Aug 9th

Lecture 2:

Recap
- brief (very brief!) course overview
- course logistics
- axiomatic derivation of entropy formula

TODAY
- entropy fn & properties
- joint entropy
- relative entropy and mutual information.

ENTROPY OF A RANDOM VARIABLE

Let X be a random variable alphabet $\mathcal{X}$, $|\mathcal{X}| < \infty$ $P_X(x) = P(x)$ is the probability mass function.

Define

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

Adopt the convention $0 \log 0 = 0$

If logs are to the base b, $H_b(X) = -\sum_{x \in \mathcal{X}} p(x) \log_b p(x)$

(logs to base 2 unless otherwise specified)

X a random variable $f(x)$ a function

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

Note that: $H(X) \triangleq \mathbb{E}_{p(x)}\left[\log \frac{1}{p(X)}\right]$ "eerily self referential" Cover & Thomas

Properties:

1) $H(X) \geq 0$

$$H(X) = \sum p(x) \underbrace{\log \frac{1}{p(x)}}_{\geq 0} \geq 0$$

clear!

2) $H_b(X)$

$$= -\sum_{x \in \mathcal{X}} p(x) \log_b p(x)$$

$$= -\sum_{x \in \mathcal{X}} p(x) \log_a p(x) \log_b a$$

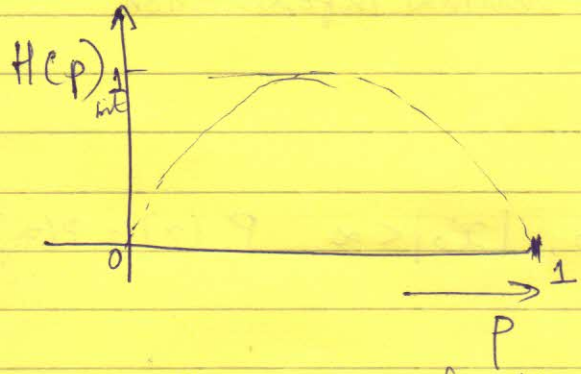
$$= H_a(X) \log_b(a)$$

$a, b > 1$

Let X be a Bernoulli random variable $\quad X = \begin{cases} 1 & \text{prob} = p \\ 0 & 1-p \end{cases}$

Then $H(X) = p \log \frac{1}{p} + (1-p) \log \frac{1}{1-p}$
$\triangleq H(p)$

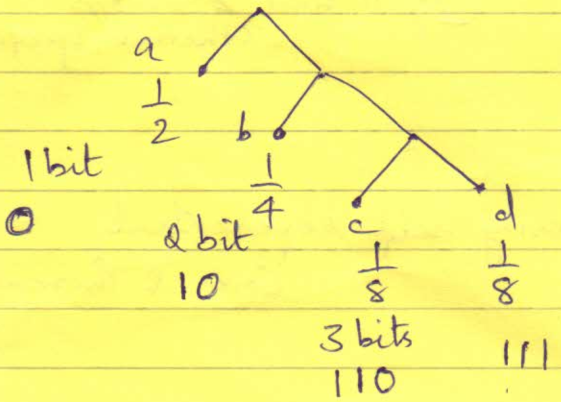
$\mathcal{X} = \{0, 1\}$.



$X = \begin{cases} a & \text{prob} = 1/2 \\ b & 1/4 \\ c & 1/8 \\ d & 1/8 \end{cases}$

$H(X) = \frac{1}{2}\log 2 + \frac{1}{4}\log 4 + \frac{1}{8}\log 8 + \frac{1}{8}\log 8$

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



$\mathcal{X} = \{a, b, c, d\}$

is it a?
is it b?
is it c?

On an average 1.75 questions will be neede[d] to determine the val[ue]

Joint & Conditional entropy:

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

JOINT ENTROPY $= E_{p(x,y)}\left[\log \frac{1}{p(X,Y)}\right]$

Ex 2.1

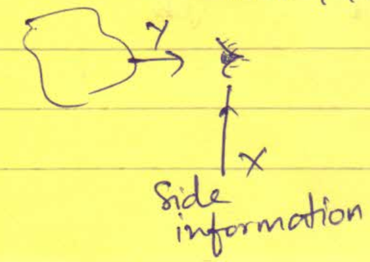
Sample space Y

X \ Y	1	2	3	4
1	1/8	1/16	1/16	1/4
2	1/16	1/8	1/16	0
3	1/32	1/32	1/16	0
4	1/32	1/32	1/16	0

$\mathcal{X} = \mathcal{Y} = \{1, 2, 3, 4\}$.

Conditional Entropy.

$H(Y|X) = \mathbb{E}_{p(x)}\left[H(Y|X=x)\right]$.



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

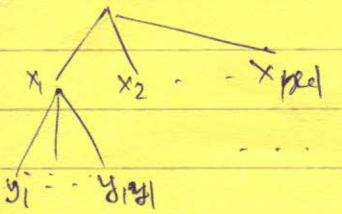
$= \sum_{x,y} p(x,y) \log \frac{1}{p(y|x)}$

Note that $H(X,Y) = \sum_{xy} p(x,y) \log \frac{1}{p(x,y)} = \sum_{xy} p(x,y) \log \frac{1}{p(x)} + \sum_{xy} p(x,y) \log \frac{1}{p(y|x)}$

(also called chain rule) $= H(X) + H(Y|X)$

Does this sound familiar?

this is precisely Grouping Axiom!



Note: $H(Y|X) = \mathbb{E}_{p(x,y)} \left\{ \log \frac{1}{p(Y|X)} \right\}$

$$H(X,Y) = \mathbb{E}_{p(x,y)} \left[\log \frac{1}{p(X,Y)} \right]$$
$$= \mathbb{E}_{p(x,y)} \left[\log \frac{1}{p(X)} + \log \frac{1}{p(Y|X)} \right]$$
$$= \mathbb{E}_{p(x,y)} \left[\log \frac{1}{p(X)} \right] + \mathbb{E}_{p(x,y)} \left\{ \log \frac{1}{p(Y|X)} \right\}$$
$$= H(X) + H(Y/X)$$

Excercise: show that

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

$H(X)$ in Example 2.1

$$P_X(x) = \begin{cases} \frac{1}{2} & x=1 \\ \frac{1}{4} & x=2 \\ \frac{1}{8} & 3 \\ \frac{1}{8} & 4 \end{cases} \qquad P_Y(y) = \begin{cases} \frac{1}{4} & y \in \mathcal{Y} \end{cases}$$

$\Rightarrow H(X) = 1.75$ bits $\qquad \Rightarrow H(Y) = 2$ bits

$H(X|Y)$

$$P_{X|Y} \quad \begin{array}{c|cccc} & & Y & & \\ \hline X & 1/2 & 1/4 & 1/4 & 1 \\ & 1/4 & 1/2 & 1/4 & 0 \\ & 1/8 & 1/8 & 1/4 & 0 \\ & 1/8 & 1/8 & 1/4 & 0 \end{array}$$

Can show
$H(X,Y) = 3.375$ bits
$H(Y|X) = \frac{13}{8}$ bits

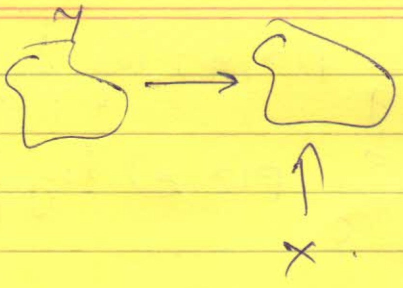
3.5+2

$H(X|Y=1) = \frac{1}{2}\log 2 + 2 \times \frac{1}{8} \log 8 + \frac{1}{4}\log 4 = 1.75$

$H(X|Y=2) = 1.75$

$H(X|Y=3) = 2, \quad H(X|Y=4) = 0.$

$\Rightarrow H(X|Y) = \frac{1.75 + 1.75 + 2}{4} = \frac{5.5}{4} = 1.375$



$H(X) - H(X|Y)$ = residual information about X after observing Y.

$\frac{7}{4} - \frac{11}{8} = 3/8$

$H(Y) - H(Y|X)$ = residual information about Y after observing X

$2 - \frac{13}{8} = \frac{3}{8}$.

$$\xrightarrow{X} \boxed{\text{channel}} \xrightarrow{Y} \qquad I(X;Y).$$

Relative entropy and mutual information

let X have alphabet $\mathcal{X}$. let $p(x)$, $q(x)$ be two distinct distributions on X. then the relative entropy w.r.t the two distributions is given by:

$$D(p\|q) = \sum_{x \in \mathcal{X}} p(x) \log \frac{p(x)}{q(x)}$$

also called the Kullback-Leibler distance between $p(x)$ and $q(x)$

though it is not symmetric it is considered as a distance measure

Adopt the convention that

$$0 \log \frac{0}{0} = 0, \quad 0 \log \frac{0}{q(x)} = 0, \quad p(x) \log \frac{p(x)}{0} = \infty$$

{in computing $p(x) \log \frac{p(x)}{q(x)}$}

Turns out if one tries to compress X using the incorrect distribution $q(x)$, then the average string length $= H(X) + D(p\|q)$ as opposite to $H(X)$!!

Mutual information:

The mutual information $I(X;Y)$ between 2 random variables having joint distribution $p(x,y)$ is defined by

$$I(X;Y) = D(p(x,y) \,\|\, p(x)p(y))$$

$$\therefore I(X;Y) = \sum_{x,y} p(x,y) \log \frac{p(x,y)}{p(x)p(y)} = E_{p(x,y)}\left[\log \frac{p(X,Y)}{p(X)p(Y)}\right]$$

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

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

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