# EE7790 Project for MEE: Advanced Image Processing and Computer Vision

## CHAPTER 1

**Entropy**

Measures the uncertainty of a process

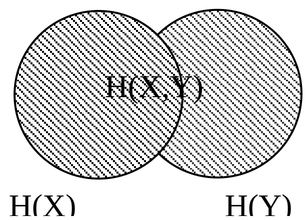
$$H(X) = \sum_{x \in X} p(x) \log_2\left(\frac{1}{p(x)}\right) = -\sum_{x \in X} p(x) \log_2(p(x))$$

For N size alphabet we have

$$H(X) = \sum_{i=1}^{N} p_i \log_2\left(\frac{1}{p_i}\right) = -\sum_{i=1}^{N} p_i \log_2(p_i)$$

**Joint Entropy**

$$H(X,Y) = \sum_{x \in X}\sum_{y \in Y} p(x,y) \log_2\left(\frac{1}{p(x,y)}\right)$$




**Conditional Entropy**

Means of additional uncertainty

$$H(X/Y) = \sum_{x \in X}\sum_{y \in Y} p(x,y) \log_2\left(\frac{1}{p(x/y)}\right) = -\sum_{x \in X}\sum_{y \in Y} p(x,y) \log_2(p(x/y))$$

$$H(X/Y) = -\sum_{x \in X}\sum_{y \in Y} p(x,y) \log_2\left(\frac{p(x,y)}{p(y)}\right) \text{ since } p(x/y) = \frac{p(x,y)}{p(y)}$$

Operating

$$H(X/Y) = -\sum_{x \in X}\sum_{y \in Y} p(x,y)\left[\log_2(p(x,y)) - \log_2 p(y)\right]$$

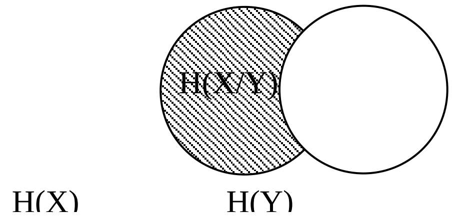
$$H(X/Y) = -\sum_{x \in X}\sum_{y \in Y} p(x,y) \log_2(p(x,y)) - p(x,y) \log_2 p(y)$$

$$H(X/Y) = -\sum_{x \in X}\sum_{y \in Y} p(x,y) \log_2(p(x,y)) - \left[-\sum_{y \in Y}\left(\sum_{x \in X} p(x,y)\right) \log_2 p(y)\right]$$

$$H(X/Y) = H(X,Y) - \left[-\sum_{y \in Y} p(y) \log_2 p(y)\right]$$

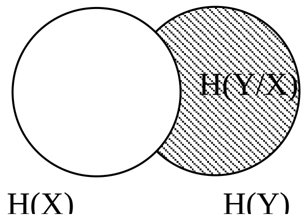
$$H(X/Y) = H(X,Y) - H(Y)$$

Meaning: the uncertainty of X given Y, is the Uncertainty of both minus the uncertainty of Y. (Graphically is the left side moon-shape)



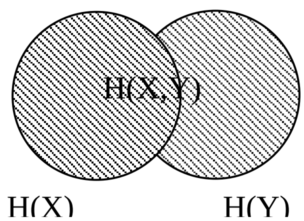

Similarly:

$H(Y/X) = H(X,Y) - H(X)$



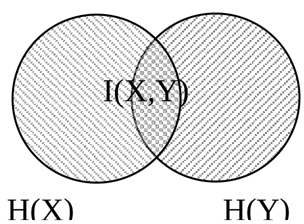

Because $H(X,Y) = H(Y,X)$ we have that:

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




Rearranging

$H(X) - H(X/Y) = +H(Y) - H(Y/X) = I(X,Y)$ is the intersection, or **mutual Information** between both RV's




From last equation we have

$H(X) = I(X,Y) + H(X/Y)$

$H(Y) = I(X,Y) + H(Y/X)$

From the previous one we have

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

$H(X,Y) = H(Y) + H(X) - I(X,Y)$ (The sum of both minus the intersection)

**Information**

The Information is defined as

$I(X,Y) = \sum_{x \in X} \sum_{y \in Y} p(x,y) \log_2 \left( \frac{p(x,y)}{p(x)p(y)} \right)$ or from the formulas:

$I(X,Y) = H(X) - H(X/Y)$

$I(X,Y) = H(Y) - H(Y/X)$

We can also find the formula of Information from

$$I(X,Y)=H(X)-H(X/Y)=\sum_{x\in X}p(x)\log_2(\frac{1}{p(x)})-\left(-\sum_{x\in X}\sum_{y\in Y}p(x,y)\log_2(\frac{p(x,y)}{p(y)})\right)$$

$$=\sum_{x\in X}\sum_{y\in Y}p(x,y)\log_2(\frac{1}{p(x)})+\sum_{x\in X}\sum_{y\in Y}p(x,y)\log_2(\frac{p(x,y)}{p(y)})=\sum_{x\in X}\sum_{y\in Y}p(x,y)\left(\log_2(\frac{1}{p(x)})+\log_2(\frac{p(x,y)}{p(y)})\right)$$

$$=\sum_{x\in X}\sum_{y\in Y}p(x,y)\left(\log_2(\frac{p(x,y)}{p(x)p(y)})\right)$$

**Let's do an exercise:**

Let's be two random variables, each of them can have these values $\{a,b,c,d\}$

| | | ***x=*** | | | | |
|---|---|---|---|---|---|---|
| | | ***a*** | ***b*** | ***c*** | ***d*** | |
| ***y=*** | ***a*** | 1/8 | 1/16 | 1/32 | 1/32 | *p(y=a)*=1/4 |
| | ***b*** | 1/16 | 1/8 | 1/32 | 1/32 | *p(y=b)*=1/4 |
| | ***c*** | 1/16 | 1/16 | 1/16 | 1/16 | *p(y=c)*=1/4 |
| | ***d*** | 1/4 | 0 | 0 | 0 | *p(y=d)*=1/4 |
| | | *p(x=a)*=1/2 | *p(x=b)*= 1/4 | *p(x=c)*= 1/8 | *p(x=d)*=1/8 | |

**Joint Probability**. We know that $p(x=a,y=a)=1/8$ etc.

**Marginal Probability.**

$$p(x=a,\forall y)=\sum_{y=a}^{d}p(x=a,y)=p(a,a)+p(a,b)+p(a,c)+p(a,d)=\frac{1}{8}+\frac{1}{16}+\frac{1}{16}+\frac{1}{4}=\frac{1}{2}\text{ etc...}$$

**Conditional Probability.**

$p(x/y)=\dfrac{p(x,y)}{p(y)}$ probability of x when y has happened.

$$p(x=a/y=c)=\frac{p(x=a,y=c)}{p(y=c)}=\frac{1/16}{1/4}=\frac{4}{16}=\frac{1}{4}$$

**Entropy**

$$H(X)=\sum_{x\in X}p(x)\log_2(\frac{1}{p(x)})=\frac{1}{2}\log_2(\frac{1}{1/2})+\frac{1}{4}\log_2 4+\frac{1}{8}\log_2 8+\frac{1}{8}\log_2 8=1.75$$

$$H(Y)=\sum_{y\in Y}p(y)\log_2(\frac{1}{p(y)})=2$$

**Joint Entropy**

$$H(X,Y)=\sum_{x\in X}\sum_{y\in Y}p(x,y)\log_2(\frac{1}{p(x,y)})=3.375$$

**Conditional Entropy**

$H(X/Y)=-\sum_{x\in X}\sum_{y\in Y}p(x,y)\log_2(p(x/y))=1.375$ or

$$H(X/Y)=H(X,Y)-H(Y)=3.375-2=1.375$$

$H(Y/X)=-\sum_{x\in X}\sum_{y\in Y}p(x,y)\log_2(p(y/x))=1.625$ or

$H(Y/X) = H(X,Y) - H(X) = 3.375 - 1.75 = 1.625$

**Information**

$I(X,Y) = \sum_{x \in X} \sum_{y \in Y} p(x,y) \log_2 (\frac{p(x,y)}{p(x)p(y)}) = 0.375$ or

$I(X,Y) = H(X) - H(X/Y) = 1.75 - 1.375 = 0.375$

$I(X,Y) = H(Y) - H(Y/X) = 2 - 1.625 = 0.375$

Other formulas we can confirm:

$H(X) = I(X,Y) + H(X/Y) = 0.375 + 1.375 = 1.75$

$H(Y) = I(X,Y) + H(Y/X) = 0.375 + 1.625 = 2$

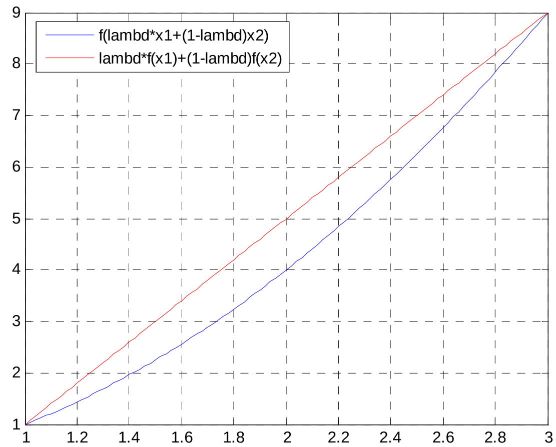
$H(X,Y) = H(Y) + H(X) - I(X,Y) = 2 + 1.75 - 0.375 = 3.375$

**Jensen's Inequality**

CONVEX FUNCTIONS

A function is said to be convex if

$f(\lambda x_1 + (1-\lambda)x_2) \le \lambda f(x_1) + (1-\lambda) f(x_2)$ like next figure




LEMMA:

If a function is convex, this holds:

$f\left(\sum_{i=1}^{2} p_i x_i\right) \le \sum_{i=1}^{2} p_i f(x_i)$ being $p_i$ the probability of $x_i$ and we know that $\sum_{i=1}^{2} p_i = 1$

Jensen's Inequality is this equation extended to N points

$$f\left(\sum_{i=1}^{N} p_i x_i\right) \le \sum_{i=1}^{N} p_i f(x_i)$$

Let's prove this by mathematical induction.

PROOF OF JENSEN'S INEQUALITY:

We know that this is true for N=2, as it is shown in the figure above.

$f\left(\sum_{i=1}^{2} p_i x_i\right) \le \sum_{i=1}^{2} p_i f(x_i)$ Because

$f(p_1 x_1 + (1-p_1)x_1) \le p_1 f(x_i) + (1-p_1) f(x_i)$ Because convergence Lemma.

Lets suppose it is true for N=K-1.

$f\left(\sum_{i=1}^{K-1} p_i x_i\right) \le \sum_{i=1}^{K-1} p_i f(x_i)$ we have to assume that $\sum_{i=1}^{K-1} p_i = 1$

Lets study for N=K

$f\left(\sum_{i=1}^{K} p_i x_i\right) \overset{?}{\le} \sum_{i=1}^{K} p_i f(x_i)$ and we assume that $\sum_{i=1}^{K} p_i = 1$

we can decompose the right side into

$$\sum_{i=1}^{K-1} p_i f(x_i) + p_K f(x_K)$$

To apply the properties of K-1, we have to normalize the probability from 1 to K-1 to be one:

$$(1-p_K)\sum_{i=1}^{K-1} \frac{p_i}{(1-p_K)} f(x_i) + p_K f(x_K)$$

We define now:

$q_i = \frac{p_i}{1-p_k}$ and we know that $\sum_{i=1}^{K-1} q_i = 1$

Therefore:

$$(1-p_K)\sum_{i=1}^{K-1} q_i f(x_i) + p_K f(x_K)$$

Looking at the left side term, and from the formula K-1 rewritten here…

$$f\left(\sum_{i=1}^{K-1} q_i x_i\right) \le \sum_{i=1}^{K-1} q_i f(x_i)$$

…we can expand the right side to the left as:

$$(1-p_K) f\left(\sum_{i=1}^{K-1} q_i x_i\right) + p_K f(x_K) \le (1-p_K)\sum_{i=1}^{K-1} q_i f(x_i) + p_K f(x_K)$$

let's be $\sum_{i=1}^{K-1} q_i x_i = y_0$

$$(1-p_K) f(y_0) + p_K f(x_K) \le (1-p_K)\sum_{i=1}^{K-1} q_i f(x_i) + p_K f(x_K)$$

Remember that for N=2 we have

$$f\left(\sum_{i=1}^{2} p_i x_i\right) \le \sum_{i=1}^{2} p_i f(x_i)$$

$f(p_1 x_1 + p_2 x_2) \le p_1 f(x_1) + p_2 f(x_2)$ and even more

$f((1-p_2)x_1 + p_2 x_2) \le (1-p_2) f(x_1) + p_2 f(x_2)$ because $p_1 + p_2 = 1$

Therefore we can write:

$$f((1-p_K)y_0 + p_K x_K) \le (1-p_K) f(y_0) + p_K f(x_K)$$

Hence:

$$f\left((1-p_K)y_0 + p_K x_K\right) \le (1-p_K)\sum_{i=1}^{K-1} q_i f(x_i) + p_K f(x_K)$$

Knowing that $\sum_{i=1}^{K-1} q_i x_i = y_0$

$$f\left((1-p_K)\sum_{i=1}^{K-1} q_i x_i + p_K x_K\right) \le (1-p_K)\sum_{i=1}^{K-1} q_i f(x_i) + p_K f(x_K)$$

And that $q_i = \frac{p_i}{1-p_k}$

$$f\left((1-p_K)\sum_{i=1}^{K-1} \frac{p_i}{1-p_k} x_i + p_K x_K\right) \le (1-p_K)\sum_{i=1}^{K-1} q_i f(x_i) + p_K f(x_K)$$

Then:

$$f\left(\sum_{i=1}^{K-1} p_i x_i + p_K x_K\right) \le (1-p_K)\sum_{i=1}^{K-1} q_i f(x_i) + p_K f(x_K)$$

Finally the left side is just

$$f\left(\sum_{i=1}^{K} p_i x_i\right) \le (1-p_K)\sum_{i=1}^{K-1} q_i f(x_i) + p_K f(x_K)$$

Of course the right side is (since the beginning)

$$f\left(\sum_{i=1}^{K} p_i x_i\right) \le \sum_{i=1}^{K} p_i f(x_i)$$ **It is proved.**

Jensen's inequality allows us to prove a lot of interesting things about the Entropy. Just note that $-\log_2()$ is a convex function.

Then, because Jensen's iquequality we have that:

$$-\log_2\left(\sum_{i=1}^{N} p_i x_i\right) \le \sum_{i=1}^{N} p_i\left[-\log_2(x_i)\right]$$

Example:
Lets be a RV uniform distributed on N. Therefore we can write in general:

$$H(X) = -\sum_{i=1}^{N} p_i \log_2(p_i) = \sum_{i=1}^{N} p_i\left(-\log_2(p_i)\right)$$

If RV is uniform then $p_i = \frac{1}{N}$

$$H_u(X) = \sum_{i=1}^{N} \frac{1}{N}\left(-\log_2(\frac{1}{N})\right) = N * \frac{1}{N}\left(-\log_2(\frac{1}{N})\right) = \log_2 N$$

We know that uniform distributed RV is the most uncertain. However, can we prove that $H(X) \le \log_2 N$ for all RV?

Let's use Jensen inequality to prove this:
First of all let me write:

$$\log_2 N = (\log_2 N)*1 = \log_2 N * \sum_{i=1}^{N} p_i = \sum_{i=1}^{N} (p_i \log_2 N) = -\sum_{i=1}^{N}\left( p_i \log_2 \frac{1}{N}\right)$$

Therefore:

$$\log_2 N = -\sum_{i=1}^{N}\left( p_i \log_2 \frac{1}{N}\right)$$

I need to prove that $H(X) \overset{?}{\leq} \log_2 N$ for all RV, or what is the same:

$$\log_2 N - H(X) \overset{?}{\geq} 0$$

Let's write

$$\log_2 N - H(X) = -\sum_{i=1}^{N}\left( p_i \log_2 \frac{1}{N}\right) - \sum_{i=1}^{N}\left( p_i \log_2 \frac{1}{p_i}\right)$$

$$= -\sum_{i=1}^{N}\left( p_i \left[\log_2 \frac{1}{N} + \log_2 \frac{1}{p_i}\right]\right) = -\sum_{i=1}^{N}\left( p_i \left[\log_2 \frac{1}{Np_i}\right]\right)$$

or

$$\log_2 N - H(X) = \sum_{i=1}^{N}\left( p_i \left[-\log_2 \frac{1}{Np_i}\right]\right)$$

Because Jensen's inequality we can expand the right side to be:

$$\log_2 N - H(X) = \sum_{i=1}^{N}\left( p_i \left[-\log_2 \frac{1}{Np_i}\right]\right) \geq -\log_2\left(\sum_{i=1}^{N}\left( p_i \frac{1}{Np_i}\right)\right)$$

$$= -\log_2\left(\sum_{i=1}^{N}\left(\frac{1}{N}\right)\right) = -\log_2\left(N\frac{1}{N}\right) = 0$$

Therefore:

$$\log_2 N - H(X) \geq 0$$

Other play (just my ideas here):
Because Jensen's inequality we can expand to the left as:

$$-\log_2\left(\sum_{i=1}^{N} p_i p_i\right) \leq \sum_{i=1}^{N} p_i(-\log_2(p_i)) = H(X)$$

$$-\log_2\left(\sum_{i=1}^{N} {p_i}^2\right) \leq \sum_{i=1}^{N} p_i(-\log_2(p_i)) = H(X)$$

Therefore, including last demonstration upper limit of the Entropy:

$$-\log_2\left(\sum_{i=1}^{N} {p_i}^2\right) \leq H(X) \leq \log_2 N$$

Therefore if $p_i = \frac{1}{N}$ (uniformly distributed)

$$-\log_2\left(\sum_{i=1}^{N} \frac{1}{N^2}\right) \leq H(X) \leq \log_2 N$$

$$-\log_2\left(N\frac{1}{N^2}\right) \leq H(X) \leq \log_2 N$$

$\log_2(N) \le H(X) \le \log_2 N$

**DATA COMPRESSION**

We have two types of data compression: lossless and lossy.
Lossy data compression has two steps: Quantization and lossless compression.

ENCODER
Given a symbol, we define a code for each symbol.

- Non-singular code. If $x_i \neq x_j \Rightarrow C(x_i) \neq C(x_j)$
- Uniquely decodable. If $C(x_1 x_2 \ldots x_n) = C(x_1)C(x_2)\ldots C(x_n)$ or $C(x_1)C(x_2)\ldots C(x_n) \neq C(y_1)C(y_2)\ldots C(y_n)$
- Prefix code: No codeword is a prefix of another codeword. This is also called instantaneous coding.

Definition: length of codeword. $l_i = length(C(x_i))$
Definition: probability of a symbol $p_i = prob(x_i)$

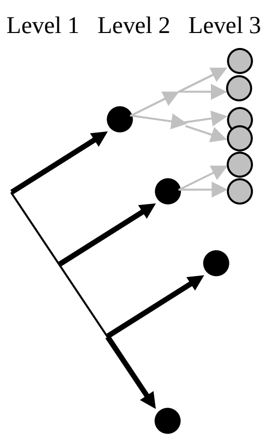
Definition: Avg. bits per symbol $L = \sum p_i l_i = R$ (Average Bit Rate, bits per symbol)

KRAFT INEQUALITY
For any instantaneous code over an alphabet of size D and lengths $l_1, l_2 \ldots l_m$ next inequality holds:

$$\sum_{i=1}^{N} D^{l_i} \le 1$$

PROOF:
Let be $l_{\max} = \max(l_1, l_2 \ldots l_N)$. Consider all branch at level $l_{\max}$. At other levels, each branch has $D^{l_{\max} - l_i}$ possible descendants leaves. (descendant defined as leaves to the right of the actual branch level.)

For example, in a binary code (*D=2*), if $l_{\max} = 3$; at level 2 each branch has $2^{3-2} = 2$ descendants leaves. at level 1 each branch has $2^{3-1} = 4$ descendants leaves. At level 3 each branch has one descendant leaf.




Each codeword branch (next to a leaf, in bold at the figure) have only a descendant leaf, because is a prefix code; hence, each codeword is not a prefix of any other.

The sum of all codeword branches (black nodes) is less than the sum of all branches at level 3 (all nodes) which is $2^{lmax}$. Therefore:

$black\ node_{level1}(0) + black\ node_{level2}(1{,}0) + black\ node_{level3}(1{,}1{,}0) + black\ node_{level3}(1{,}1{,}1) \le \sum_{level3} nodes$ Each branch has descendant leaves depending on its level, therefore

$$2^{l_{\max}-1} + 2^{l_{\max}-2} + 2^{l_{\max}-3} + 2^{l_{\max}-3} \le 2^{l_{\max}}$$
$$\sum_{i=1}^{N} D^{l_{\max}-l_i} \le D^{l_{\max}}$$
$$\sum_{i=1}^{N} D^{l_{\max}} D^{-l_i} \le D^{l_{\max}}$$

$$D^{l_{\max}} \sum_{i=1}^{N} D^{-l_i} \le D^{l_{\max}}$$

$\sum_{i=1}^{N} D^{-l_i} \le 1$ the proof.

MINIMIZATION OF BIT RATE, OR EXPECTED LENGTH.

For binary codes we have:

$L = \sum_{i=1}^{N} p_i l_i = R$ and we know this holds $\sum_{i=1}^{N} 2^{-l_i} \le 1$

We want to minimize L. We will use the Lagrange multiplier method.

In that method we need to minimize $f(x)$ with a constrain $g(x) = 0$

$$J(l_1,...l_N) = \sum_{i=1}^{N} p_i l_i + \lambda\left[\sum_{i=1}^{N} 2^{-l_i} - 1\right]$$

The minimum of this function is also the minimum of $L = \sum_{i=1}^{N} p_i l_i = R$. Therefore lets derivate to find the minimum.

$$\frac{\partial J(l_1,...l_N)}{\partial l_i} = \frac{\partial\left(\sum_{i=1}^{N} p_i l_i + \lambda\left[\sum_{i=1}^{N} 2^{-l_i} - 1\right]\right)}{\partial l_i}$$
$$= \frac{\partial\left(\sum_{i=1}^{N} p_i l_i\right)}{\partial l_i} + \lambda \frac{\partial\left(\left[\sum_{i=1}^{N} 2^{-l_i} - 1\right]\right)}{\partial l_i} = p_i - \lambda \frac{\partial\left(2^{-l_i}\right)}{\partial l_i} = p_i + \lambda(-1)2^{-l_i} \ln 2$$

We equate this to zero and solve for $l_i$

$$0 = p_i + \lambda(-1)2^{-l_i} \ln 2$$
$$2^{-l_i} = \frac{p_i}{\lambda \ln 2}$$

Plugging this value at the constrain equation we have

$$\sum_{i=1}^{N} 2^{-l_i} - 1 = 0 \Leftrightarrow \sum_{i=1}^{N} \frac{p_i}{\lambda \ln 2} - 1 = 0$$

Solving for $\lambda$

$\frac{1}{\lambda \ln 2} \sum_{i=1}^{N} p_i = 1$ and because $\sum_{i=1}^{N} p_i = 1$ we have

$$\lambda = \frac{1}{\ln 2}$$

Pluggin this value of $\lambda$ in the derivate equation to solve for $l_i$

$$0 = p_i + \lambda(-1)2^{-l_i} \ln 2$$

$$0 = p_i + \frac{1}{\ln 2}(-1)2^{-l_i} \ln 2$$

$$2^{-l_i} = p_i$$

$-l_i = \log_2 p_i$ or $l_i = \log_2 \frac{1}{p_i}$

Finally we plug this value to find the minimum of L

$$L_{\min} = \sum_{i=1}^{N} p_i l_i \Big|_{l_i = \log_2 \frac{1}{p_i}} = R_{\min}$$

$$\boxed{L_{\min} = R_{\min} = \sum_{i=1}^{N} p_i \log_2 \frac{1}{p_i} = H(X)}$$

The minimum average length of bit rate is the Entropy itself.

EXAM POSSIBLE QUESTIONS

1. Find all possible relations between Information and Entropy.
2. Prove Jensen's inequality
3. Prove $H(X) \le \log_2 N$ for all Random Variables
4. Find the theoretical minimum for Entropy.
5. Prove Kraft Inequality
6. Find the minimum average bit rate for instantaneous coding.
7. Find the minimum average bit rate for instantaneous coding using Jensen's inequality.
8. What is the implementation of the Shannon codeing?
9. What is the implementation of Huffman coding?
10. Find the bounds of the average bit rate for Vector coding

**New Questions**

11. Prove that the Shannon Coding agree with Kraft Inequality.

**EOD**