Information Theory Assignment 2

Q.1 Say $\vec{p_x} = (p, 1-p)$

$$\Rightarrow \vec{p_y} = \vec{p_x} Q_\Gamma$$

$$= (p, 1-p) \begin{pmatrix} f & 1-f \\ f & 1-f \end{pmatrix}$$

$$= (pf + (1-p)f, \; p(1-f) + (1-p)(1-f))$$

$$= (f, 1-f) \Rightarrow H(Y) = H_2(f)$$

$$\Rightarrow H(Y|X) = pH(Y|X=0) + (1-p)H(Y|X=1)$$

$$= p\,H(f, 1-f) + (1-p)H(f, 1-f)$$

$$= H_2(f)$$

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

$$= H_2(f) - H_2(f) = \underline{\underline{0}}$$

$$\Rightarrow C(\Gamma) = \max_{\vec{p_x}} I(X;Y) = 0$$

Q.2 We need to check first if $\vec{q} = \vec{p}T$ is a probability vector

Indeed, $\sum_{i=1}^{n} q_i$

$$= \sum_{i=1}^{n} \left(\sum_{j=1}^{n} p_j T_{ji} \right)$$

$$= \sum_{i=1}^{n} \sum_{j=1}^{n} p_j T_{ji}$$

$$= \sum_{j=1}^{n} \sum_{i=1}^{n} p_j T_{ji} \quad \text{(finite sum)}$$

$$= \sum_{j=1}^{n} p_j \underbrace{\sum_{i=1}^{n} T_{ji}}_{=1}$$

$$= \sum_{j=1}^{n} p_j = 1$$

and of course, all entries of $\vec{q}$ are non-negative.

Now,

$$H(\vec{q}) - H(\vec{p})$$

$$= -\sum_{j=1}^{n} q_j \log_2 q_j + \sum_{i=1}^{n} p_i \log_2 p_i$$

$$\cancel{= \sum_{i=1}^{n} p_i \log_2 p_i - \sum_{i=1}^{n} \left(\sum_{j=1}^{n} p_j T_{ji}\right) \log_2\left(\sum_{j=1}^{n} p_j T_{ji}\right)}$$

$$\cancel{= \sum_{i=1}^{n} p_i \left(\sum_{j=1}^{n} T_{ji}\right) \log_2 p_i - \sum_{i=1}^{n} \sum_{j=1}^{n} p_j}$$

$$= \sum_{i=1}^{n} p_i \log_2 p_i - \sum_{j=1}^{n}\left(\sum_{i=1}^{n} p_i T_{ij}\right)\log_2\left(\sum_{l=1}^{n} p_l T_{lj}\right)$$

$$= \sum_{i=1}^{n} p_i \left(\sum_{j=1}^{n} T_{ij}\right) \log_2 p_i$$

$$- \sum_{i=1}^{n}\sum_{j=1}^{n} p_i T_{ij} \log_2\left(\sum_{l=1}^{n} p_l T_{lj}\right)$$

$$= \sum_{i=1}^{n}\sum_{j=1}^{n} p_i T_{ij}\left(\log_2 p_i - \log_2\left(\sum_{l=1}^{n} p_l T_{lj}\right)\right)$$

Note that if $\sum_{l=1}^{n} p_l T_{lj} = 0$,

$$p_l T_{lj} = 0 \quad \forall \ 1 \leq l \leq n.$$

$$\Rightarrow \quad p_i T_{ij} = 0$$

So in all cases, the sum is well defined. WLOG, take $p_i > 0$, $T_{ij} > 0$.

$$\Rightarrow \quad = \sum_{i=1}^{n}\sum_{j=1}^{n} p_i T_{ij} \log_2\left(\frac{p_i}{\sum_{l=1}^{n} p_l T_{lj}}\right)$$

$$= \sum_{i=1}^{n}\sum_{j=1}^{n} p_i T_{ij} \log_2\left(\frac{p_i T_{ij}}{T_{ij} \sum_{l=1}^{n} p_l T_{lj}}\right)$$

See that, over $\{1, 2, \ldots, n\}^2$, the sets $\{p_i T_{ij}\}_{i,j \in \{1,2,\ldots,n\}}$

& $\left\{\left(\sum_{l=1}^{n} p_l T_{lj}\right) T_{ij}\right\}_{i,j \in \{1,2,\ldots,n\}}$

form probability distributions

since

$$\sum_{i=1}^{n}\sum_{j=1}^{n} p_i T_{ij} = \sum_{i=1}^{n} p_i \sum_{j=1}^{n} T_{ij}$$
$$= \sum_{i=1}^{n} p_i = 1$$

& $$\sum_{i=1}^{n}\sum_{j=1}^{n}\left(T_{ij}\sum_{l=1}^{n} p_l T_{lj}\right)$$
$$= \sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{l=1}^{n} p_l T_{lj} T_{ij}$$
$$= \sum_{j=1}^{n}\sum_{l=1}^{n} p_l T_{lj}\left(\sum_{i=1}^{n} T_{ij}\right)$$
$$= \sum_{j=1}^{n}\sum_{l=1}^{n} p_l T_{lj} = 1$$

$\Rightarrow$ $H(\vec{q}) - H(\vec{p})$ is the KL divergence b/w these 2 distrib^n's

$\Rightarrow H(\vec{q}) - H(\vec{p}) \geq 0$

$\Rightarrow H(\vec{p}T) \geq H(\vec{p})$

Q.3

a) Put $|A| = M$, $S = \sum_{i=1}^{M} 2^{-l_i}$

$\Rightarrow S^N =$

$$\sum_{i_1=1}^{M} \sum_{i_2=1}^{M} \cdots \sum_{i_N=1}^{M} 2^{-\left(\sum_{k=1}^{N} i_k\right)}$$

This is the length of the encoding of the string $x = a_{i_1} \cdot a_{i_2} \cdot \ldots \cdot a_{i_N}$

$\Rightarrow$ For every string of length N, say x, there is exactly one term in the big summation, which is $2^{-\text{encoded length}(x)}$ (since code is uniquely decodable)

Let $f(l)$ count the number of N length strings whose encodings have length l.

Now define

$$l_{min} = \min_{x \text{ is an } N \text{ length string}} \text{encoded length}(x)$$

$$l_{max} = \max_{x \text{ is an } N \text{ length string}} \text{encoded length}(x)$$

Clearly,

$$l_{min} = N \min_i l_i$$

$$l_{max} = N \max_i l_i$$

$$\Longrightarrow S^N = \sum_{l=l_{min}}^{l_{max}} 2^{-l} f(l)$$

But see that $f(l) \leq 2^l$. To see why, note that if the encoded length of a string is l, the length of the string itself is atmost l (since ~~each~~ the encoding of each symbol of non-zero probability is at least 1 bit and so the number of such strings is atmost 2^l.

$$\Rightarrow S^N \leq \sum_{l=l_{min}}^{l_{max}} 2^{-l} 2^l$$

$$= l_{max} - l_{min}$$

$$= N \left(\max_i l_i - \min_i l_i \right)$$

$$\Rightarrow S \leq \left(N \left(\max_i l_i - \min_i l_i \right) \right)^{1/N}$$

$$\Rightarrow S \leq \sup_N \left(N \left(\max_i l_i - \min_i l_i \right) \right)^{1/N}$$

$$= 1$$

Q.3 b) Let the lengths be, WLOG,

$$l_1 \leq l_2 \leq \ldots \leq l_n$$

Given that $\{l_i\}_{1 \leq i \leq n}$ satisfies the Kraft-McMillan inequality,

$$\sum_{i=1}^{M} 2^{-l_i} \leq 1 \quad \{\text{assume } M \geq 2\}$$

$$\Rightarrow \sum_{i=1}^{j} 2^{-l_i} \leq 1 \quad \forall\, 2 \leq j \leq M$$

$$\Rightarrow \sum_{i=1}^{j} 2^{l_j - l_i} \leq 2^{l_j}$$

$$\Rightarrow 1 + \sum_{i=1}^{j-1} 2^{l_j - l_i} \leq 2^{l_j}$$

$$\Rightarrow \sum_{i=1}^{j-1} 2^{l_j - l_i} < 2^{l_j} \quad \forall\, 2 \leq j \leq M$$

Now say we try & construct an instantaneous code sequentially.

1) There are 2^{l_1} codewords to choose from

2) There would have been 2^{l_2} codewords to choose from,

but because of the prefix constraints, $2^{l_2 - l_1}$ of these are forbidden.

eg. Say $l_1 = 1$, $l_2 = 3$ and say

we choose $c(a_1) = '0'$. Now we are unable to put $c(a_2) = '000', '001', '010', '011'$ since then code would not be instantaneous. $\Rightarrow 4 = 2^{l_2 - l_1}$ words forbidden.

3) There would have been

2^{l_3} words to choose from, but

$2^{l_3 - l_2} + 2^{l_3 - l_1}$ are forbidden.

☆ No two codewords can forbid the same codeword of a larger length since that would mean the smaller of the two was a prefix of the larger.

☆ At each step j, we are able to assign instantaneous codewords iff the number of choices available (2^{l_j}) outnumber the number forbidden by ~~[illegible]~~ smaller codelengths i.e. iff $\forall j \in \{2, \ldots, M\}$,

$$2^{l_j} > \sum_{i=1}^{j-1} 2^{l_j - l_i}$$

which is what we showed earlier.

c) Define $z = \sum_{i=1}^{M} 2^{-l_i}$ and $q_i = \frac{2^{-l_i}}{z}$. This is the distribution we shall work with.

$$L(C,X) = \sum_{i=1}^{M} p_i l_i$$

$$= \sum_{i=1}^{M} p_i \left(-\log_2 (q_i z)\right)$$

$$= -\sum_{i=1}^{M} p_i \log_2 q_i - \underbrace{\sum_{i=1}^{M} p_i \log_2 z}_{= \log_2 z}$$

$$= -\log_2 z - \sum_{i=1}^{M} p_i \log_2 p_i$$

$$+ \sum_{i=1}^{M} p_i \log_2 p_i - \sum_{i=1}^{M} p_i \log_2 q_i$$

$$= -\log_2 z + H(X) + D(p\|q)$$

But $z \leq 1 \Rightarrow \log_2 z \leq 0$

$$\Rightarrow -\log_2 z \geq 0$$

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

$$\Rightarrow L(C,X) \geq H(X)$$

Clearly, equality occurs iff

$$\underbrace{-\log_2 z}_{\downarrow} = \underbrace{D(p\|q) = 0}$$

$z = 1 \Rightarrow$ code is complete

$$\Rightarrow p = q \Rightarrow p_i = \frac{2^{-l_i}}{z} = 2^{-l_i}$$

$$\Rightarrow l_i = -\log_2 p_i$$

$$\Rightarrow l(a) = -\log_2 \varphi^{-1}(a)$$

d) For complete codes,

$$L(C,X) = H(X) + D(p\|q)$$

$$\Rightarrow L(C,X) - H(X) = D(p\|q)$$

$\Rightarrow$ The cost of not using the most optimal encoding is the KL divergence b/w p and the implicitly defined distribution q.

e) Set $l_i = \lceil -\log_2 p_i \rceil$. Since X has equal support & image, $\log_2 p_i$ is well defined. So,

$$-\log_2 p_i \leq l_i \Rightarrow \log_2 p_i \geq -l_i$$

$$\Rightarrow 2^{\log_2 p_i} \geq 2^{-l_i}$$

$$\Rightarrow \sum_{i=1}^{M} p_i \geq \sum_{i=1}^{M} 2^{-l_i}$$

$$\Rightarrow \sum_{i=1}^{M} 2^{-l_i} \leq 1$$

$\Rightarrow$ Kraft-McMillan inequality is satisfied.

$\rightarrow$ There is an instantaneous code with codelengths $\lceil -\log_2 p_i \rceil$

For this code,

$$L(C,X) = \sum_{i=1}^{M} p_i \lceil -\log_2 p_i \rceil$$

$$< \sum_{i=1}^{M} p_i (-\log_2 p_i + 1)$$

$$= \underline{H(X) + 1}$$