Let dist. of X be $\{p_1, p_2, \dots p_n\}$

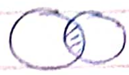
$$\Rightarrow H(X) = -\sum_{i=1}^{n} p_i \log(p_i)$$

$$X = \left\{ \begin{array}{l} a \longrightarrow 1/2 \\ b \longrightarrow 1/4 \\ c \longrightarrow 1/8 \\ d \longrightarrow 1/8 \end{array} \right\} \Rightarrow H(X) = \frac{1}{2}\log 2 + \frac{1}{4}\log 4 + \frac{1}{8}\times 2 \log 8$$

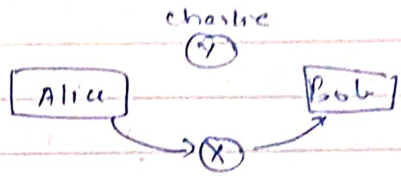
$$= \frac{1}{2} + \frac{1}{2} + \frac{3}{4} = 1 + \frac{3}{4} = \frac{7}{4} \text{ bits}$$

P.T. $H(X|Y) \leq H(X)$

$P(X|Y) = \frac{P(X,Y)}{P(Y)}$ and $P(X)$ $\qquad P(X,Y) \leq P(X)$.



Conditional Entropy



X: R.V. that A will visit B

Y: R.V. that A's car got punctured.

A and B are now dependent

Let $q_{kl} = P(B_l | A_k)$, and $\pi_{kl} = P(A_k, B_l)$

$\Rightarrow$

$$\pi_{kl} = p_k \cdot q_{kl}$$

$$-H(AB) = \sum_k \sum_l \pi_{kl} \log(\pi_{kl})$$

$$= \sum_k \sum_l p_k q_{kl} \{\log(p_k) + \log(q_{kl})\}$$

$$= \underbrace{\sum_k p_k \log(p_k)}_{-H(A)} \cdot \underbrace{\sum_l q_{kl}}_{1} + \sum_k p_k \sum_l q_{kl} \log(q_{kl}).$$

$$-H(AB) = -H(A) + \sum_k p_k \sum_l q_{kl} \log(q_{kl}).$$

$\Rightarrow$

$$H(AB) = H(A) + \sum_k p_k \left(-\sum_l q_{kl} \log(q_{kl})\right).$$

$$H(AB) = H(A) + \sum_k p_k (H_k(B))$$

$$H(AB) = H(A) + H(B|A). \quad \text{——— (1)}.$$

P.T. $H(B|A) \leq H(B)$.

$$H(BA) = H(B) + H(A|B). \qquad -\cancel{\sum_k p_k \log p_k + \sum_k \sum_l p_k q_k}$$

$$H(B|A) = H(B) + H(A|B) - H(A).$$

$$= H(B) - (H(A) - H(A|B))$$

$H(B) - H(B|A)$

$= -\sum_{l} q_l \log q_l + \sum_{k}\sum_{l} p_k q_{kl} \log (q_{kl})$

for concave ~~convex~~ funct. (cont.),

$$\sum_k \lambda_k f(x_k) \geq f\left(\sum \lambda_k x_k\right). \text{ if } \lambda_k \geq 0 \text{ and } \sum \lambda_k = 1$$

$\begin{cases} f(x) = x\log x \\ \lambda_k = p_k, \; x_k = q_{kl} \end{cases}$

$\Rightarrow$

$$\sum_k p_k q_{kl} \log(q_{kl}) \geq \sum_k \underbrace{p_k q_{kl}}_{q_l} \log\left(\underbrace{\sum_k p_k q_{kl}}_{q_l}\right)$$

$$\Rightarrow \sum_k p_k q_{kl} \log(q_{kl}) \geq q_l \log(q_l)$$

$$\Rightarrow -\sum_k p_k H_k(B) \geq -H(B)$$

$$\Rightarrow -H(B|A) \geq -H(B)$$

$$\boxed{H(B) \geq H(B|A)}$$

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

$\frac{3}{16} + \frac{4}{16}$

find $H(X)$, $H(X|Y)$, $H(Y|X)$
$H(XY)$

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

$H(X) = -\sum_x p_x \log(p_x)$

$= \frac{1}{2}.1 + \frac{1}{4}.2 + \frac{1}{8} \times 2 \times 3$

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

$H(XY) = \sum_x\sum_y p_{xy} \log(p_{xy})$.

$= H(Y|X) = \sum_x p_x \left(-\sum_y p_{y|x} \log(p_{y|x})\right)$.

~~$H_1(Y) = \frac{1}{8} \times 3 + \frac{1}{16} \times 2 \times 4 \times 2 + \frac{1}{4} \times \frac{2}{2} = 1 + \frac{3}{8} = \frac{11}{8}$~~

$H_2(Y) =$

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

$: H(X'Y) - H(Y) : \frac{27}{8} - \frac{2 \times 8}{8} = \frac{11}{8}$

$$H(Y|X) = \sum_x \sum_y P(x,y) \log\left(\frac{P(x)}{P(x,y)}\right).$$

$$= H(Y)_1 = \frac{1}{8}\log\left(\frac{\cancel{8}^{4}}{\cancel{2}}\right) + \frac{1}{16}\log\left(\frac{\cancel{16}^{8}}{\cancel{2}}\right) + \frac{1}{16}\log\left(\frac{\cancel{16}^{8}}{\cancel{2}}\right) + \frac{1}{4}\log\left(\frac{\cancel{4}^{2}}{\cancel{2}}\right)$$

$$= \frac{\cancel{2}^{1}}{\cancel{8}_{4}} + \frac{3}{16} + \frac{3}{16} + \frac{1}{4} = \frac{1}{2} + \frac{3}{8} = \frac{7}{8}$$

$$H(Y)_2 = \frac{1}{\cancel{16}}\log\left(\frac{\cancel{16}}{\cancel{4}}\right) \times \cancel{2} + \frac{1}{8}\log\left(\frac{\cancel{8}}{\cancel{4}}\right) = \frac{2}{4} \cdot \frac{1}{8} = \frac{3}{8}.$$

$$H(Y)_3 = \frac{1}{\cancel{32}} \times \frac{\cancel{2}}{\cancel{16}_{8}} \times \log\left(\frac{\cancel{32}^{4}}{\cancel{8}}\right) + \frac{1}{16}\log\left(\frac{\cancel{16}^{2}}{\cancel{8}}\right) = \frac{1}{8} + \frac{1}{16} = \frac{3}{16}$$

$$H(Y)_4 = \frac{1}{\cancel{32}} \times \frac{\cancel{2}}{\cancel{16}_{8}} \times \log\left(\frac{\cancel{32}^{4}}{\cancel{8}}\right) + \frac{1}{16}\log\left(\frac{\cancel{16}^{2}}{\cancel{8}}\right) = \frac{1}{16} + \frac{1}{8} = \frac{5}{16}$$

$$P(Y|X) = \frac{7}{8} \times \frac{1}{2} + \frac{3}{8} \times \frac{1}{4} + \frac{1}{8} \times \frac{3}{16} + \frac{1}{8} \times \frac{3}{16}$$

$$= \frac{7}{16} + \frac{3}{32} + \frac{2 \times 3}{4\,\cancel{8} \times 16} = \frac{7}{16} + \frac{3}{32} + \frac{3}{64}$$

$$= \frac{28 + 6 + 3}{64} =$$

$$H(X|Y) = \sum_x P_x\, H(X|Y=y)$$

$$H(X|Y=1) = H\left(\frac{1}{2}\ \frac{1}{4}\ \frac{1}{8}\ \frac{1}{8}\right) = \qquad H(X|Y=2) = \frac{1}{4}\ \frac{1}{2}\ \frac{1}{8}\ \frac{1}{8})$$

$$H(X|Y=3) = H\left(\frac{1}{4}\ \frac{1}{4}\ \frac{1}{4}\ \frac{1}{4}\right) \qquad H(X|Y=4) = (1, 0, 0, 0)$$

$$\frac{1}{2} + \frac{2}{4} + \frac{3}{\cancel{8}4} = \frac{7}{4}, \qquad \cancel{\frac{2}{4}} \cancel{\frac{1}{2}} + \cancel{\frac{1\,2\,2+3}{84}} \quad \frac{7}{4}$$

$$2 \qquad\qquad 0$$

$$H(X|Y) = \frac{1}{4} \cdot \frac{7}{4} + \frac{1}{4} \cdot \frac{7}{4} + \frac{1}{4} \cdot \frac{2 \cdot 4}{4} = \frac{8 + 7 + 7}{16} = \frac{22}{16} = \frac{11}{8}$$

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

$$= 2 + \frac{11}{8} = \frac{7}{4} + H(Y|X)$$

$$\begin{array}{r} 27 \\ -14 \\ \hline 13 \end{array}$$

$$H(Y|X) = \frac{16 + 11}{8} - \frac{7}{4} = \frac{27 - 14}{8} = \frac{13}{8}.$$

$H(Y|X) = \sum_x p_x \, H(Y|X=x).$

$H_1(Y) = H\left(\frac{1}{4}, \frac{1}{8}, \frac{1}{8}, \frac{1}{2}\right) = \frac{2}{4} + \frac{3}{4} + \frac{1}{2} = \frac{7}{4}$

$H_2(Y) = H\left(\frac{1}{4}, \frac{1}{2}, \frac{1}{4}, 0\right) = \frac{1}{2} + 2 \times \frac{2}{4} = \frac{3}{2}$

$H_3(Y) = H\left(\frac{1}{4}, \frac{1}{4}, \frac{1}{2}, 0\right) = \frac{3}{2}$

$H_4(Y) = H\left(\frac{1}{4}, \frac{1}{4}, \frac{1}{2}, 0\right) = \frac{3}{2}$

$H(Y|X) = \frac{1}{2} \cdot \frac{7}{4} + \frac{1}{4} \cdot \frac{3}{2} + \frac{1}{8} \times \cancel{2} \times \frac{3}{\cancel{2}}$

$= \frac{7}{8} + \frac{3}{8} + \frac{3}{8} = \frac{13}{8}$

$\frac{16}{8} \quad \frac{13}{8}$

Remember $\boxed{H(XY) = H(X) + H(Y|X) = H(Y) + H(X|Y)}$

$H(X,Y)$

* The joint entropy $H(XY)$ corresponding to a pair of discrete random variables (X,Y) with prob. dist. $p(x,y)$ is given by

$$H(X,Y) = -\sum_x \sum_y p(x,y) \cdot \log(p(x,y)) = -\mathbb{E}_{X,Y}\left[\log(p(x,y))\right]$$

and $H(XY) = H(X) + H(Y)$ when X, Y are mut. ind.

* Conditional Entropy : If there are 2 random variables $X, Y \sim p(x,y)$ then $H(Y|X)$ is given by

$H(X|Y)$

$$H(Y|X) = \sum_x p(x) \cdot H(Y|X=x) = -\sum_x p(x) \sum_y p(y|x) \log p(y|x)$$

$$= -\sum_x \sum_y p(x) \cdot p(y|x) \log(p(y|x))$$

$$= -\sum_x \sum_y p(x,y) \cdot \log(p(y|x))$$

$$= -\mathbb{E}_{x,y} \log(p(y|x))$$

P.T $H(X,Y|Z) = H(X|Z) + H(Y|X,Z)$

$H(X,Y|Z) = \sum_z p(z).\ H(X,Y|Z=z)$

$= -\sum_x \sum_y \sum_z p(z).P(x,y|z).\log(P(x,y|z))$

$= -\sum_x \sum_y \sum_z \cancel{p(z)}.P(x,y,z)/\cancel{P(z)}.\{\log P(x,y,z) + \log P(z)\}$

$= -\sum_x \sum_y \sum_z \cancel{P(z)}\ P(x,y,z)\ \cancel{P(z)}\ \log P(x,y,z)$

$- \sum_x \sum_y \sum_z \cancel{P(z)}\ P(x,y,z)\ \cancel{P(z)}\ \log(P(z))$.

$= \cancel{\sum_z p(z).\ P(x,y.}$

RHS $= -\sum_x \sum_z P(x,z)\ \cancel{P(x)}\ \log(P(x|z))$

$+ -\sum_x \sum_y \sum_z p(x,y,z).\log\left(\frac{y,x,z}{P(x,z)}\right)$.

Soln. $H(X,Y|Z) = H(X,Y,Z) - H(Z)$

$H(X|Z) = H(X,Z) - H(Z)$

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

$= H(X,Y,Z) - \{H(X,Z)\}$

$\Rightarrow H(X,Y|Z) = H(X,Z) + H(Y|X,Z)$.

(K.L. Divergence)

▶ Relative Entropy : It is the measure of distance between 2 probability distributions. It is also the measure of inefficiency of assuming one dist. while other dist. is ~~given~~ true.

$D(p\|q)$.

$D(p\|q) \rightarrow$ 'q' is assumed while 'p' is the true distribution

$$D(p\|q) = \sum_{x \in X} p(x).\log\frac{p(x)}{q(x)} = \mathbb{E}_{p(x)}.\log\left\{\frac{p(x)}{q(x)}\right\}.$$

1) when $p(x)=0$ $D(p\|q)=0$

2) when $q(x)=0$ $D(p\|q) \rightarrow \infty$

↳ can $D(p\|q) = \infty$ even if $q(x) != 0$?

$p \cdot \{0,1\} = \{1-r, r\}$
$q \{0,1\} = \{1-s, s\}$

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

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

$$\left.\right\} D(p\|q) \neq D(q\|p)$$

* Relative Entropy is a non-negative quantity

▶ Information Inequality ($D(p\|q) \geq 0$)

We have 2 probability distributions $p(x)$ & $q(x)$ (pmf).

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

with equality iff $p(x) = q(x) \;\forall x \in X$

Let $A = \{x : p(x) > 0\}$ be the support set of $p(x)$

Then

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

for concave ~~convex~~ function $\log(x)$, $f'(x) = 1/x \leftarrow$ dec

$E[f(x)) \leq f(E[\cancel{f}(x)]] \longrightarrow (1)$

using (1).

$$-D = \sum_{x \in A} p(x) \log\left(\frac{q(x)}{p(x)}\right) \leq \log\left(\sum_{x \in A} p(x) \cdot \frac{q(x)}{p(x)}\right) \leq \log\left(\sum_{x \in X} q(x)\right) \leq 0$$

$\Rightarrow \boxed{D \geq 0}$

P.T. $\ln x \leq x - 1$

$e^x \geq 1 + x \Rightarrow$ Let $x = \ln x$

$\Rightarrow e^{\ln x} \geq 1 + \ln x \Rightarrow \boxed{\ln(x) \leq x - 1} \longrightarrow (11)$

We can prove $D \geq 0$ using (11) also

Mutual Information ($I(X:Y)$)

Consider X and Y to be 2 r.v. with a joint prob. mass function $p(x,y)$. Let $p(x)$ & $p(y)$ be marginal prob dist. (pmf).
The mutual information $I(X;Y)$ is given as the relative entropy between the joint prob dist. $p(x,y)$ & $p(x).p(y)$.

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

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

$$= \sum_x \sum_y p(x,y)\log\left(\frac{p(x|y)}{p(x)}\right).$$

$$= -\sum_x \sum_y p(x,y).\log p(x) - \left(-\sum_x \sum_y p(x,y).\log(p(x|y))\right)$$

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

$\Rightarrow$ (*) $\boxed{I(X:Y) = H(X) - H(X|Y)}$

similarly $I(X:Y) = H(Y) - H(Y|X)$.

$\Rightarrow$. $H(X,Y) = H(X) + H(Y|X)$

$I(X:Y) = H(\cancel{X}\,Y) - (H(X,Y) - H(X))$

(*) $\boxed{I(X:Y) = H(X) + H(Y) - H(X,Y)}$ } sth. like Inclusion exclusion principle.

(*) $\boxed{I(X:X) = H(X) - H(X|X) = H(X) - 0 = H(X)}$

$\Rightarrow$ we can call entropy ($H(X)$) as self information $I(X:X)$

Theorem: Chain Rule of Entropy:

We have 'n' r.v. $X_1, X_2, X_3 \ldots X_n$ drawn acc. to prob. dist. $p(x_1, x_2, \ldots x_n)$. Then.

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

Proof

$$\left\{ H(X_1, X_2 \ldots X_n) = \sum_{i=1}^{n} H(X_i \mid X_{i-1} \ldots X_1) \right\}$$

$$H(X_1, X_2) = H(X_2 \mid X_1) + H(X_1)$$ → Base case

Let $H(X_1 \ldots X_k) = \sum_{i=1}^{k} H(X_i \mid X_{i-1} \ldots X_1)$.

$$\begin{aligned} H(X_1 \ldots X_k X_{k+1}) &= H(X_{k+1} \mid X_k, X_{k-1} \ldots X_1) + H(X_k, X_{k-1} \ldots X_1) \\ &= H(X_{k+1} \mid X_k, \ldots X_1) + \sum_{i=1}^{k} H(X_i \mid X_{i-1} \ldots X_1) \\ &= \sum_{i=1}^{k+1} H(X_i \mid X_{i-1} \ldots X_1) \end{aligned}$$ → Ind. case

Hence Proved

- **Conditional Mutual Information** $I(X;Y \mid Z)$

Let X, Y, Z be 3 random variables, then conditional mutual info. b/w X & Y given Z is given as:-

$$\begin{aligned} I(X;Y \mid Z) &= H(X \mid Z) - H(X \mid Y, Z). \\ &= \mathbb{E}_{p(x,y,z)}\left[\log \frac{p(x,y \mid z)}{p(x \mid z) \cdot p(y \mid z)}\right] \end{aligned}$$

- **Chain Rule for Mutual Info.**

$$I(X_1, X_2, \ldots X_n; Y) = \sum_{i=1}^{n} I(X_i; Y \mid X_1, \ldots X_{i-1})$$

$$\begin{aligned} I(X_1, X_2 \ldots X_n; Y) &= H(X_1, X_2 \ldots X_n) - H(X_1, \ldots X_n \mid Y) \\ &= \sum_{i=1}^{n} H(X_i \mid X_{i-1} \ldots X_1) - \sum_{i=1}^{n} H(X_i \mid X_1 \ldots X_n, Y) \\ &= \sum_{i=1}^{n} H(X_i ; Y \mid X_1 \ldots X_1). \end{aligned}$$

▶ Conditional Relative Entropy : between 2 condⁿ prob dist $p(y|x)$ & $q(y|x)$ is given by

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

$$= \sum_{p(x,y)} p(x,y) \cdot \log\left[\frac{p(y|x)}{q(y|x)}\right] = \underset{x,y}{E}\left[\log \frac{p(y|x)}{q(y|x)}\right]$$

Show: $\boxed{D(p(x,y) \| q(x,y))}$

$$= D(p(x) \| q(x)) + D(p(y|x) \| q(y|x)).$$

$$\text{LHS} = \sum_{x,y} p(x,y) \cdot \log \frac{p(x,y)}{q(x,y)}$$

$\sum_y p(y) \log$

$$= \sum_{x,y} p(\cancel{y|x}\, x,y) \log\left(\frac{p(y|x) \cdot p(x)}{q(y|x)\, q(x)}\right)$$

$$= \sum_{x,y} p(x,y) \cdot \cancel{p(x)} \log\left(\frac{p(x)}{q(x)}\right) + \sum_{x,y} p(x,y) \log\left(\frac{p(y|x)}{q(y|x)}\right)$$

$$= D(p(x) \| q(x)) + D(p(y|x) \| q(y|x)).$$

Eg

Y \ X	1	2	
1	0	3/4	$\frac{3}{4}$
2	1/8	1/8	$\frac{1}{4}$
	$\frac{1}{8}$	$\frac{7}{8}$	

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

$H(X) = \frac{1}{8} \times 3 + \frac{7}{8} \log\left(\frac{8}{7}\right) = 0.544$ bits

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

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

$H(X,Y) = \frac{3}{4} \log \frac{4}{3} + 2 \times \frac{1}{8} \times 3 = \frac{3}{4}\left(1 + \log \frac{4}{3}\right)$

$H(X|Y=1) = 0$

$H(X|Y=2) = 1$

$H(X|Y) = \frac{3}{4} + \cancel{\frac{3}{4} \log \frac{4}{3}} - \cancel{\frac{3}{4} \log \frac{4}{3}} - \frac{2}{4} \times 2$

$= \frac{1}{4}$.

Note that $H(X|Y=1) > H(X)$

but $H(X|Y)$ will always be $\leq H(X)$.

① for any 2 R.V. X, Y $I(X;Y) \geq 0$
(equality holds when X & Y are independent).

② ~~D(p(y||x)~~ $D(p(y|x) \,\|\, q(y|x)) \geq 0$
(equality when $p(y|x) = q(y|x)$ with $p(x) > 0$).

③. $I(X;Y|Z) \geq 0$ with equality iff X & Y are cond. independent given Z

Create examples for equalities in ①, ② & ③.

Q.) I am tossing a coin twice. Let X be the number of heads that occur. Then I toss it twice again, Y be no. of heads. now, this time

i) find $P(X<2, Y>1)$.

ii) find $I(X:Y)$.

H $\frac{1}{8}\times 3$

X \ Y	0	1	2
0	1/4	1/2	1/4
1	1/2		
2	1/4		

i) $P(X<2) * P(Y>1)$
$= \left(1-\frac{1}{4}\right)\cdot\frac{1}{4} = \frac{3}{16}$

x:	0	1	2
	1/4	1/2	1/4
y:	1/4	1/2	1/4

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

<u>Show that:</u>
$I(X:Y,Z) = I(X:Z) + I(X:Y|Z)$ —— ⓘ
$I(X:Y,Z) = I(X:Y) + I(X:Z|Y)$. —— ⓘⓘ

①. $I(X:Y,Z) = \sum p(x,y,z)\cdot\log\frac{p(x,y,z)}{p(x)\cdot p(y,z)}$

$I(X:Z) = \sum p(x,z)\cdot\log\frac{p(x,z)}{p(x)\cdot p(z)}$

$I(X:Y|Z) = \sum p(x,y,z)\cdot\log\frac{p(x,y|z)}{p(x|z)\cdot p(y|z)}$

(i)

$I(X:Z) = H(X) - H(X|Z)$

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

adding

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

$= I(X:Y,Z) \longrightarrow$ Thus proved.

Similarly prove (ii)

(ii)

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

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

$\Sigma \quad = H(X) - H(X|Y,Z) = I(X:Y,Z)$ Hence Proved

Let there are 3 r.v. X, Y, Z. They will form a markov chain in the order $X \rightarrow Y \rightarrow Z$ if the conditional distribution of Z depends only upon Y and ~~the~~ is conditionally independent on X.

$$\boxed{p(x,y,z) = p(x) \cdot p(y|x) \cdot p(z|y).}$$

$$\Rightarrow p(x,z|y) = \frac{p(x,y,z)}{p(y)} = \frac{p(x) \cdot p(y|x) \cdot p(z|y)}{p(y)}$$

$$= \frac{p(x,y)}{p(y)} \cdot p(z|y) = p(x|y) \cdot p(z|y).$$

Data Processing Inequality: If we have a markov chain $X \rightarrow Y \rightarrow Z$ then $I(X:Y) \geq I(X:Z)$.

Proof.

$I(X:Y) = I(X:Y,Z) - I(X:Z|Y) \quad$ — (I)

$I(X:Z) = I(X:Y,Z) - I(X:Y|Z). \quad$ — (II)

Note Z|Y and X are ind. (since Markov chain).

$\Rightarrow I(X:Z|Y) = 0$ & $I(X:Y|Z)$ exists ≥ 0

$\Rightarrow I(X:Y) \geq I(X:Z)$

HW.

(1) Let $X = x_1 x_2$ $\quad$ x_i & y_j are random bits
$Y = y_1 y_2$ $\quad$ such that

$$x_1 \oplus x_2 = y_1 \oplus y_2 = z.$$

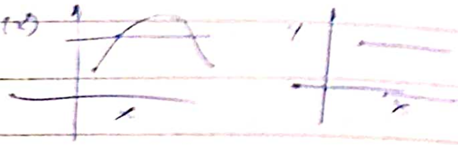
$0 = 0 \qquad\qquad 1 = 1$

(1,1)	(0,0)	(1,0)	(1,0)
(0,0)	(1,1)	(0,1)	(0,1)
(1,1)	(1,1)	(1,0)	(0,1)
(0,0)	(0,0)	(0,1)	(1,0)
X	Y	X	Y

Find $I(X:Y|Z)$

$I(X:Z|Y)$

$I(Y:Z|X)$



$\cancel{H(Y) \le H(X)}$

HW

(2) Let X & Y be 2 r.v. $\quad Y = g(X)$ $\quad$ Which of $H(X)$ or $H(Y)$ is greater

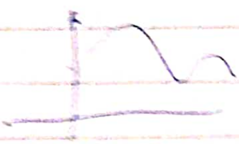
$$H(X) = \sum_x p(x) \log \frac{1}{p(x)}$$

Let $g(x_1) = g(x_2) \ldots g(x_k) = y$

$$H(Y) = \sum_y \left(\sum_{x : y = g(x)} p(x) \log \frac{1}{\sum_{x : y = g(x)} p(x)} \right).$$

$$\underbrace{\sum_y \sum_{x : y = g(x)} p(x)}$$

$$\hookrightarrow \sum_x p(x) . \log \left(\frac{1}{\sum_{x : y = g(x)} p(x)} \right)$$



$$H(Y) = \sum_y \left(\sum_{x : y = g(x)} p(x) . \log \left[\frac{1}{\sum_{x : y = g(x)} p(x)} \right] \right) \le H(X) \; ?$$

Prove

Jensen's Inequality

A function $f(x)$ will be called a convex function over an interval (a,b) if for $x_1, x_2 \in (a,b)$ and $0 \le \lambda < 1$

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

Then $-f$ is concave.

Thm: If the functions 2nd derivative is ≥ 0 everywhere, then we call function to be convex.

Proof: $f(x) = f(x_0) + f'(x_0)(x-x_0) + \frac{f''(x^*)}{2}(x-x_0)^2 \rightarrow$ Taylor exp. around x_0

$x_0 < x^* < x$

$\Rightarrow f(x) - f(x_0) - f'(x_0)\cdot(x-x_0)$.

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

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

take $x = x_1$

$x_0 = \lambda x_1 + (1-\lambda)x_2$

$x_1 - \lambda x_1 - x_2 + \lambda x_2$

$= (x_1 - x_2) - \lambda(x_1 - x_2) = (1-\lambda)(x_1 - x_2)$.

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

2nd case:

$\left.\begin{array}{l} x = x_2 \\ x_0 = \lambda x_1 + (1-\lambda)x_2 \end{array}\right\} \rightarrow \quad f(x_2) \geq f(x_0) + f'(x_0)[\lambda(x_2 - x_1)]$ —— (II).

multiplying (I) by λ & (II) by $(1-\lambda)$ & then adding

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

$\Rightarrow \lambda f(x_1) + (1-\lambda) f(x_2) \geq \cancel{\lambda f}\ f(x_0)$.

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

$\Rightarrow \boxed{\lambda f(x_1) + (1-\lambda) f(x_2) \geq f(\lambda x_1 + (1-\lambda)x_2)}$

$\rightarrow f$ is convex.

(7)

1) we have a convex function f and X is a random variable, then

$$\boxed{E[f(X)] \geq f(E[X])} \rightarrow \text{Jensens Inequality.}$$

Proof: for 2 mass point distribution, it directly follows from being convex.

$$[\, p_1 f(x_1) + p_2 f(x_2) \geq f(p_1 x_1 + p_2 x_2)\,]$$

Let this be true for k-1 mass point & let $p_i' = \frac{p_i}{1-p_k}$

$$\Rightarrow \sum_{i=1}^{k} p_i f(x_i) = p_k f(x_k) + (1-p_k)\left\{ \sum_{i=1}^{k-1} p_i' f(x_i) \right\}$$

$$\geq p_k f(x_k) + (1-p_k) f\left(\sum_{i=1}^{k-1} p_i' x_i \right)$$

$$\geq f\left(p_k x_k + (1-p_k) \sum_{i=1}^{k-1} p_i' x_i \right)$$

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

Now since

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

$$\Rightarrow \boxed{E[f(X)] \geq f(E[X])}$$

Note

$$\sum p_i' = \frac{p_1 + p_2 \cdots p_{k-1}}{1-p_k} = \frac{1-p_k}{1-p_k} = 1$$

since $\boxed{\sum_{i} p_i = 1}$

Q. Show that $H(X) \leq \log|\mathcal{X}|$

where $|\mathcal{X}|$ denotes no. of elements in the range of X with equality iff X has uniform distribution.

Let X have uniform dist:

$$H(X) = -\left(\frac{1}{n}\log\left(\frac{1}{n}\right) + \frac{1}{n}\log\left(\frac{1}{n}\right) + \cdots \frac{1}{n}\log\left(\frac{1}{n}\right)\right)$$

$p_1 \quad p_1 \quad p_2 \quad p_2$

Let $\left.\begin{matrix} p_1 \rightarrow p_1 + \epsilon \\ p_2 \rightarrow p_2 - \epsilon \end{matrix}\right\} \Rightarrow -H'(X) = -\left(\left(\frac{1}{n}+\epsilon\right)\log\left(\frac{1}{n}+\epsilon\right) + \left(\frac{1}{n}-\epsilon\right)\log\left(\frac{1}{n}-\epsilon\right) + \cdots\right)$

$\sum \quad p(x) \log n\, p(x)$

$\leq \quad p(x) \log n \cdot p(x) \log p(x)$

$\log n - H(X)$

Solⁿ Let $u(x)$ be uniform dist $u(x) = \frac{1}{n} \; \forall \; (x \in \mathcal{X})$ $(n = |\mathcal{X}|)$

$p(x)$ be any prob dist of $\mathcal{X}$.

$$\Rightarrow D(p \| u) = \sum_{x \in \mathcal{X}} p(x) \cdot \log \frac{(p(x))}{(u(x))} = \sum p(x) \cdot \log (n \cdot p(x))$$

$$= \sum p(x) \log n + \sum p(x) \log (p(x))$$

$$D(p \| u) = \log n - H(x) \geq 0$$

$$\Rightarrow \boxed{H(x) \leq \log n}$$