UNIT-V

Information theory and coding

Entropy

Let X be a discrete random variable and p(x) be the probability mass function. The entropy H(x) of x is defined by

$$H(X) = -\sum_{x} p(x) \log_2 p(x) = H(P)$$

↳ amount of information

Quantity entropy is measured in bits ($\log_2$). If the base of the logarithm is b we denote the entropy as $H_b(x)$. If base is e then entropy is measured in nats.

Entropy depends not on the value taken by the random varible X, but on the probability of occurance.

Ex: Entropy of a fair coin toss is 1 bits

& $0 \log 0 = 0$ as $x \log x = 0$ as $x \to 0$

According turns of zero probability does not change the entropy.

Expected value: Expected value of a random varible g(x) is denoted by

$E[g(x)] = \sum g(x) p(x)$ where p(x) is probability distribution (pmf) of X.

Note: If $g(x) = \log\left[\frac{1}{p(x)}\right]$ then

$$E\left[\log\left(\frac{1}{p(x)}\right)\right] = \sum p(x) \log\left(\frac{1}{p(x)}\right) = H(p)$$

i.e entropy of X can also be interpreted as the expected value of the random variable $\log\left(\frac{1}{p(x)}\right)$

Properties of $H(x)$

Lemma ① $H(x) \geq 0$

Proof: By defn $H(x) = -\sum_{x} p(x) \log[p(x)]$

$$H(x) = \sum p(x) \log\left(\frac{1}{p(x)}\right)$$

since $0 \leq p(x) \leq 1 \Rightarrow \log\left(\frac{1}{p(x)}\right) \geq 0$

$$\therefore \boxed{H(x) \geq 0}$$

Lemma ② $H_b(x) = (\log_b a) H_a(x)$

Proof: By defn $H_b(x) = -\sum p(x) \log_b p(x)$

$$H_b(x) = -\sum p(x) \frac{\log p(x)}{\log b} \times \frac{\log a}{\log a}$$

$$= -\sum p(x) \log_a p(x) \log_b a$$

$$= -\log_b a \sum p(x) \log_a p(x)$$

$$\boxed{H_b(x) = (\log_b a) H_a(x)}$$

Examples

① Let $X = \begin{cases} 0 & \text{with } p=\frac{1}{2} \\ 1 & \text{with } p=\frac{1}{2} \end{cases}$ then find $H(X)$

Solution:

$$H(X) = -\sum p(x)\log_2 p(x)$$

$$H(X) = -\frac{1}{2}\log_2\frac{1}{2} - \frac{1}{2}\log_2\frac{1}{2}$$

$$H(X) = -\log_2\frac{1}{2} = 1 \text{ bit}$$

① $\log_2 2 = \frac{\log 2}{\log 2} = 1$

② $\log\frac{1}{2} = \log 2^{-1} = -\log 2$

③ $-\log_2\frac{1}{2} = -\frac{(-\log 2)}{\log 2} = 1$

i.e tossing of a coin will have one bit of information

② Let X denote the value shown up when a die is rolled with equally likely for all six events, find H(X)

Solution:

X:	1	2	3	4	5	6
p(x)	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$

$$H(X) = -\sum p(x)\log_2 p(x) = -6\times\frac{1}{6}\log_2\frac{1}{6}$$

$$= -\log_2\frac{1}{6} = \log_2 6 = \frac{\log 6}{\log 2} = 2.585 \text{ bits}$$

③ Let $X = \begin{cases} a & \text{with } p=\frac{1}{2} \\ b & \text{with } p=\frac{1}{4} \\ c & \text{with } p=\frac{1}{8} \\ d & \text{with } p=\frac{1}{8} \end{cases}$

(OR)

X	a	b	c	d
p(x)	$\frac{1}{2}$	$\frac{1}{4}$	$\frac{1}{8}$	$\frac{1}{8}$

, then find entropy of X

Solution:

$$H(X) = -\sum p(x)\log p(x)$$

$$= -\left[\frac{1}{2}\log\frac{1}{2} + \frac{1}{4}\log\frac{1}{4} + \frac{1}{8}\log\frac{1}{8} + \frac{1}{8}\log\frac{1}{8}\right]$$

$$= -\left[-\frac{1}{2}\log 2 - \frac{1}{4}\log 4 - \frac{1}{8}\log 8 - \frac{1}{8}\log 8\right]$$

$$= \frac{1}{2}\log 2 + \frac{2}{4}\log 2 + \frac{3}{4}\log 2 + \frac{3}{4}\log 2$$

$$= \left(\frac{1}{2} + \frac{1}{2} + \frac{3}{4} + \frac{3}{4}\right)\log_2 2$$

$$= \frac{14}{8}$$

$$= \frac{7}{4} = 1.75 \text{ bits}$$

$\log m^n = n\log m$
$\log 8 = \log 2^3$
$= 3\log 2$

4. A fair coin is tossed untill the first head occurs. Let X denotes the number of tosses required. Find the entropy $H(X)$ is bits. Use $\sum_{n=0}^{\infty} r^n = \frac{1}{1-r}$ & $\sum_{n=0}^{\infty} n r^n = \frac{r}{(1-r)^2}$

Solution:

1) $(1-a)^{-1} = 1 + a + a^2 + a^3 + \cdots\cdots$

2) $(1-a)^{-2} = 1 + 2a + 3a^2 + 4a^3 + \cdots$

Solution: Here 'X' is a geometric distribution

$\therefore f(x) = f(n) = (1-p)^{n-1}p = q^{n-1}p, \; n = 1, 2, 3 \cdots$

$$H(X) = -\sum p(x)\ln_2 p(x)$$

$$H(X) = -\sum_{n=1}^{\infty} q^{n-1} p \, \ln_2 (q^{n-1} p)$$

$$= -\sum_{n=1}^{\infty} q^{n-1} p \left[(n-1) \ln_2 q + \ln_2 p \right]$$

$$= -\sum_{n=1}^{\infty} q^{n-1} p (n-1) \ln_2 q - \sum_{n=1}^{\infty} q^{n-1} p \ln_2 p$$

put $n-1=t$ $\therefore$ $n=1$ $t=0$

$n=\infty$ $t=\infty$

$$= \sum_{t=0}^{\infty} q^{t} t p \ln_2 q - \sum_{t=0}^{\infty} q^{t} p \ln_2 p$$

$$= -p \ln_2 q \sum_{t=0}^{\infty} t q^{t} - p \ln_2 p \sum_{t=0}^{\infty} q^{t}$$

$$= -p \ln_2 q \frac{q}{(1-q)^2} - p \ln_2 p \frac{1}{1-q}$$

$$= -\cancel{p} q \frac{\ln_2 q}{p^{\cancel{2}}} - \cancel{p} \frac{\ln_2 p}{\cancel{p}}$$

$$= -\frac{1}{p} \left[p \ln_2 p + q \ln_2 q \right]$$

$$H(X) = \frac{1}{p} H(p) \text{ bits}$$

where $H(p) = -p \ln_2 p - (1-p) \ln_2 (1-p)$

<u>Note</u> If $p = \frac{1}{2}$ then $H(X) = 2$

If $p = \frac{1}{4}$ then $H(X) = 0.81127 \times 4 = 3.2451$

If $p = \frac{1}{8}$ then $H(X) = 0.5435 \times 8 = 4.35$

Ex: Let $X = \begin{cases} 1 & \text{with pr } p \\ 0 & \text{with pr } (1-p) \end{cases}$

Soln: By defn

$$H(X) = -p\log_2 p - (1-p)\log_2(1-p) = H(p)$$

When $p = \frac{1}{2}$

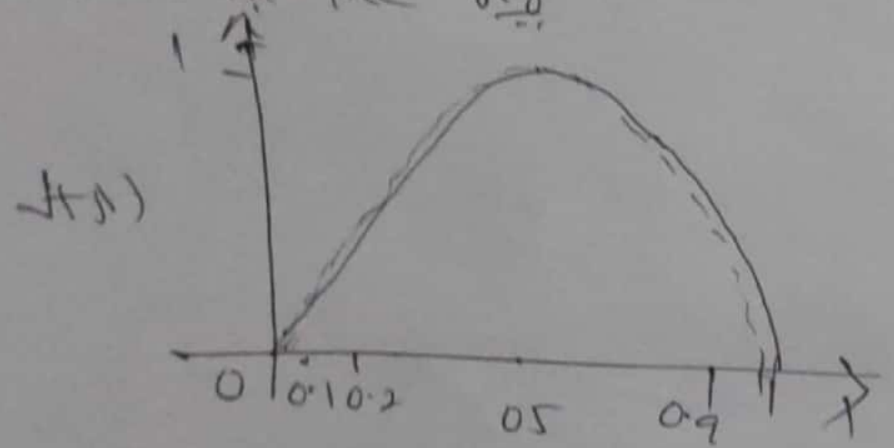
$$H(X) = -\frac{1}{2}\log_2\left(\frac{1}{2}\right) - \left(\frac{1}{2}\right)\log_2\left(\frac{1}{2}\right)$$

$$= \frac{1}{2}\log_2 2 + \frac{1}{2}\log_2 2$$

$$= \frac{1}{2} + \frac{1}{2} = 1$$

$\log_2\left(\frac{1}{2}\right)$
$= \log_2 1 - \log_2 2$
$= -\log_2 2$

The graph of the function $H(p)$ is as shown in the fig



Joint Entropy

Joint entropy $H(X,Y)$ is a pair of discrete random variables (X,Y) with a joint distribution $p(x,y)$ is defined as

$$H(X,Y) = -\sum_{x}\sum_{y} p(x,y) \log [p(x,y)]$$

which can ~~b~~also be written as

$$H(X,Y) = -E[\log p(X,Y)] = E\left[\log \frac{1}{p(X,Y)}\right]$$

Conditional Entropy:

Conditional entropy for a pair of discrete random variables (X,Y) with a joint distribution $p(x,y)$ is denoted by $H(Y/X)$ is defined by

$$H(Y/X) = \sum_{x} p(x)\, H(Y/X=x)$$

$$= -\sum_{x} p(x) \sum_{y} p(y/x) \log p(y/x)$$

$$= -\sum_{x}\sum_{y} p(x)\, p(y/x) \log p(y/x)$$

$$= -\sum_{x}\sum_{y} p(x,y) \log p(y/x)$$

$$= -E[\log p(Y/X)]$$

Note x & y are independent events

$$P(x,y) = P(x)\, P(y/x)$$

Theorem: Chain Rule

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

Proof consider

$$H(X,Y) = -\sum_x \sum_y P(x,y) \log P(x,y)$$

$$= -\sum_x \sum_y P(x,y) \log [P(x)\, P(y/x)]$$

$$= -\sum_x \sum_y P(x,y) \log P(x) - \sum_x \sum_y P(x,y) \log P(y/x)$$

$$= -\sum_x P(x) \log P(x) - \sum_x \sum_y P(x,y) \log (y/x)$$

$$\boxed{H(X,Y) = H(X) + H(Y/X)}$$

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

Note ① $H(X/Y) \neq H(Y/X)$

② $H(X,Y/Z) = H(X/Z) + H(Y/X,Z)$

③ $H(X) - H(X/Y) = H(Y) - H(Y/X)$

Examples

1) Let (X, Y) have the following joint distribution

$Y \backslash X$	1	2	3	4
1	$\frac{1}{8}$	$\frac{1}{16}$	$\frac{1}{32}$	$\frac{1}{32}$
2	$\frac{1}{16}$	$\frac{1}{8}$	$\frac{1}{32}$	$\frac{1}{32}$
3	$\frac{1}{16}$	$\frac{1}{16}$	$\frac{1}{16}$	$\frac{1}{16}$
4	$\frac{1}{4}$	0	0	0

Compute $H(X,Y)$, $H(Y/X)$, $H(X/Y)$, $H(X)$, $H(Y)$

Solution: The marginal distribution of X & Y are

X	1	2	3	4
$f(x)$	$\frac{1}{2}$	$\frac{1}{4}$	$\frac{1}{8}$	$\frac{1}{8}$

Y	1	2	3	4
$g(y)$	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{4}$

$$H(X) = -\sum p(x) \log_2 p(x)$$

$$H(X) = -\left[\frac{1}{2}\log\left(\frac{1}{2}\right) + \frac{1}{4}\log\left(\frac{1}{4}\right) + \frac{1}{8}\log\left(\frac{1}{8}\right) + \frac{1}{8}\log\left(\frac{1}{8}\right)\right]$$

$$H(X) = -\left[-\frac{1}{2}\log 2 - \frac{1}{4}\log 4 - \frac{1}{8}\log 8 - \frac{1}{8}\log 8\right]$$

$$H(X) = \frac{1}{2}\log 2 + \frac{1}{2}\log 2 + \frac{3}{8}\log 2 + \frac{3}{8}\log 2$$

$$= \frac{1}{2} + \frac{1}{2} + \frac{3}{8} + \frac{3}{8} = \frac{7}{4} = 1.75$$

$$H(y) = -\sum P(y)\, \ln_2 P(y)$$

$$= -\left[\frac{1}{4}\ln_2\left(\frac{1}{4}\right) + \frac{1}{4}\ln_2\left(\frac{1}{4}\right) + \frac{1}{4}\ln_2\left(\frac{1}{4}\right) + \frac{1}{4}\ln_2\left(\frac{1}{4}\right)\right]$$

$$= -\left[-\frac{1}{4}\log 4 - \frac{1}{4}\log 4 - \frac{1}{4}\log 4 - \frac{1}{4}\log 4\right]$$

$$= \left[\frac{1}{2} + \frac{1}{2} + \frac{1}{2} + \frac{1}{2}\right]\log 2$$

$$\boxed{H(y) = 2}$$

$$H(X/y) = \sum P(y)\, H(X/y=y)$$

$$= \frac{1}{4}H\left(\frac{\frac{1}{8}}{\frac{1}{4}}, \frac{\frac{1}{16}}{1/4}, \frac{\frac{1}{32}}{1/4}, \frac{\frac{1}{32}}{\frac{1}{4}}\right) + \frac{1}{4}H\left(\frac{\frac{1}{16}}{\frac{1}{4}}, \frac{\frac{1}{8}}{\frac{1}{4}}, \frac{\frac{1}{32}}{\frac{1}{4}}, \frac{\frac{1}{32}}{\frac{1}{4}}\right)$$

$$+ \frac{1}{4}H\left(\frac{\frac{1}{16}}{\frac{1}{4}}, \frac{\frac{1}{16}}{\frac{1}{4}}, \frac{\frac{1}{16}}{\frac{1}{4}}, \frac{\frac{1}{16}}{\frac{1}{4}}\right) + \frac{1}{4}H\left(\frac{\frac{1}{4}}{\frac{1}{4}}, \frac{0}{\frac{1}{4}}, \frac{0}{\frac{1}{4}}, \frac{0}{\frac{1}{4}}\right)$$

$$= \frac{1}{4}\left[H\left(\frac{1}{2}, \frac{1}{4}, \frac{1}{8}, \frac{1}{8}\right) + H\left(\frac{1}{4}, \frac{1}{2}, \frac{1}{8}, \frac{1}{8}\right) + H\left(\frac{1}{4}, \frac{1}{4}, \frac{1}{4}, \frac{1}{4}\right) + H(1, 0, 0, 0)\right]$$

$$= \frac{1}{4}\left[1.75 + 1.75 + 2 + 0\right] = 1.375$$

$$\boxed{H(X/y) = 1.375}$$

$$H(Y/X) = P(x)\ H(Y/X=x)$$

$$= \frac{1}{2} H\left(\frac{\frac{1}{8}}{\frac{1}{2}}, \frac{\frac{1}{16}}{\frac{1}{2}}, \frac{\frac{1}{16}}{\frac{1}{2}}, \frac{\frac{1}{4}}{\frac{1}{2}}\right) + \frac{1}{4} H\left(\frac{\frac{1}{16}}{\frac{1}{4}}, \frac{\frac{1}{8}}{\frac{1}{4}}, \frac{\frac{1}{16}}{\frac{1}{4}}, \frac{0}{\frac{1}{4}}\right)$$

$$+ 2 \times \frac{1}{8} H\left(\frac{\frac{1}{32}}{\frac{1}{8}}, \frac{\frac{1}{32}}{\frac{1}{8}}, \frac{\frac{1}{16}}{\frac{1}{8}}, \frac{0}{\frac{1}{8}}\right)$$

$$= \frac{1}{2} H\left(\frac{1}{4}, \frac{1}{8}, \frac{1}{8}, \frac{1}{2}\right) + \frac{1}{4} H\left(\frac{1}{4}, \frac{1}{2}, \frac{1}{4}, 0\right)$$

$$+ \frac{1}{4} H\left(\frac{1}{4}, \frac{1}{4}, \frac{1}{2}, 0\right)$$

$$H(Y/X) = \frac{1}{2}(1.75) + \frac{1}{4}(1.5) + \frac{1}{4}(1.5) = 1.625$$

$$\boxed{H(Y/X) = 1.625}$$

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

$$H(X,Y) = 1.75 + 1.625 = 3.375$$

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

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

(2) Let $p(x, y)$ be given by

$x \backslash y$	0	1
0	$\frac{1}{3}$	$\frac{1}{3}$
1	0	$\frac{1}{3}$

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

Solution: The marginal distributions of X & Y are

X	0	1
$p(x)$	$\frac{2}{3}$	$\frac{1}{3}$

Y	0	1
$q(y)$	$\frac{1}{3}$	$\frac{2}{3}$

$$H(X) = -\sum_x p(x)\log p(x)$$

$$= -\frac{2}{3}\log\frac{2}{3} - \frac{1}{3}\log\left(\frac{1}{3}\right) = -\frac{2}{3}\left[\log 2 - \log 3\right] - \frac{1}{3}\left[\log 1 - \log 3\right]$$

$$= \frac{2}{3}\log 3 + \frac{1}{3}\log 3 - \frac{2}{3}$$

$$= \frac{2}{3}\frac{\log 3}{\log 2} + \frac{1}{3}\frac{\log 3}{\log 2} - \frac{2}{3}$$

$$= 0.9183$$

$$H(Y) = -\sum_y p(y)\log p(y)$$

$$= 0.9183$$

$$H(X/Y) = \sum p(y)\, H(X/Y=y)$$

$$= \frac{1}{3}H\left(\frac{\frac{1}{3}}{\frac{1}{3}}, \frac{0}{\frac{1}{3}}\right) + \frac{2}{3}H\left(\frac{\frac{1}{3}}{\frac{2}{3}}, \frac{\frac{1}{3}}{\frac{2}{3}}\right)$$

$$= \frac{1}{3}H(1, 0) + \frac{2}{3}H\left(\frac{1}{2}, \frac{1}{2}\right)$$

$$= \frac{1}{3}(0) + \frac{2}{3}(1)$$

$$= \frac{2}{3}$$

(12)

$$H(Y/X) = \sum P(x)\, H(Y/X=x)$$

$$= \frac{2}{3} H\left(\frac{\frac{1}{3}}{\frac{2}{3}}, \frac{\frac{1}{3}}{\frac{2}{3}}\right) + \frac{1}{3} H\left(\frac{0}{\frac{1}{3}}, \frac{\frac{1}{3}}{\frac{1}{3}}\right)$$

$$= \frac{2}{3} H\left(\frac{1}{2}, \frac{1}{2}\right) + \frac{1}{3} H(0, 1)$$

$$= \frac{2}{3}(1) + \frac{1}{3}(0) = \frac{2}{3} = 0.6666$$

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

$$= 0.9183 + 0.6666$$

$$= 1.5849$$

$$H(Y) - H(Y/X) = 0.9183 - 0.6666 = 0.2517$$

$$H(X) - H(X/Y) = 0.9183 - 0.6666 = 0.2517$$

OR

$$H(X,Y) = -\sum_x \sum_y p \log_2 p$$

$$= -\left[\frac{1}{3}\log\frac{1}{3} + \frac{1}{3}\log\frac{1}{3} + \frac{1}{3}\log\frac{1}{3}\right]$$

$$= -3 \times \frac{1}{3}\log\left(\frac{1}{3}\right) = -\log\frac{1}{3} = -[\log 1 - \log 3]$$

$$= \log_2 3 = \frac{\log 3}{\log 2} = 1.5849$$

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

$$= 1.5849 - 0.9182 = 0.6666 = \frac{2}{3}$$

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

$$= 1.5849 - 0.9182 = 0.6666 = \frac{2}{3}$$

Relative Entropy

It is a measure of distance between two distributions. Relative entropy is also called Kullback-Leibler distance.

Definition:

The relative entropy between two pmf's $p(x)$ & $q(x)$ is denoted by $D(p\|q)$ & is defined by

$$D(p\|q) = \sum_x p(x) \log \frac{p(x)}{q(x)} = E\left[\log \frac{p(x)}{q(x)}\right]$$

Relative entropy is a measure of the inefficiency of assuming that the distribution is q when the true distribution is p.

If $p(x) > 0$ & $q(x) = 0$ then $D(p\|q) = \infty$

Note 1) If $p = q$, then $D(p\|q) = \sum_x p(x) \log 1 = 0$

2) We use the convention that

$$0 \log \frac{0}{0} = 0, \quad 0 \log \frac{0}{p} = 0, \quad 0 \log \frac{p}{0} = \infty$$

Mutual Information:

Consider two random variables X & Y with a joint distribution $p(x,y)$ & marginal distribution $p(x)$ & $p(y)$.

The mutual information $I(X:Y)$ is the "relative entropy between joint distribution and product distribution $p(x)p(y)$"

$$I(X,Y) = \sum_x \sum_y p(x,y) \log \frac{p(x,y)}{p(x)p(y)} = D\left(p(x,y) \,||\, p(x)p(y)\right)$$

"Mutual information is a measure of the amount of information that one random variable contains about another random variable".

It is reduction of uncertainity of one random variable due to the knowledge of the other.

<u>Relation between entropy and mutual information:</u>

Note 1) $p(x,y) = p(y)\, p(x/y)$
2) $p(x,y) = p(x)\, p(y/x)$

we know that

$$I(X,Y) = \sum_x \sum_y p(x,y) \log \frac{p(x,y)}{p(x)p(y)}$$

$$= \sum_x \sum_y p(x,y) \log \frac{p(x,y)/p(y)}{p(x)}$$

$$= \sum_x \sum_y p(x,y) \log \frac{p(x/y)}{p(x)}$$

$$I(X,Y) = \sum_x \sum_y p(x,y) \log p(x/y) - \sum_x \sum_y p(x,y) \log p(x)$$

$$= -H(X/Y) - \sum_x p(x) \log p(x)$$

$$= -H(X/Y) + H(X) \quad \therefore \boxed{I(X,Y) = H(X) - H(X/Y)}$$

Reduction in uncertainity of X due to knowledge of Y.

Note. ① $I(X,Y) = H(X) - H(X/Y)$

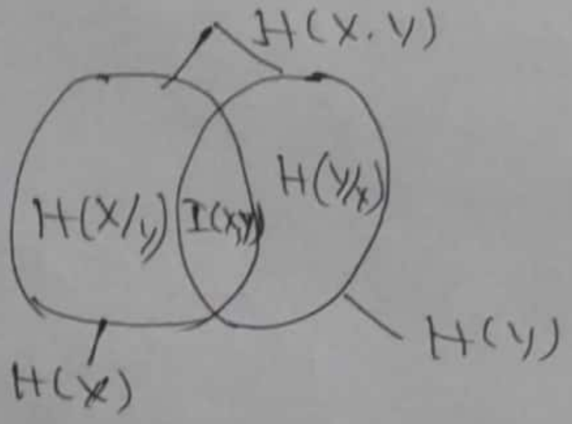
② $I(X,Y) = H(Y) - H(Y/X)$

③ $I(X,Y) = H(X) + H(Y) - H(X,Y)$

④ $I(X,Y) = I(Y,X)$

⑤ $I(X,Y) = H(X)$

Venn diagram relating to various quantities



Examples

1) Let $X = \{0, 1\}$ and consider two distribution p and q on X.

Let $p(0) = 1-r$, $p(1) = r$

$q(0) = 1-s$, $q(1) = s$

find $D(p||q)$, $D(q||p)$

Solution:

$$D(p||q) = \sum_x p(x) \log \frac{p(x)}{q(x)}$$

$$D(p||q) = (1-r)\log\left(\frac{1-r}{1-s}\right) + r\log\left(\frac{r}{s}\right)$$

$$D(q||p) = \sum_x q(x) \log \frac{q(x)}{p(x)}$$

$$D(q||p) = (1-s)\log\left(\frac{1-s}{1-r}\right) + s\log\left(\frac{s}{r}\right)$$

If $p = q$, then $D(p||q) = 0$, $D(q||p) = 0$

If $r = \frac{1}{2}$, $s = \frac{1}{3}$

$$D(p||q) = \frac{1}{2}\log\frac{\frac{1}{2}}{\frac{2}{3}} + \frac{1}{2}\log\left(\frac{\frac{1}{2}}{\frac{1}{3}}\right) = \frac{1}{2}\left[\log\frac{3}{4} + \log\frac{3}{2}\right]$$

$$= \frac{1}{2}\left[\log 3 - \log 4 + \log 3 - \log 2\right]$$

$$= \frac{1}{2}\left[2\log 3 - \log 4 - \log 2\right]$$

$$= \frac{1}{2}\left[2\frac{\log 3}{\log 2} - \frac{\log 4}{\log 2} - 1\right]$$

$$= 0.085$$

$D(p||q) = 0.085$ bit

$$D(q||p) = \frac{2}{3}\log\frac{\frac{2}{3}}{\frac{1}{2}} + \frac{1}{3}\log\frac{\frac{1}{3}}{\frac{1}{2}}$$

$D(q||p) = 0.082$ bit

i.e $D(p||q) \neq D(q||p)$

② calculate the mutual information for Example ① & ② (Q-B-12)

Solution 1) $I(X,Y) = H(X) - H(X/Y) = 0.375$

2) $I(X,Y) = 0.25163$

Note: ① $D(P\|Q) = \frac{1}{2}\left[\log\left(\frac{3}{4}\right) + \log\left(\frac{3}{2}\right)\right]$

$= \frac{1}{2}\left[\log 3 - \log 4 + \log 3 - \log 2\right]$

$= \frac{1}{2}\left[2\log 3 - \log 4 - \log 2\right]$

$= \frac{1}{2}\left[2\frac{\log 3}{\log 2} - \frac{\log 4}{\log 2} - 1\right]$

$= 0.085$

② $D(Q\|P) = \frac{2}{3}\log\frac{4}{3} + \frac{1}{3}\log\frac{2}{3}$

$= \frac{2}{3}\left[\log 4 - \log 3\right] + \frac{1}{3}\left[\log 2 - \log 3\right]$

$= \frac{2}{3}\log 4 - \frac{2}{3}\log 3 + \frac{1}{3} - \frac{1}{3}\log 3$

$= \frac{1}{3} + \frac{2}{3}\log 4 - \log 3.$

$= \frac{1}{3} + \frac{2}{3}\left[\frac{\log 4}{\log 2}\right] - \left[\frac{\log 3}{\log 2}\right]$

$= 0.082$

Chain Rules for entropy, relative entropy and mutual information:

Theorem: Let $X_1, X_2 \cdots X_n$ be drawn according to $p(x_1, x_2, \cdots x_n)$ then

$$H(X_1, X_2, \cdots X_n) = \sum_{i=1}^{n} H(X_i / X_{i-1}, X_{i-2} \cdots X_1)$$

Proof:

$$H(X_1, X_2 \cdots X_n) = -\sum_{x} p(x_1 x_2 \cdots x_n) \log p(x_1, x_2 \cdots x_n) \rightarrow (1)$$

but
$$p(x_1, x_2) = p(x_1) p(x_2/x_1)$$
$$= \prod_{i=1}^{2} p(x_i/x_{i-1})$$

$$p(x_1, x_2 x_3) = p(x_1) p(x_2/x_1) \cdot p(x_3/x_2 x_1)$$
$$= \prod_{i=1}^{3} p(x_i/x_{i-1} \cdots x_1)$$

$$p(x_1 x_2 x_3 \cdots x_n) = \prod_{i=1}^{n} p(x_i/x_{i-1}, x_{i-2} \cdots x_1)$$

$$\log p(x_1, x_2) = \log \prod_{i=1}^{2} p(x_i/x_{i-1}) = \sum_{i=1}^{2} p(x_i/x_{i-1})$$

$$\log p(x_1, x_2 \cdots x_n) = \log \prod_{i=1}^{n} p(x_i/x_{i-1}, x_{i-2} \cdots x_1)$$
$$= \sum_{i=1}^{n} p((x_i/x_{i-1}, x_{i-2} \cdots x_1)$$

Substitute in eqn (1)

$$H(X_1, X_2 \cdots X_n) = -\sum_x p(x_1, x_2 \cdots x_n) \times \sum_{i=1}^{n} p(x_i / x_{i-1} \cdots x_1)$$

$$= -\sum_{i=1}^{n} \sum_x p(x_1, x_2 \cdots x_n)\, p(x_i / x_{i-1} \cdots x_1)$$

$$= \sum_{i=1}^{n} H(X_i / X_{i-1}, X_{i-2} \cdots X_1)$$

i.e entropy of collection of random variables is the sum of the conditional entropies.

Definition:-

The conditional mutual information of random variables X & Y given Z is defined by

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

Definition:

For joint probability mass function $p(x, y)$ & $q(x, y)$ the conditional relative entropy $D[p(y/x) \| q(y/x)]$ is the average of the relative entropies between the conditional probability mass function $p(y/x)$ and $q(y/x)$ averaged over the probability mass function $p(x)$.

i.e $$D[p(y/x) \| q(y/x)] = \sum_x p(x) \sum_y p(y/x) \log \frac{p(y/x)}{q(y/x)}$$

Jensen's Inequality and its consequences:-

Definition:

A function $f(x)$ is said to be convex over an interval (a,b) if for every $x_1, x_2 \in (a,b)$ and $0 \leq \lambda \leq 1$

$$f[\lambda x_1 + (1-\lambda)x_2] \leq \lambda f(x_1) + (1-\lambda) f(x_2)$$

function is strictly convex if equality holds i.e $\lambda = 0$ and $\lambda = 1$

Definition:

A function is concave if f is convex

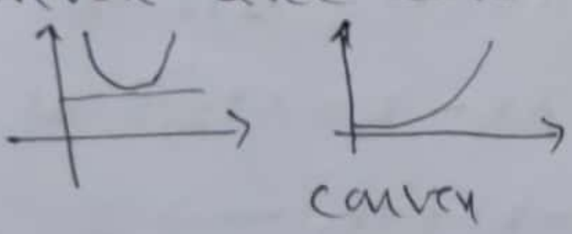
Note: 1) A function is convex if it is below a chord

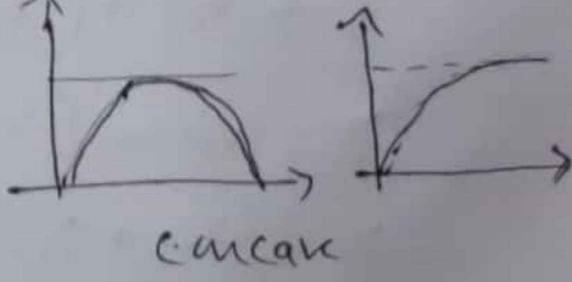
2) A function is concave if it always lies above any chord

Ex: x^2, $|x|$, e^x, $x\log x$ $(x \geq 0)$ are convex. $\cup$

$\log x$, $\sqrt{x}$ are concave $\cap$

The linear functions $ax+b$ are both convex and concave





Theorem:-

If the function f has a second derivative that is positive over an interval then the function is convex over that interval.

Proof: consider the Taylors series expansion of $f(x)$ about x_0

i-e $f(x) = f(x_0) + (x-x_0)f'(x_0) + \frac{(x-x_0)^2}{\lfloor 2} f''(x_0) + \cdots \rightarrow (1)$

where $x^* \in [x_0, x]$

By hypothesis $f''(x^*) \geq 0$, thus last term is +ve, $\forall x$

Let $x_0 = \lambda x_1 + (1-\lambda)x_2$ & take $x = x_1$ in (1)

$$f(x_1) \geq f(x_0) + (x_1 - x_0) f'(x_0)$$

$$\geq f(x_0) + [x_1 - \lambda x_1 - (1-\lambda)x_2] f'(x_0)$$

$$\geq f(x_0) + [x_1(1-\lambda) - (1-\lambda)x_2] f'(x_0)$$

$$f(x_1) \geq f(x_0) + (1-\lambda)(x_1 - x_2) f'(x_0) \longrightarrow (2)$$

Taking $x = x_2$ in (1)

$$f(x_2) \geq f(x_0) + (x_2 - x_0) f'(x_0)$$

$$\geq f(x_0) + [x_2 - \lambda x_1 - (1-\lambda)x_2] f'(x_0)$$

$$\geq f(x_0) + [x_2 - \lambda x_1 - x_2 + \lambda x_2] f'(x_0)$$

$$\geq f(x_0) + \lambda(x_2 - x_1) f'(x_0) \longrightarrow (3)$$

$② \times \lambda + ③ (1-\lambda) \Longrightarrow$

$$\lambda f(x_1) + (1-\lambda) f(x_2) \geq \lambda f(x_0) + \lambda(\underline{1-\lambda})(x_1 - x_2) f'(x_0)$$
$$+ (1-\lambda) f(x_0) + \lambda(\underline{1-\lambda})(x_2 - x_1) f'(x_0)$$

$$\lambda f(x_1) + (1-\lambda) f(x_2) \geq \lambda f(x_0) + (1-\lambda) f(x_0)$$
$$\geq f(\underline{x_0})$$

$$\lambda f(x_1) + (1-\lambda) f(x_2) \geq f(\lambda x_1 + (1-\lambda) x_2)$$

$$\therefore f(\lambda x_1 + (1-\lambda) x_2) \leq \lambda f(x_1) + (1-\lambda) f(x_2)$$

$\therefore f$ is convex

Theorem:- Jensen's Inequality:

If f is a convex function and X is a random variable, then prove that $Ef(X) \geq f(EX)$. Moreover if f is strictly convex then $X = EX$ with probability 1. (i.e X is a constant)

Proof:- consider a finite discrete distribution $X = (x_1, x_2, \cdots x_n)$ with pmf $p(x_1)\, p(x_2), \cdots p(x_n)$

$$Ef(x) \geq f(EX) \longrightarrow ①$$

$$p_1 f(x_1) + p_2 f(x_2) + \cdots + p_n f(x_n) \geq f(x_1 p_1 + x_2 p_2 + \cdots + x_n p_n) \longrightarrow ②$$

we prove ② by Mathematical Induction

for $n=2$, $p_1 f(x_1) + p_2 f(x_2) \geq f(x_1 p_1 + x_2 p_2)$

by defn of f convex.

$\therefore$ Result ② is true for $n=k$

$$p_1 f(x_1) + p_2 f(x_2) + \cdots + p_k f(x_k) \geq f(x_1 p_1 + x_2 p_2 + \cdots + x_k p_k) \longrightarrow ③$$

we prove ② for $n = k+1$

consider L.H.S of Eqn ②

$$= \sum_{i=1}^{k+1} p_i f(x_i)$$

$$= \sum_{i=1}^{k} p_i f(x_i) + p_{k+1} f(x_{k+1})$$

Put $p_i = p_i'(1-p_{k+1})$ for $i=1,2\cdots k$

LHS of ② $= \sum_{i=1}^{k} f(x_i)\, p_i'(1-p_{k+1}) + p_{k+1} f(x_{k+1})$

$= (1-p_{k+1}) \sum_{i=1}^{k} p_i' f(x_i) + p_{k+1} f(x_{k+1})$

$\geq p_{k+1} f(x_{k+1}) + (1-p_{k+1}) f\left(\sum_{i=1}^{k} p_i' x_i\right]$ from ③

$\geq f\left[\sum_{i=1}^{k+1} p_i x_i\right]$ since f is convex

Thus the result is true for $n=k+1$

Hence by mathematical induction ① is true for all n.

<u>Theorem</u> [<u>Information Inequality</u>]

Let $p(x), q(x)$ for $x \in X$ be two pmf's then $D(p\|q) \geq 0$ with Equality iff $p(x)=q(x)\ \forall x$

<u>Proof</u>: By <u>defn</u>

$$D(p\|q) = \sum_x p(x) \log \frac{p(x)}{q(x)}$$

$$-D(p\|q) = \sum_x p(x) \log \frac{q(x)}{p(x)}$$

$$\leq \log \sum_x p(x) \cdot \frac{q(x)}{p(x)}$$

$$\leq \log \sum_x q(x) = \log 1 = 0$$

$-D(p\|q) \leq 0$

$\therefore D(p\|q) \geq 0$

$\log x$ is concave but $-\log x$ is convex

$\sum p(x) \log \frac{q(x)}{p(x)}$

$= E\left[\log \frac{q(x)}{p(x)}\right]$

$\leq \log E\left[\frac{q(x)}{p(x)}\right]$

Corollary: (Non negativity of mutual information).

For any two random variables X, Y, $I(X, Y) \geq 0$ with equality iff X & Y are independent.

Proof: $I(X, Y) = D(p(x,y) \| p(x)p(y)) \geq 0$

$p(x,y) = p(x)p(y)$ if X & Y are independent

Theorem:- (conditioning reduces entropy / ∴ Information can't hurt)

prove that

$H(X/Y) \leq H(X)$ with equality iff X & Y are independent

Proof: we have

$$I(X, Y) = H(X) - H(X/Y) \geq 0$$

$$\therefore H(X) \geq H(X/Y) \quad \text{OR}$$

$$H(X/Y) \leq H(X)$$

Theorem: (Independent bound on entropy)

Let $X_1, X_2 \cdots X_n$ be drawn accordingly to $p(x_1, x_2, \cdots x_n)$ then $H(X_1, X_2, \cdots X_n) \leq \sum_{i=1}^{n} H(X_i)$

Proof: By chain rule for entropies

$$H(X_1, X_2, \cdots X_n) = \sum_{i=1}^{n} H(X_i / X_{i-1} \cdots X_1)$$

$$\leq \sum_{i=1}^{n} H(X_i) \text{ (by above theorem)}$$

State and Prove log sum Inequality

Theorem: For non negative number $a_1, a_2, \cdots a_n$ and $b_1, b_2, \cdots b_n$

$$\sum_{i=1}^{n} a_i \log \frac{a_i}{b_i} \geq \sum_{i=1}^{n} a_i \log \frac{\sum_{i=1}^{n} a_i}{\sum_{i=1}^{n} b_i}$$

with Equality iff $\frac{a_i}{b_i} = \text{constant}$

Proof: Given $a_i > 0$ and $b_i > 0$

The function $f(t) = t \log t$ is convex $\rightarrow$ ①

since $f'(t) = 1 + \log t$ & $f'' = \frac{1}{t} > 0 \ \forall t > 0$

By Jensen's inequality

$$\sum \alpha_i f(t_i) \geq f(\sum \alpha_i t_i)$$

for $\alpha_i \geq 0$, $\sum \alpha_i = 1$

Let $\alpha_i = \frac{b_i}{\sum_{j=1}^{n} b_j}$ & $t_i = \frac{a_i}{b_i}$ in ①

$$\sum_{i=1}^{n} \frac{b_i}{\sum_{j=1}^{n} b_j} \times \frac{a_i}{b_i} \log \frac{a_i}{b_i} \geq \sum_{i=1}^{n} \frac{b_i}{\sum_{j=1}^{n} b_j} \frac{a_i}{b_i} \log \sum_{i=1}^{n} \frac{b_i}{\sum_{j=1}^{n} b_j} \frac{a_i}{b_i}$$

$$\frac{1}{\sum_{j=1}^{n} b_j} \sum_{i=1}^{n} a_i \log \frac{a_i}{b_i} \geq \frac{1}{\sum b_j} \sum a_i \log \frac{\sum a_i}{\sum b_i}$$

$$\sum_{i=1}^{n} a_i \log \frac{a_i}{b_i} \geq \sum a_i \log \frac{\sum_{i=1}^{n} a_i}{\sum_{i=1}^{n} b_i}$$

Using log sum inequality show that (27)

$D(p\|q) \geq 0$

Soln: By def^n

$$D(p\|q) = \sum p(x) \log \frac{p(x)}{q(x)}$$

$$\geq \sum p(x) \log \left[\frac{\sum p(x)}{\sum q(x)}\right]$$

$$\geq 1 \log \frac{1}{1} = 0$$

$\therefore D(p\|q) \geq 0$

Let the random variable X has 3 possible outcomes {a, b, c}. Consider two distributions on this random variable

X	a	b	c
p(x)	$\frac{1}{2}$	$\frac{1}{4}$	$\frac{1}{4}$
q(x)	$\frac{1}{3}$	$\frac{1}{3}$	$\frac{1}{3}$

find H(P), H(q), D(p‖q), D(q‖p)

& show that $D(p\|q) \neq D(q\|p)$

Solution:

$$H(p) = -p \log p$$

$$= -\frac{1}{2}\log\frac{1}{2} - \frac{1}{4}\log\frac{1}{4} - \frac{1}{4}\log\frac{1}{4} \Rightarrow$$

$$= -\frac{1}{2}\log\frac{1}{2} - \frac{1}{2}\log\frac{1}{4} = \frac{1}{2}\log 2 + \frac{1}{2}\log 4$$

$$= \frac{1}{2}\log 2 + \frac{2}{2}\log 2 = \frac{1}{2} + 1 = \frac{3}{2}$$

$$H(q) = -\frac{1}{3}\log\frac{1}{3} - \frac{1}{3}\log\frac{1}{3} - \frac{1}{3}\log\frac{1}{3} = -\frac{3}{3}\log\frac{1}{3}$$

$$= \log 3 = \left(\frac{\log 3}{\log 2}\right) = 1.585$$

$$D(p\|q) = \Sigma\, p(x) \log\frac{p(x)}{q(x)}$$

$$= \frac{1}{2}\log\left(\frac{\frac{1}{2}}{\frac{1}{3}}\right) + \frac{1}{4}\log\left(\frac{\frac{1}{4}}{\frac{1}{3}}\right) + \frac{1}{4}\log\left(\frac{\frac{1}{4}}{\frac{1}{3}}\right)$$

$$= \frac{1}{2}\log\left(\frac{3}{2}\right) + \frac{1}{4}\log\left(\frac{3}{4}\right) + \frac{1}{4}\log\left(\frac{3}{4}\right) = 0.085$$

$$D(q\|p) = \Sigma\, q(x) \log\frac{q(x)}{p(x)}$$

$$= \frac{1}{3}\log\left(\frac{\frac{1}{3}}{\frac{1}{2}}\right) + \frac{1}{3}\log\left(\frac{\frac{1}{3}}{\frac{1}{4}}\right) + \frac{1}{3}\log\left(\frac{\frac{1}{3}}{\frac{1}{4}}\right)$$

$$= \frac{1}{3}\log\left(\frac{2}{3}\right) + \frac{1}{3}\log\left(\frac{4}{3}\right) + \frac{1}{3}\log\left(\frac{4}{3}\right)$$

$$= \frac{1}{3}\log\frac{2}{3} + \frac{2}{3}\log\frac{4}{3}$$

$$= \frac{1}{3}\left[\log_2 2 - \log 3\right] + \frac{2}{3}\left[\log 4 - \log 3\right]$$

$$= \frac{1}{3} - \frac{1}{3}\log 3 + \frac{2}{3}\log 4 - \frac{2}{3}\log 3$$

$$= \frac{1}{3} + \frac{2}{3}\log 4 - \log 3$$

$$= \frac{1}{3} + \frac{2}{3}\left[\frac{\log 4}{\log 2}\right] - \frac{\log 3}{\log 2}$$

$$= 0.082$$

Information theory and coding

1. Define entropy.
2. Write any two properties of entropy.
3. Prove that H (x, y) = H(x) + H(y/x) (Chain rule)
4. Define conditional entropy.
5. Define joint entropy.
6. State and prove any two properties of entropy
7. Let $x = \begin{cases} 0 & with\ p = \frac{1}{2} \\ 1 & p = \frac{1}{2} \end{cases}$, then find H(X).
8. Let x denote the value shown up when a die is rolled with equally likely for all six events. Find H(x).
9. Let X= $x = \begin{cases} a & p = \frac{1}{2} \\ b & p = \frac{1}{4} \\ c & p = \frac{1}{8} \\ d & p = \frac{1}{8} \end{cases}$, then find entropy of X.
10. A fair coin is tossed until the first head occurs. Let X denotes the number of tosses required. Find the entropy H(X) in bits.
11. Prove that $H(X,Y) = H(X) + H\left(\frac{Y}{X}\right)$
12. Compute $H(X)$, $H(Y)$, $H\left(\frac{Y}{X}\right)$, $H\left(\frac{X}{Y}\right)$, $H(X,Y)$ for the following

Y/x	1	2	3	4
1	$\frac{1}{8}$	$\frac{1}{16}$	$\frac{1}{32}$	$\frac{1}{32}$
2	$\frac{1}{16}$	$\frac{1}{8}$	$\frac{1}{32}$	$\frac{1}{32}$
3	$\frac{1}{16}$	$\frac{1}{16}$	$\frac{1}{16}$	$\frac{1}{16}$
4	$\frac{1}{4}$	0	0	0

x/y	a	b	c
1	$\frac{1}{6}$	$\frac{1}{12}$	$\frac{1}{12}$
2	$\frac{1}{12}$	$\frac{1}{6}$	$\frac{1}{12}$
3	$\frac{1}{12}$	$\frac{1}{16}$	$\frac{1}{12}$

x/y	0	1
0	$\frac{1}{3}$	$\frac{1}{3}$
1	0	$\frac{1}{3}$

13. Define relative entropy and Mutual information.

14. Prove that $I(X,Y) = H(X) - H\left(\frac{X}{Y}\right)$

15. Let $X = \{0, 1\}$ and consider two distribution p and q on X.
Let p (0) =1-r, p (1) = r and q (0) =1-s, q (1) =s. Find $D(p//q)$ $D(q//p)$

16. Let $X_1, X_2, ... X_n$ be drawn according to $P(X_1, X_2, ... X_n)$, then prove that $H(X_1, X_2, ... X_n) = \sum_{i=1}^{n}\left(\frac{X_i}{X_{i-1}, X_{i-2}, ... X_1}\right)$

17. Define concave and convex function.

18. If the function f has a second derivative that is positive over an interval, then the function is convex over that interval.

19. If f is a convex function and X is a random variable, then prove that $Ef(X) \geq f(E\ X)$. Moreover, if f is strictly convex then $X = E\ X$ with probability 1.

20. Let p(x). q(x) for $x \in X$ be two probability mass functions, then $D(p//q)$=0 with equality if and only if $p(x) = q(x) \quad \forall x$.

21. Prove that $H\left(\frac{X}{Y}\right) \leq H(X)$ with equality iff X and Y are independent.

22. State and prove log sum inequality.

23. Using log sum inequality show that $D(p//q) \geq 0$.

24. Let the random variable X has three possible outcomes {a, b, c}. consider two distributions on this random variable.

x	a	b	c
P(x)	$\frac{1}{2}$	$\frac{1}{4}$	$\frac{1}{4}$
q(x)	$\frac{1}{3}$	$\frac{1}{3}$	$\frac{1}{3}$

Find H(p), H(q), $D(p//q)$, $D(q//p)$ and show that $D(p//q) \neq D(q//p)$.