ECE 562: Information Theory	**Spring 2006**

Lecture 1 — Jan 26th 2006.

Lecturer: Sergio D. Servetto *Scribe: Animashree Anandkumar*

1.1 Shannon's Information Measures(Yeung,22)

Definition for Entropy :

A random variable X (unless otherwise mentioned discrete) taking values from $\mathcal{X} = \{1, \ldots, n\}$ with PMF $p(X) = \mathbb{P}(X = i) = p_i,\ i = 1, \ldots, n$ has entropy $H(X)$ given by,

$$H(X) = \sum_{i=1}^{n} p_i \log \frac{1}{p_i}. \tag{1.1}$$

Entropy is expressed in bits when base of logarithm is 2.

Let $g : \mathcal{X} \mapsto \Re$ and X is a r.v on $\mathcal{X}$ define a new r.v $Y = g(X)$. We know that,

$$\begin{aligned} \mathbb{E}Y &= \mathbb{E}g(X) \\ &= \sum_{x \in \mathcal{X}} \mathbb{P}(X = x) g(x). \end{aligned}$$

We can express entropy as an expectation,

$$H(X) = \mathbb{E} \log \frac{1}{p(X)}, \tag{1.2}$$

where we have $g(x) = \frac{1}{p(x)}$.

Intuitively, Entropy is the average amount of uncertainty of a random variable. More the uncertainty, more is the entropy.

Entropy is a function of the distribution and not of the alphabet itself.

Example : Binary Entropy $h_b(p)$

Let $0 \leq p \leq 1$. Then,

$$h_b(p) = \left\{ \begin{array}{cc} -p \log p - (1-p) \log(1-p) & \text{for} 0 < p < 1. \\ 0 & \text{o.w} \end{array} \right. \tag{1.3}$$

When $p = 0$ or $p = 1$, we use calculus to evaluate the limits. Intuitively, for these values, there is no uncertainty in the random variable, *i.e.,* it is deterministic.

Maximum of $h_b(p)$ occurs when $p = \frac{1}{2}$ with $h_b(\frac{1}{2}) = 1$. Intuitively, this is because uncertainty is maximum when the outcomes are equally likely.
$h_b(\cdot)$ is a concave function of p (as log is a concave function).

Multiple Random Variables

Joint Entropy : Let X, Y be r.vs taking values in $\mathcal{X}, \mathcal{Y}$. The joint entropy $H(X, Y)$ is given by,

$$\begin{aligned} H(X,Y) &= -\mathbb{E} \log p(X,Y). && (1.4) \\ &= -\sum_{(x,y)\in\mathcal{X}\times\mathcal{Y}} p(x,y)\log p(x,y). && (1.5) \end{aligned}$$

Joint entropy denotes the uncertainty for two random variables. Note that it is symmetric wrt to the random variables.

Conditional Entropy :

$$\begin{aligned} H(X|Y) &= -\mathbb{E} \log p(X|Y). && (1.6) \\ &= -\sum_{(x,y)\in\mathcal{X}\times\mathcal{Y}} p(x,y)\log p(x|y). && (1.7) \end{aligned}$$

Intuitively, the conditional entropy is the uncertainty of a random variable given another random variable.

Note that by Bayes rule,

$$p(x,y) = p(y)p(x|y). \qquad (1.8)$$

Thus we have,

$$\begin{aligned} H(X|Y) &= -\sum_{y\in\mathcal{Y}}\sum_{x\in\mathcal{X}} p(y)p(x|y)\log p(x|y) \\ &= -\sum_{y\in\mathcal{Y}} p(y) \sum_{x\in Xc} p(x|y)\log p(x|y) \\ &\triangleq \sum_{y\in\mathcal{Y}} p(y)H(X|Y=y) \\ &\triangleq H(X|Y) \end{aligned}$$

Example :

For X, Y the joint distribution $p(X, Y)$ is given in the table.

X \ Y	a	b
0	0.25	0.25
1	0.5	0

$$\begin{aligned} H(Y|X=0) &= 1 \\ H(Y|X=1) &= 0 \\ \therefore\ H(Y|X) &= \mathbb{P}(X=0)H(Y|X=0) + \mathbb{P}(X=1)H(Y|X=1) \\ &= \frac{1}{2} \end{aligned}$$

Relationship between Joint and Conditional Entropy :

$$\begin{aligned} H(X,Y) &= H(X|Y) + H(Y) &\quad (1.9) \\ &= H(Y|X) + H(X). &\quad (1.10) \end{aligned}$$

Proof :

$$\begin{aligned} H(X,Y) &= -\mathbb{E}\log p(X,Y) \\ &= -\mathbb{E}\log p(X)p(Y|X) \\ &= -\mathbb{E}(\log p(X) + \log p(Y|X)) \\ &= -\mathbb{E}\log p(X) - \mathbb{E}\log(p(Y|X) \text{ Linearity of expectation.} \\ &= H(X) + H(Y|X). \end{aligned}$$

Similarly we can prove the second identity.

Mutual Information :

$$\begin{aligned} I(X;Y) &= \sum_{x,y \in \mathcal{X}\times\mathcal{Y}} p(x,y) \log \frac{p(x,y)}{p(x)p(y)} &\quad (1.11) \\ &= \mathbb{E}\log \frac{p(X,Y)}{p(X)p(Y)} &\quad (1.12) \end{aligned}$$

Intuitively, Mutual information is a measure of how far two random variables are from independence.

1. For $X \perp Y$ (*i.e.,* X and Y are independent) then $I(X;Y) = 0$.

2. If $\mathbb{P}(X = Y) = 1$ then $I(X;Y) = H(X) = H(Y)$. In other words, for any random variable X, $I(X;X) = H(X)$. Thus entropy is also known as *Self Information.*

3. We have,

$$
\begin{aligned}
I(X;Y) &= H(X) - H(X|Y). && (1.13) \\
&= H(Y) - H(Y|X). && (1.14)
\end{aligned}
$$

It is useful to think of the information measures in terms of Venn Diagrams. Then mutual information can be thought of as the intersection, the conditional entropy as difference and the joint entropy as the union of two sets respectively. With this, the above identities can be derived easily.