \sum_{i=1}^{M}\sum_{j=1}^{L} p(x_i,y_j)\left(\log_2\frac{1}{p(x_i)} + \log_2 p(x_i|y_j)\right)$$

$$= \sum_{i=1}^{M}\sum_{j=1}^{L} p(x_i,y_j)\log_2\frac{1}{p(x_i)} + \sum_{i=1}^{M}\sum_{j=1}^{L} p(x_i,y_j)\log_2 p(x_i|y_j)$$

$$\overset{\log_2 x = -\log_2 \frac{1}{x}}{=} \sum_{i=1}^{M}\log_2\frac{1}{p(x_i)}\underbrace{\sum_{j=1}^{L} p(x_i,y_j)}_{p(x_i)} - \sum_{i=1}^{M}\sum_{j=1}^{L} p(x_i,y_j)\log_2\frac{1}{p(x_i|y_j)}$$

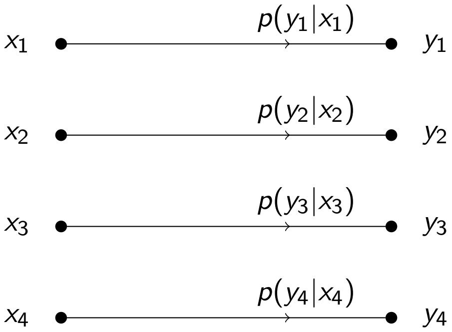
$$= \underbrace{\sum_{i=1}^{M} p(x_i)\log_2\frac{1}{p(x_i)}}_{H(X)} - \underbrace{\sum_{i=1}^{M}\sum_{j=1}^{L} p(x_i,y_j)\log_2\frac{1}{p(x_i|y_j)}}_{H(X|Y)} = H(X) - H(X|Y)$$

Likewise, it can be shown that

$$I(X,Y) = H(Y) - H(Y|X).$$

# Mutual information: example

Let us consider the DMC




$p(x_1)$, $p(x_2)$, $p(x_3)$, $p(x_4)$ are known
$X \in \{x_1, x_2, x_3, x_4\}$ is transmitted, $Y \in \{y_1, y_2, y_3, y_4\}$ is received and, in this case, there is one-to-one mapping between $X$ and $Y$.

Here the conditional probabilities are

$$p(y_i|x_i) = 1 \ \forall i,$$
$$p(y_j|x_i) = 0 \ \forall i \neq j.$$

The entropy of $X$ is given by

$$H(X) = \sum_{i=1}^{4} p(x_i) log_2 \frac{1}{p(x_i)}.$$

Since $p(y_i) = p(x_i)$ we have

$$H(Y) = \sum_{i=1}^{4} p(y_i) log_2 \frac{1}{p(y_i)} = \sum_{i=1}^{4} p(x_i) log_2 \frac{1}{p(x_i)} = H(X)$$

$$I(X,Y) = \sum_{i=1}^{4}\sum_{j=1}^{4} p(x_i,y_j)\log_2 \frac{p(x_i,y_j)}{p(x_i)p(y_j)}$$

$$= \sum_{i=1}^{4}\sum_{j=1}^{4} p(y_j|x_i)p(x_i)\log_2 \frac{p(y_j|x_i)\cancel{p(x_i)}}{\cancel{p(x_i)}p(y_j)}$$

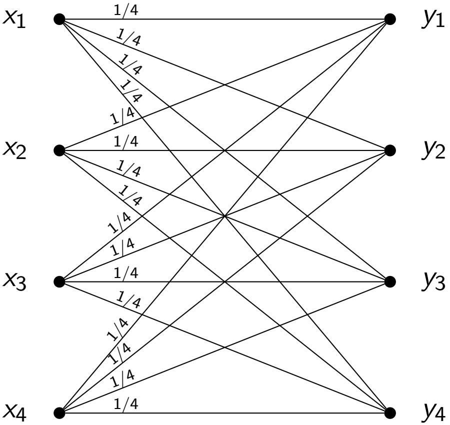
$$\overset{\substack{p(y_j|x_i)=0\\ \forall j\neq i}}{=} \sum_{i=1}^{4} p(y_i|x_i)p(x_i)\log_2 \frac{p(y_i|x_i)}{p(y_i)}$$

$$\overset{p(y_i|x_i)=1}{=} \sum_{i=1}^{4} p(x_i)\log_2 \frac{1}{p(y_i)}$$

$$\overset{p(y_i)=p(x_i)}{=} \sum_{i=1}^{4} p(x_i)\log_2 \frac{1}{p(x_i)} = H(X) = H(Y)$$

Information at the output of the channel, $I(X,Y)$, is equal to that at the input $H(X)$

What happens if...




?

Notice

$$p(y_j|x_i) = \frac{1}{4}\ \forall i, j.$$

$$\begin{aligned} I(X,Y) &= \sum_{i=1}^{4}\sum_{j=1}^{4} p(x_i,y_j)\log_2 \frac{p(x_i,y_j)}{p(x_i)p(y_j)} \\ &= \sum_{i=1}^{4}\sum_{j=1}^{4} p(y_j|x_i)p(x_i)\log_2 \frac{p(y_j|x_i)\cancel{p(x_i)}}{\cancel{p(x_i)}p(y_j)} \\ &= \sum_{i=1}^{4}\sum_{j=1}^{4} \frac{1}{4}p(x_i)\log_2 \frac{1/4}{p(y_j)}, \end{aligned} \tag{1}$$

and we need

$$p(y_j) = \sum_{i=1}^{4} p(y_j,x_i) = \sum_{i=1}^{4} p(y_j|x_i)p(x_i) = \frac{1}{4}\sum_{i=1}^{4} p(x_i) = \frac{1}{4},$$

which does not depend on the input!!

Going back to (1),

$$I(X,Y) = \sum_{i=1}^{4}\sum_{j=1}^{4}\frac{1}{4}p(x_i)\log_2\frac{1/4}{p(y_j)} = \sum_{i=1}^{4}\sum_{j=1}^{4}\frac{1}{4}p(x_i)\log_2\frac{\cancel{1/4}}{\cancel{1/4}} = 0,$$

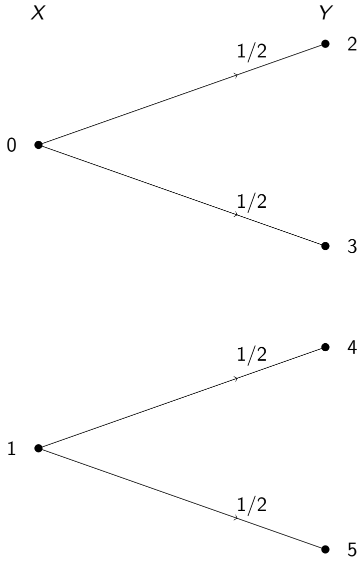
i.e. **no** information goes through the channel: $Y$ does not provide any information about $X$ because whatever $X$ is transmitted every $Y$ is equally likely.

Safety check:

$$p(x_i|y_j) = \frac{p(y_j|x_i)p(x_i)}{p(y_j)} = \frac{1/4p(x_i)}{1/4} = p(x_i) \text{ (Bayes Theorem).}$$

That is, $x_i$ and $y_j$ are independent

# Capacity: example




$$
\begin{aligned}
p(X=0) &= \alpha \\
p(X=1) &= 1-\alpha
\end{aligned}
$$

Mathematically,

$$C = \max I(X, Y),$$

and, following the usual approach (here, the hard way!!),

$$I(X, Y) = H(Y) - H(Y|X).$$

For computing $H(Y)$, we need

$$p(Y=2) = p(2|0)p(X=0) + \cancelto{0}{p(2|1)p(X=1)} = \frac{1}{2}\alpha$$

$$p(Y=3) = p(3|0)p(X=0) + \cancelto{0}{p(3|1)p(X=1)} = \frac{1}{2}\alpha$$

$$p(Y=4) = \cancelto{0}{p(4|0)p(X=0)} + p(4|1)p(X=1) = \frac{1}{2}(1-\alpha)$$

$$p(Y=5) = \cancelto{0}{p(5|0)p(X=0)} + p(5|1)p(X=1) = \frac{1}{2}(1-\alpha).$$

$$H(Y) = p(Y=2)\log_2\frac{1}{p(Y=2)} + p(3)\log_2\frac{1}{p(3)} + p(4)\log_2\frac{1}{p(4)} + p(5)\log_2\frac{1}{p(5)}$$

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

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

$$= \alpha\underbrace{\log_2 2}_{1} + \alpha\log_2\frac{1}{\alpha} + (1-\alpha)\underbrace{\log_2 2}_{1} + (1-\alpha)\log_2\frac{1}{1-\alpha}$$

$$= \cancel{\alpha} + \alpha\log_2\frac{1}{\alpha} + 1 - \cancel{\alpha} + (1-\alpha)\log_2\frac{1}{1-\alpha} = 1 + H_b(\alpha).$$

$$H(Y|X) = \sum_{i=1}^{M} p(x_i)H(Y|X=x_i) = p(X=0)H(Y|X=0) + p(X=1)H(Y|X=1)$$

$$= \alpha\left(p(2|0)\log_2\frac{1}{p(2|0)} + p(3|0)\log_2\frac{1}{p(3|0)} + \underbrace{p(4|0)\log_2\frac{1}{p(4|0)}}_{0} + \right.$$

$$\left. + \; + \underbrace{p(5|0)\log_2\frac{1}{p(5|0)}}_{0}\right) + (1-\alpha)\left(\underbrace{p(2|1)\log_2\frac{1}{p(2|1)}}_{0} + \right.$$

$$\left. + \; \underbrace{p(3|1)\log_2\frac{1}{p(3|1)}}_{0} + p(4|1)\log_2\frac{1}{p(4|1)} + p(5|1)\log_2\frac{1}{p(5|1)}\right)$$

$$= \alpha\underbrace{\left(\frac{1}{2}\log_2 2 + \frac{1}{2}\log_2 2\right)}_{1} + (1-\alpha)\underbrace{\left(\frac{1}{2}\log_2 2 + \frac{1}{2}\log_2 2\right)}_{1} = 1.$$

## Putting it all together

$$I(X, Y) = H(Y) - H(Y|X) = 1 + H_b(\alpha) - 1 = H_b(\alpha)$$

An easier way to obtain $I(X, Y)$:

$$I(X, Y) = H(X) - H(X|Y) = H(X) = H_b(\alpha),$$

where it has been used that

- $H(X|Y) = 0$ because it is *the uncertainty we have about $X$ once we know $Y$*
- $H(X) = H_b(\alpha)$ because $X$ is a binary r.v.

The **capacity** is then

$$C = \max I(X, Y) = \max H_b(\alpha) = 1 \frac{\text{bit}}{\text{channel use}},$$

and it is attained when $\alpha = 1/2$ (uniform distribution).