E2 205 Error-Control Coding
Lecture 12

Naveeta Maheswari

september 23, 2019

1 Entropy

Let X be a discrete random variable having finite alphabets $\mathcal{X}$ and $P_X(x)$= p(x) be the pmf of X.

Then, the entropy H(X) of X in bits per symbol is given by

$$H(X) = \sum_{x \in \mathcal{X}} p(x) \log \frac{1}{p(x)} = H_2(X)$$

$$H_b(X) = \sum_{x \in \mathcal{X}} p(x) \log_b \frac{1}{p(x)} = \log_b(2)\ H_2(X)$$

Example 1 *Let $\mathcal{X} = \{0, 1\}$ and $P_X(x) = p(x)$.*

Then,

$$\begin{aligned} H(X) &\triangleq H_2(p) = H_2(p, (1-p)) \\ &= p \log \frac{1}{p} + (1-p) \log \frac{1}{(1-p)} \end{aligned}$$

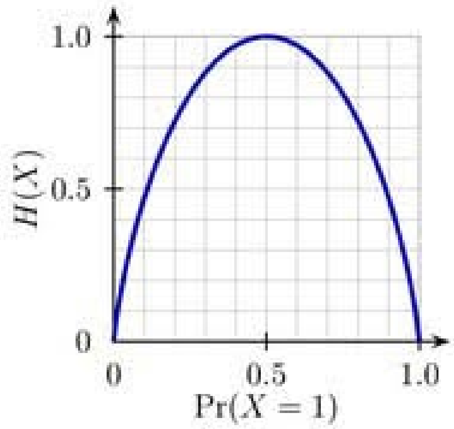


Figure 1: Binary entropy function

Thus, H(X) is maximum when {0,1} are equally likely. This is in general true.

Example 2 *Let* $\mathcal{X} = \{0, 1, 2..., M-1\}$.
X has pmf p(x) and Y be uniform over $\mathcal{X}$

Hence,

$$p(y) = \frac{1}{M} \ , \quad all \ \ y \in \mathcal{X}$$

The entropy of Y is given by

$$H(Y) = \sum_{y \in \mathcal{X}} p(y) \log \frac{1}{p(y)}$$

$$= \sum_{y=0}^{M-1} \frac{1}{M} \log(M)$$

$$H(Y) = \log M$$

Claim: $\quad H(X) \leq H(Y), \quad$ *with equality iff* $p(x) = \frac{1}{M}$, $\forall \ x \in \mathcal{X}$

Proof:

$$
\begin{aligned}
H(X) - H(Y) &= \sum_{x \in \mathcal{X}} p(x) \log \frac{1}{p(x)} - \sum_{y \in \mathcal{X}} p(y) \log \frac{1}{p(y)} \\
&= \sum_{x \in \mathcal{X}} p(x) \log \frac{1}{p(x)} - \sum_{y \in \mathcal{X}} \frac{1}{M} \log(M) \\
&= \sum_{x \in \mathcal{X}} p(x) \log \frac{1}{p(x)} - \sum_{x \in \mathcal{X}} p(x) \log(M) \\
&= \sum_{x \in \mathcal{X}} p(x) \log \frac{1}{Mp(x)} \\
&\leq \sum_{x \in \mathcal{X}} p(x) \left[\frac{1}{Mp(x)} - 1 \right] \quad (\textit{ see aside }) \\
&= \sum_{x \in \mathcal{X}} \frac{1}{M} - \sum_{x \in \mathcal{X}} p(x) \\
&= 1 - 1 = 0
\end{aligned}
$$

Hence, $H(X) \leq H(Y)$

Equality holds iff $\frac{1}{Mp(x)} = 1 \ , \ \forall \ x \in \mathcal{X}$.

$$\Rightarrow p(x) = \frac{1}{M} \ , \ \forall \ x \in \mathcal{X}$$

Aside: From the linear approximation of ln(x) it is known that

$$\ln(x) \leq (x - 1)$$

with equality if and only if x=1.

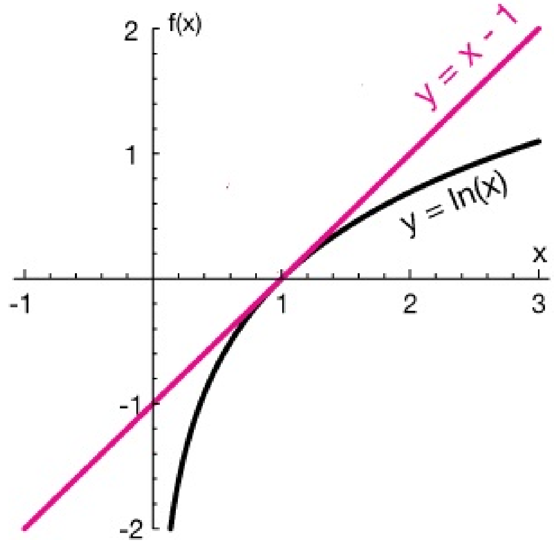


Figure 2: Linear approximation of ln(x)

2 Conditional Entropy

The conditional entropy is a measure of the average uncertainty remaining about random variable X after observing another random variable Y.

$$H(Y|X) \triangleq \sum_{(x,y)} p(x,y) \log \frac{1}{p(y|x)}$$

$$= \sum_{x \in \mathcal{X}} \sum_{y \in \mathcal{Y}} p(x,y) \log \frac{1}{p(y|x)}$$

Example 3 *Binary Symmeric Channel*
The binary symmetric channel(BSC) is defined by the channel diagram shown in fig 3.The common transition probability is denoted by ε *.*

$$\begin{aligned} H(Y|X) &= P_X(0)H(Y|X=0) + P_X(1)H(Y|X=1) \\ &= P_X(0)H_2(\varepsilon, 1-\varepsilon) + P_X(1)H_2(\varepsilon, 1-\varepsilon) = H_2(\varepsilon, 1-\varepsilon) \end{aligned}$$

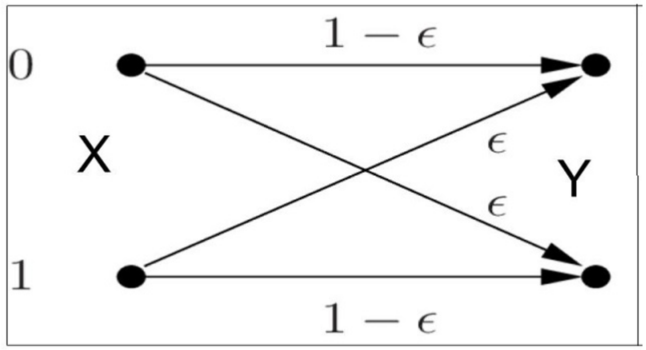


Figure 3: Binary symmetric channel

Note:

$$H(X) = \mathbb{E}\big[\log\frac{1}{P(X)}\big]$$

Consider Y as a function of X, i.e

$$Y = f(X)$$

By the expectation value rule,

$$\mathbb{E}[Y] = \sum_{x \in \mathcal{X}} f(x)p(x)$$

$$\therefore H(Y|X) = \mathbb{E}\left[\log\frac{1}{P(Y|X)}\right]$$

3 Joint Entropy

The join entropy H(X,Y) is the average uncertainty of the random variables X and Y as a whole.
$(x_1, x_2, ..., x_n) \in (\mathcal{X}_1 \times \mathcal{X}_2 \times ...\mathcal{X}_n)$

$$H(X_1, X_2, ..., X_n) \triangleq \sum_{(x_1,x_2,....,x_n)} p(x_1, x_2, ..., x_n) \log\frac{1}{p(x_1, x_2, ..., x_n)}$$

3.1 Chain Rule For Joint Entropy

$$H(X_1, X_2, ..., X_n) = \mathbb{E}\left[\log \frac{1}{P(X_1, X_2, ..., X_n)}\right]$$

$$= \mathbb{E} \log \frac{1}{p(X_1) \prod_{i=2}^{n} p(X_i | X_{[i-1]})}$$

where, $X_{[i-1]} = X_1, X_2, ..., X_{i-1}$.

$$= \mathbb{E} \sum_{i=1}^{n} \log \frac{1}{p(X_i | X_{[i-1]})}$$

$$= \sum_{i=1}^{n} \mathbb{E} \log \frac{1}{p(X_i | X_{[i-1]})}$$

$$= \sum_{i=1}^{n} H(X_i | X_{[i-1]})$$

Thus, in particular

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

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

4 Mutual Information

The mutual information I(X;Y) is the reduction in entropy defined by:

$$I(X;Y) \triangleq H(Y) - H(Y|X)$$

$$= \sum_{y} p(y) \log \frac{1}{p(y)} - \sum_{(x,y)} p(x, y) \log \frac{1}{p(x|y)}$$

Note:

$$\mathbb{E}_{P_Y}\left[\log \frac{1}{P_Y(y)}\right] = \sum_{y \in \mathcal{Y}} p(y) \log \frac{1}{p(y)}$$

$$= \sum_{(x,y)} p(x,y) \log \frac{1}{p(y)}$$

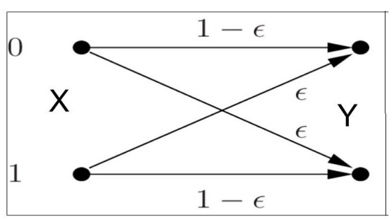
This is the marginalisation sum over all x

Now,

$$\begin{aligned} I(X;Y) &= \mathbb{E}\left[\log \frac{1}{P(Y)} - \log \frac{1}{P(Y|X)}\right] \\ &= \mathbb{E}\left[\log \frac{p(Y|X)}{P(Y)}\right] \\ &= \mathbb{E}\left[\log \frac{P(X,Y)}{P(X)P(Y)}\right] \\ &= H(X) - H(X|Y) \quad (\, from\ the\ symmetry) \end{aligned}$$

$$\therefore \quad H(Y) - H(Y|X) = I(X;Y) = H(X) - H(X|Y)$$

Example 4 *Consider a binary symmetric channel with transition probability ε.*



$$\begin{aligned} I(X;Y) &= H(Y) - H(Y|X) \\ &= H(Y) - H_2(\varepsilon, 1-\varepsilon) \\ &\leq 1 - H_2(\varepsilon, 1-\varepsilon) \end{aligned}$$

equality achieved iff $p_X(0) = p_X(1) = 1/2$

Claim: *I(X;Y)* ≥ 0

Proof:

$$I(X;Y) = \mathbb{E}\left[\log \frac{p(x,y)}{p(x).p(y)}\right]$$

$$-I(X;Y) = \mathbb{E}\left[\log\frac{p(x).p(y)}{p(x,y)}\right]$$
$$\leq \ \mathbb{E}\left[\frac{p(x)p(y)}{p(x,y)} - 1\right] \ (\ from\ aside\ )$$
$$= \sum_{x,y}\frac{p(x,y).p(x).p(y)}{p(x,y)} - \sum_{x,y}p(x.y)$$
$$= 1 - 1 = 0$$

Hence, $I(X;Y) \geq 0$

Corollary:

$$1.H(X|Y) \leq H(X)$$
$$2.H(Y|X) \leq H(Y)$$

Note:

$$\begin{aligned} I(X;Y) &= \mathbb{E}\log\frac{p(x,y)}{p(x).p(y)} \\ &= \mathbb{E}\log\frac{1}{p(x)} + \mathbb{E}\log\frac{1}{p(y)} - \mathbb{E}\log\frac{1}{p(x,y)} \end{aligned}$$

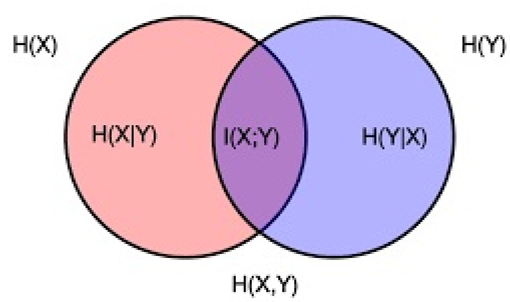


Figure 4: Venn-diagram of mutual information

$$I(X;Y) = H(X) + H(Y) - H(X,Y)$$

5 Conditional Mutual Information

$$I(X;Y|Z) \triangleq H(Y|Z) - H(Y|X,Z)$$
$$= \mathbb{E}\Big[\log \frac{P(Y|X,Z)}{P(Y|Z)}\Big]$$

5.1 Chain Rule of Conditional Mutual Information

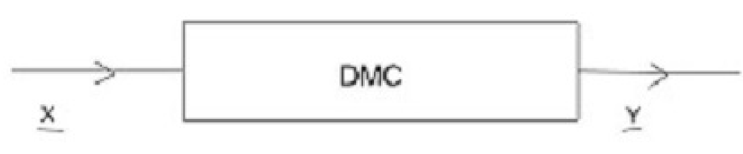
$$I(X_{[n]};Y) = \sum_{i=1}^{n} I(X_i;Y|X_{[i-1]})$$

Proof:

$$I(X_{[n]};Y) = \mathbb{E}\big[\log \frac{p(X_{[n]}|Y)}{p(X_{[n]})}\big]$$
$$= \mathbb{E}\Big[\log \prod_{i=1}^{n} [\frac{p(X_i|Y,X_{[i-1]})}{p(X_i|X_{[i-1]})}]\Big]$$
$$= \sum_{i=1}^{n} \mathbb{E}\Big[\log(\frac{p(X_i|Y,X_{[i-1]})}{p(X_i|X_{[i-1]})})\Big]$$
$$= \sum_{i=1}^{n} [H(X_i|X_{i-1}) - H(X_i|Y,X_{[i-1]})]$$
$$\therefore I(X_{[n]};Y) = \sum_{i=1}^{n} I(X_i;Y|X_{[i-1]})$$

6 Channel Capacity

Consider a discrete memoryless channel



The input $\underline{X}$ consist of input symbols $x_1, x_2, .., x_n$ and the output $\underline{Y}$ consists of output symbols $y_1, y_2, ..., y_n$.

$$P_{\underline{Y}|\underline{X}} = \prod_{i=1}^{n} P_{Y_i|X_i}(y_i|x_i)$$

$$= \prod_{i=1}^{n} P(y_i|x_i)$$

The channel capacity per symbol of a DMC is defined as

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

Thus, the BSC has capacity

$$C = 1 - H_2(\varepsilon, 1-\varepsilon)$$

The capacity has operational meaning as the largest rate R at which information can be reliably transmitted across the channel.
Saying that one is able to transmit reliably at rate R across a DMC is equivalent to saying that there exist a sequence of (n,M=2^{nR}) codes whose associated probability $P_e^{(n)}$ of codeword error goes zero in the limit i.e ,

$$\lim_{n\to\infty} P_e^{(n)} = 0$$

Example 5 *Consider binary erasure channel shown in fig5.*

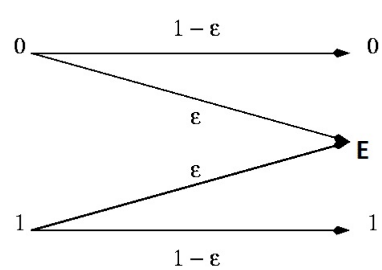


Figure 5: Binary erasure channel

Claim:

$$Capacity(C) = (1 - \varepsilon)$$

Proof:

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

$$H(Y|X) = H_2(\varepsilon, 1 - \varepsilon)$$

To compute H(Y), introduce the random variable Z such that

$$Z = \begin{cases} 1, & Y{=}E \\ 0, & else \end{cases}$$

Then,

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

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

$$\therefore H(Y) = H(Z) + H(Y|Z)$$

$$= H_2(\varepsilon, 1 - \varepsilon) + H(Y|Z = 0)P_Z(0) + H(Y|Z = 1)P_Z(1)$$

$$H(Y|Z) = (1 - \varepsilon)H(Y|Z = 0)$$

$$\leq (1 - \varepsilon)$$

Select $x = \{$ *0,1* $\}$ *with equal probability to get*

$$H(Y|Z) = (1 - \varepsilon)$$

Concluding

$$C = H_2(\varepsilon, 1 - \varepsilon) + (1 - \varepsilon) - H_2(\varepsilon, 1 - \varepsilon)$$

$$= (1 - \varepsilon)$$