Tutorial - 4

Information &

Communication

Problem - 1 (Independent Random Variables)

Let X, X' be independent random variables with $X \sim p(x)$ and $X' \sim r(x)$, $x, x' \in \mathcal{X}$.

Then prove that

$$P(X = X') \geq 2^{-H(p) - D(p\|r)}$$

$$P(X = X') \geq 2^{-H(r) - D(r\|p)}$$

To prove: $P(X = X') \geq 2^{-H(p) - D(p\|r)}$

$$P(X = X') = \sum_{x \in \mathcal{X}} P(X = x, X' = x)$$

$$= \sum_{x \in \mathcal{X}} P(X = x) \cdot P(X' = x)$$

$$= \sum_{x \in \mathcal{X}} p(x) \cdot r(x)$$

Let $L = 2^{-H(p) - D(p\|r)}$

$$\Rightarrow \log_2(L) = -(H(p) + D(p\|r))$$

$$= -\sum_{x \in \mathcal{X}} p(x) \log_2 \frac{1}{p(x)} - \sum_{x \in \mathcal{X}} p(x) \log_2 \frac{p(x)}{r(x)}$$

$$= \sum_{x \in \mathcal{X}} p(x) \log_2 r(x)$$

$$\Rightarrow L = 2^{\sum_{x \in \mathcal{X}} p(x) \log r(x)}$$

Using Jensen's Inequality,

$$2^{\sum \lambda_z z} \leq \sum \lambda_z 2^z \quad \Big| \quad \sum \lambda_z = 1$$

2^z is a convex function

$$\Rightarrow L \leq \sum_{x \in \mathcal{X}} p(x) \cdot 2^{\log_2 r(x)}$$

$$= \sum_{x \in \mathcal{X}} p(x) \cdot r(x)$$

$$= P(X = X')$$

$$\therefore P(X = X') \geq 2^{-H(p) - D(p \| r)}$$

Alternate method I

Credits: Avyukta Manjunatha Vummintala

By A.M.-G.M. inequality, for positive numbers p_i and r_i, we have:

$$\Sigma p_i r_i \geq \Pi r_i^{p_i}$$

Since *log* is a increasing function, we can infer:

$$log(\Sigma p_i r_i) \geq \Sigma log(r_i^{p_i})$$

$$log(\Sigma p_i r_i) \geq \Sigma p_i log(r_i)$$

$$-log(\Sigma p_i r_i) \leq -\Sigma p_i log(r_i)$$

$$log(\Sigma \frac{1}{p_i r_i}) \leq \Sigma p_i log(\frac{1}{r_i})$$

$$log(\Sigma \frac{1}{p_i r_i}) \leq \Sigma p_i log(\frac{1}{p_i}) + \Sigma p_i log(\frac{p_i}{r_i})$$

Since positive exponentiation is a increasing function, we can infer that:

$$\Sigma \frac{1}{p_i r_i} \leq 2^{\Sigma p_i log(\frac{1}{p_i}) + \Sigma p_i log(\frac{p_i}{r_i})}$$

Credits: Yarramaneni Jaishnav

Alternate Method-2

Q1:-

$$-H(X) - D(p\|q)$$

I took $r(x)$ as $q(x)$

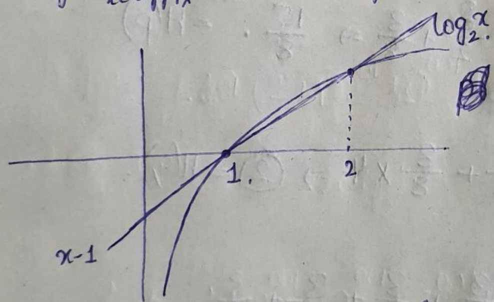
$$\Rightarrow -\left(\sum_{x \in supp P_X} \left(p(x)\log\left(\frac{1}{p(x)}\right) + p(x)\log\left(\frac{p(x)}{q(x)}\right)\right)\right)$$

$$\Rightarrow -\left(\sum_{x \in supp(P_X)} \left(p(x)\log\left(\frac{1}{q(x)}\right)\right)\right)$$

$$\Rightarrow \left(\sum_{x \in supp P_X} p(x)\log(q(x))\right).$$

$$\log_2(P(X=x')) - \sum_{x \in supp P_X} p(x)\log(q(x)).$$

$$\log_2\left(\sum_{x \in supp P_X} p(x)\cdot q(x)\right) - \sum_{x \in supp P_X} p(x)\log(q(x)) \quad (\because P(X'=x)=q(x)).$$



$$\because \sum_{x \in supp P_X} p(x)\cdot q(x) \leq 1, \quad \log_2\left(\sum_{x \in supp(P_X)} p(x)\cdot q(x)\right) \leq \left[\sum_{x \in supp(P_X)} p(x)\cdot q(x)\right] - 1.$$

$$\left(\sum_{x \in supp P_X} p(x)\cdot q(x)\right) - 1 - \sum_{x \in supp(P_X)} p(x)\log_2(q(x)).$$

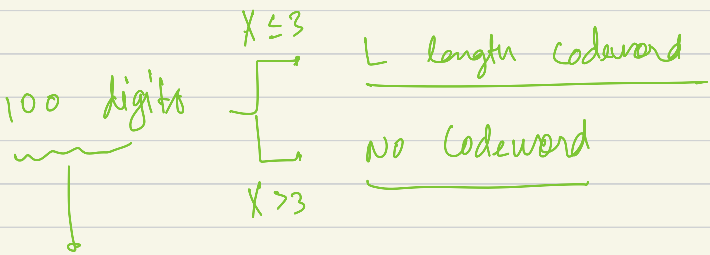
$$\text{let } T = \sum_{x \in supp(P_X)} p(x)\left[q(x) - 1 - \log_2(q(x))\right]$$

$$\because q(x) \leq 1 \qquad q(x) \geq 1 + \log_2(q(x)).$$

$$\therefore T \geq 0.$$

Problem-2 (Fixed length source coding)

A source emits a sequence of independent binary digits with probabilities $p(1) = 0.005$ and $p(0) = 0.995$. The digits are taken 100 at a time and a binary codeword is provided for every sequence of 100 digits containing three or fewer 1's.
Assuming that all codewords are the same length, find the minimum length required to provide unique codewords for all sequences with three or fewer 1's. Other sequences need not be assigned any codeword



X is the random variable denoting the number of 1's

Number of codewords

$$|C| = \binom{100}{0} + \binom{100}{1} + \binom{100}{2} + \binom{100}{3}$$

Minimum length (L)

$$L^{*} = \left\lceil \log_2(|C|) \right\rceil$$

Problem - 3 (Variable Length Source Coding)

Let the range of random variable X be $\{0, 1, 2, 3, 4\}$. Consider the two distributions $p(x)$ and $q(x)$ on this random variable.

Codes for random variable X

Symbol	$p(x)$	$q(x)$	$C_1(x)$	$C_2(x)$
1	$1/2$	$1/2$	0	0
2	$1/4$	$1/8$	10	100
3	$1/8$	$1/8$	110	101
4	$1/16$	$1/8$	1110	110
5	$1/16$	$1/8$	1111	111

(a) Calculate $H(p)$, $H(q)$, $D(p\|q)$ and $D(q\|p)$

(b) Check if C_1 and C_2 are prefix-free codes.

(c) Verify that average length of C_1 under p is equal to the entropy $H(p)$. Thus, C_1 is optimal for p. Verify that C_2 is optimal for q.

(d) Now assume that we use C_2 when distribution is p. What is the average length of the codewords? By how much does it exceed entropy $H(p)$?

Solution Credits : Yarramaneni Jaishnav

P_3 :-

(a) (1) $H(P) = -\sum_{x \in P_x} p(x) \log(p(x))$.

$\Rightarrow \frac{1}{2}\cancel{\log 2} + \frac{1}{4}\cancel{\log 4}^{2} + \frac{1}{8}\cancel{\log 8}^{3} + \frac{1}{16}\cancel{\log 16}^{4}$

$+ \frac{1}{16}\cancel{\log 16}^{4}$

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

$\frac{4 \times 3}{4 \times 2} + \frac{3}{8} \Rightarrow \frac{15}{8}$

(2) $H(q) = \sum_{x \in q_x} q(x) \log(q(x))$.

$\Rightarrow \frac{1}{2} + \frac{3}{8} + \frac{3}{8} + \frac{3}{8} + \frac{3}{8}$.

$\frac{12}{8} + \frac{4}{8} \Rightarrow \frac{16}{8} = 2$

(3). $\sum p(x) \log\left(\frac{p(x)}{q(x)}\right)$

$\Rightarrow \frac{1}{2}\cancel{\log(1)}^{0} + \frac{1}{4}\log(2) + \frac{1}{8}\log 1$

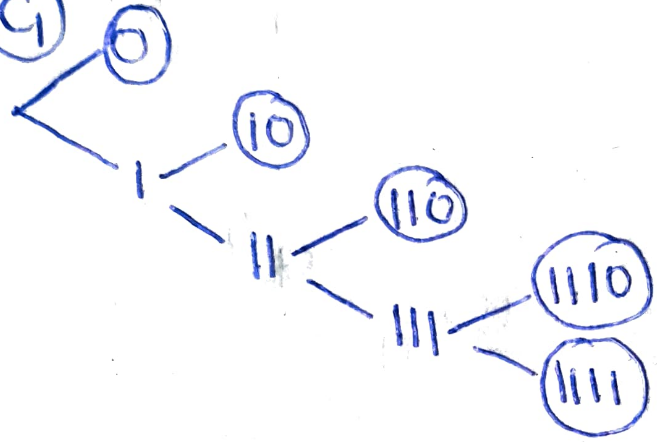
$+ \frac{1}{16}\log\left(\frac{1}{2}\right) + \frac{1}{16}\log\left(\frac{1}{2}\right)$.

$\frac{1}{4} - \frac{1}{8} \Rightarrow \frac{1}{8}$.

(4). $\cancel{\frac{1}{2}\log 1} + \frac{1}{8}\log\left(\frac{1}{2}\right) + \frac{1}{8}\log(2) + \frac{1}{8}\log 2.$

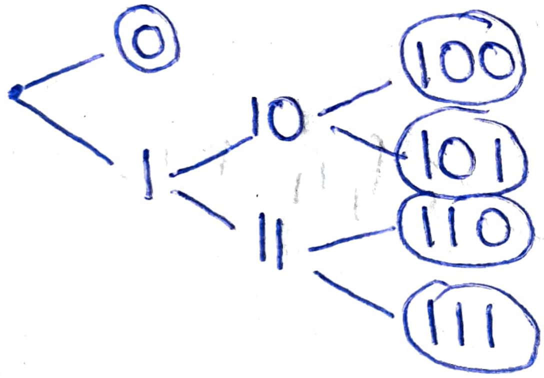
$$\frac{1}{8} + \frac{1}{4} \Rightarrow \frac{1}{8}.$$

(b) (i) (C_1)



∴ P-F notation.

(ii) (C_2)



∴ P-F notation.

(c) (1) $1 \cdot \frac{1}{2} + \frac{2}{4} + \frac{3}{8} + \frac{4}{16} + \frac{4}{16}$.

$\frac{3}{2} + \frac{3}{8} \Rightarrow \frac{15}{8} = H(p)$

(2) $\frac{1}{2} + \frac{3}{8} \times 4 \Rightarrow 2 = H(q)$.

(d) $\frac{1}{2} + \frac{3}{4} + \frac{3}{8} + \frac{3}{16} + \frac{3}{16}$.

$\frac{1}{2} + \frac{3}{4} + \frac{3}{8} + \frac{3}{8}$

$\frac{1}{2} + \frac{3}{2} \Rightarrow 2$

$2 - \frac{15}{8}$

$\Rightarrow \frac{1}{8}$ bits

exceded