Homework 1 solutions

1.

$P(X=0) = 1-p$, $0 \le p \le 1$

$P(X=1) = p$

(a) $h(p) = H_2(X) = -P(X=0)\log_2 P(X=0) - P(X=1)\log_2 P(X=1)$

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

(b) Observe that $h(1-p) = h(p)$.

$\therefore$ $h(p)$ is symmetric about $p = \frac{1}{2}$.

(c) $h(p) = -(1-p)\dfrac{\log(1-p)}{\log 2} - p\dfrac{\log p}{\log 2}$

$$h'(p) = \frac{d}{dp}h(p)$$

$$= \cancel{\frac{-(1-p)}{\log 2}\cdot\frac{1}{(1-p)}\cdot(-1)} + \frac{\log(1-p)}{\log 2} - \cancel{\frac{p}{\log 2}\cdot\frac{1}{p}} - \frac{\log p}{\log 2}$$

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

2.

$X = (\mathcal{X}, p)$

$Y = (\mathcal{Y}, q)$

Define $Z = (\mathcal{Z}, r)$

where $\mathcal{Z} = \mathcal{X} \times \mathcal{Y}$

$r(z) = p(x)\,q(y)$ for $z = (x, y)$

$$H_2(Z) = -\sum_{z\in\mathcal{Z}} r(z)\log_2 r(z)$$

$$= -\sum_{x\in\mathcal{X}}\sum_{y\in\mathcal{Y}} p(x)\,q(y)\log_2\big(p(x)\,q(y)\big)$$

$$= -\sum_{x\in\mathcal{X}} \sum_{y\in\mathcal{Y}} p(x)\, q(y) \log_2 p(x) \quad - \sum_{x\in\mathcal{X}} \sum_{y\in\mathcal{Y}} p(x)\, q(y) \log_2 q(y)$$

$$= -\sum_{x\in\mathcal{X}} p(x) \log_2 p(x) \cdot \underbrace{\sum_{y\in\mathcal{Y}} q(y)}_{=1} \quad - \sum_{y\in\mathcal{Y}} q(y) \log_2 q(y) \underbrace{\sum_{x\in\mathcal{X}} p(x)}_{=1}$$

$$= H_2(X) + H_2(Y)$$

$$\boxed{\therefore\ H_2(Z) = H_2(X) + H_2(Y)}$$

3. $Z = (\mathcal{Z}, r)$, $\mathcal{Z} = \mathcal{X} \times \mathcal{Y}$

Define $X = (\mathcal{X}, p)$, $Y = (\mathcal{Y}, q)$ with

$$p(x) \triangleq \sum_{y\in\mathcal{Y}} r(x,y), \quad x \in \mathcal{X}$$

$$q(y) \triangleq \sum_{x\in\mathcal{X}} r(x,y), \quad y \in \mathcal{Y}$$

(a) No. $Z = (\mathcal{Z}, r)$ can be determined from $X = (\mathcal{X}, p)$ and $Y = (\mathcal{Y}, q)$ only when

$$r(x,y) = p(x) \cdot q(y), \quad x \in \mathcal{X}, y \in \mathcal{Y}.$$

(b) $$H_2(Z) = -\sum_{\substack{x\in\mathcal{X}\\ y\in\mathcal{Y}}} r(x,y) \log_2 r(x,y)$$

$$= -\sum_{x,y} r(x,y) \log_2 \left(p(x) \cdot q(y|x)\right)$$

(with some slight abuse of notation to represent the conditional distribution).

$$= -\sum_{x,y} r(x,y) \log_2 p(x) \quad - \sum_{x,y} r(x,y) \log_2 q(y|x)$$

$$= -\sum_{x\in\mathcal{X}} \underbrace{\left(\sum_{y\in\mathcal{Y}} r(x,y)\right)}_{=p(x)} \log_2 p(x) \quad - \sum_{x\in\mathcal{X}} p(x) \sum_{y\in\mathcal{Y}} q(y|x) \log_2 q(y|x)$$

$$\therefore \; H_2(Z) = H_2(X) - \sum_x p(x) \sum_y q(y|x) \log_2 q(y|x) \quad \text{—} ①$$

For arbitrary pmfs $p(\cdot), q(\cdot)$, we know

$$D(p\|q) \geq 0 \quad , \text{ with } D(p\|q) = 0 \text{ iff } p = q$$

Hence, $\forall x$

$$D(q(\cdot|x) \| q(\cdot)) \geq 0$$

i.e. $$\sum_y q(y|x) \log_2 \frac{q(y|x)}{q(y)} \geq 0$$

i.e. $$-\sum_y q(y|x) \log_2 q(y|x) \leq -\sum_y q(y|x) \log_2 q(y) \quad \text{—} ②$$

Using ② in ①,

$$H_2(Z) \leq H_2(X) - \sum_{x,y} \overbrace{p(x) q(y|x)}^{\substack{r(x,y) \\ \|}} \log_2 q(y)$$

$$= H_2(X) - \sum_{y \in \mathcal{Y}} \underbrace{\left(\sum_{x \in \mathcal{X}} r(x,y) \right)}_{= q(y)} \log_2 q(y)$$

$$= H_2(X) + \underline{\underline{H_2(Y)}}$$

<u>Intuitively</u>, $H_2(X)$ is the entropy of first component of Z, while $H_2(Y)$ is the entropy of the second component alone. The entropy of Z can not exceed the sum of the entropies of each component.

(c) Note that

$H_2(Z) = H_2(X) + H_2(Y)$ with equality only when

$\forall x,$

$$-\sum_y q(y|x) \log_2 q(y|x) = -\sum_y q(y|x) \log_2 q(y) \quad \text{in } ②.$$

$\Leftrightarrow \forall x,$

$$D(q(\cdot|x) \| q(\cdot)) = 0$$

$\Leftrightarrow \quad \forall x, \; q(\cdot|x) = q(x)$

or $q(y|x) = q(y) \;, y \in \mathcal{Y}.$

In other words X and Y are independent and

$$r(x,y) = p(x) \cdot q(y) \quad , x \in \mathcal{X}, y \in \mathcal{Y}.$$

In Problem 2, this condition is satisfied and hence we obtain

$$H_2(Z) = H_2(X) + H_2(Y).$$

4. Let X denote the outcome in a single trial.

$$P(X = \text{"H"}) = P(X = \text{"T"}) = \frac{1}{2}.$$

$$\therefore H_2(X) = \log_2 2 = 1 \text{ bit}$$

To calculate entropy for 10 trials, look at the extended source of order 10.

$$H(X^{10}) = 10\, H(X) = 10 \text{ bits}.$$

$$\left(\text{Since } H(X^n) = n\, H(X)\right).$$

5. (a) Observe that the value of the sum $\sum_{x \in \mathcal{X}} 2^{-l(x)}$ decreases as the value of $l(x)$ increases for any $x \in \mathcal{X}$.

Hence, for any finite set $\mathcal{X}$, it is always possible to find $\{l(x), x \in \mathcal{X}\}$ s.that

$$\sum_{x \in \mathcal{X}} 2^{-l(x)} \leq 1$$

by choosing the lengths to be large enough.

Thus, Kraft inequality says there exists a prefix code $C : \mathcal{X} \to \mathbb{B}^*$ s.that

$$l_c(x) = l(x) \quad , \quad x \in \mathcal{X}$$

i.e. $$\sum_{x \in \mathcal{X}} 2^{-l_c(x)} \leq 1 . \qquad \text{—— (1)}$$

Now, if we want to construct a prefix code $C_\lambda : \mathcal{X} \to \mathbb{B}^*$ s. that

$$\sum_{x \in \mathcal{X}} 2^{-l_{c_\lambda}(x)} \leq \lambda \quad , \quad \lambda \in (0,1) \qquad \text{—— (2)}$$

we just need to pad the code $C : \mathcal{X} \to \mathbb{B}^*$ (from (1)) with the required no. of 0's (or 1's). Note that such padding will not affect the prefix condition.

For e.g., if we pad all codewords in $C : \mathcal{X} \to \mathbb{B}^*$ with "a" zeroes to get $C_\lambda : \mathcal{X} \to \mathbb{B}^*$, then

$$\sum_{x \in \mathcal{X}} 2^{-l_{c_\lambda}(x)} = \sum_{x \in \mathcal{X}} 2^{-(l_c(x) + a)} = 2^{-a} \sum_{x \in \mathcal{X}} 2^{-l_c(x)}$$

$$\leq 2^{-a} \quad \text{(from (1))}$$

$\therefore$ by choosing "a" s.that

$$2^{-a} \leq \lambda \quad \text{i.e.} \quad a \geq -\log_2 \lambda$$

we can ensure condition (2) is satisfied.

Hence, this is <u>NOT</u> an interesting fact.

(b) For a given alphabet $\mathcal{X}$, if we can find $\{l(x), x \in \mathcal{X}\}$ s.that

$$\sum_{x \in \mathcal{X}} 2^{-l(x)} = 1 \quad \text{—(3)}$$

then by Kraft inequality there exists at least one prefix code which satisfies the stated condition. So, the question is really whether we can find $\{l(x), x \in \mathcal{X}\}$ satisfying (3) for any given $\mathcal{X}$. Obviously, this depends only on the no. of symbols in $\mathcal{X}$ i.e. $|\mathcal{X}|$.

For $|\mathcal{X}|=1$, i.e. there only a single symbol, there is no need for any code and (3) is trivially satisfied by zero codeword length.

For $|\mathcal{X}|=2$, the codeword length $\{1, 1\}$ satisfy (3).

We will now use an <u>induction</u> argument to show such lengths can be found for any (finite) alphabet size.

Suppose there exist $\{l(x), x \in \mathcal{X}\} = \{l_1, \ldots, l_k\}$ satisfying (3) for $|\mathcal{X}| = k$.

We want to show there exists $\{l_1', \ldots, l_{k+1}'\}$ that satisfy (3) for $|\mathcal{X}| = k+1$. Simply choose $l_1' = l_1, \ldots, l_{k-1}' = l_{k-1}$, $l_k' = l_k + 1$, $l_{k+1}' = l_k + 1$.

Then, $$\sum_{i=1}^{k+1} 2^{-l_i'} = \sum_{i=1}^{k-1} 2^{-l_i} + \cancel{2} \times 2^{-(l_k+\cancel{1})} = \sum_{i=1}^{k} 2^{-l_i} = 1 \text{ (by assumption).}$$

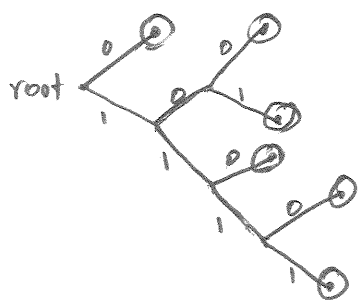
$\therefore$ it is indeed <u>TRUE</u> that there exists at least one prefix code $c: \mathcal{X} \to \mathcal{B}^*$ s. that $\sum_x 2^{-l_c(x)} = 1$.

6. Given lengths: 1, 3, 3, 3, 4, 4.

Kraft inequality satisfied?

$$2^{-1} + (2^{-3} \times 3) + (2^{-4} \times 2) = 1 \quad , \quad \text{YES!}$$

∴ there exists a prefix code with given codeword lengths.



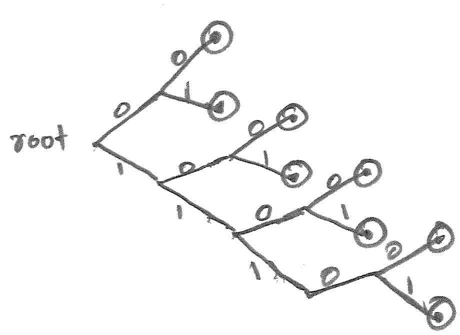
0
1 0 0
1 0 1
1 1 0
1 1 1 0
1 1 1 1

7. Given lengths: 2, 2, 3, 3, 4, 4, 5, 5

Kraft inequality satisfied?

$$(2^{-2} \times 2) + (2^{-3} \times 2) + (2^{-4} \times 2) + (2^{-5} \times 2) = \frac{15}{16} \leq 1 \quad . \quad \text{YES!}$$

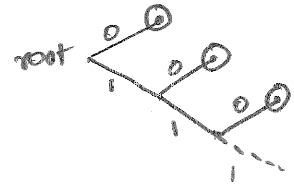
∴ there exists a prefix code with given codeword lengths



0 0
0 1
1 0 0
1 0 1
1 1 0 0
1 1 0 1
1 1 1 0 0
1 1 1 0 1

8. The prefix code contains

0
1 0
1 1 0.



So, the only prefix for the other codewords is 111.

Since we need the maximal number of codewords of <u>length</u> 5, we may include all those with prefix 111 :

11100
11101
11110
11111

as they satisfy the <u>prefix</u> condition.

$\therefore$ maximal no. of codewords of length $5 = 4$

and total codewords $= 3 + 4 = 7$.

Since each symbol in $\mathcal{X}$ is assigned a unique codeword

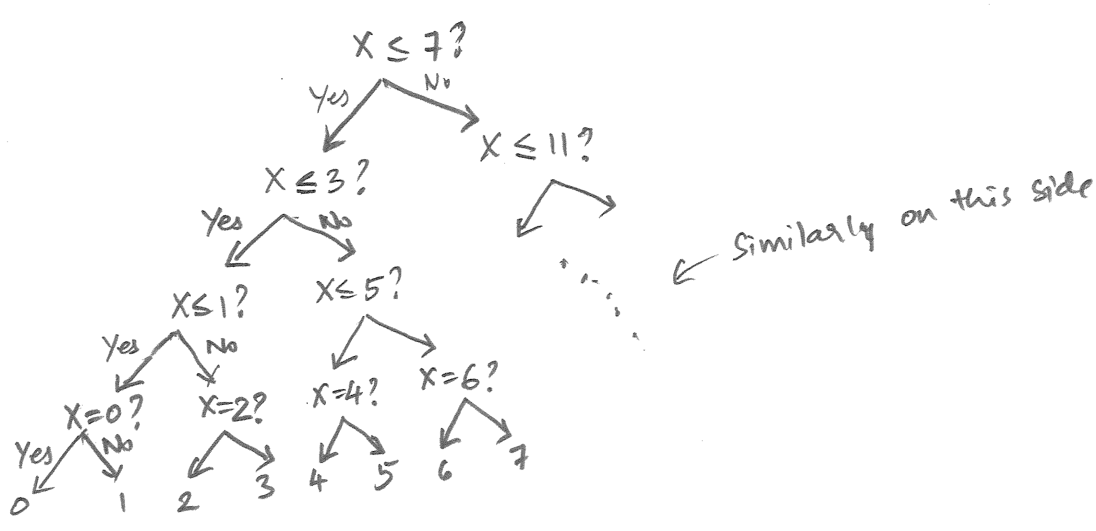
$$|\mathcal{X}| \geq 7.$$

9. To begin with, we have 16 possibilities for X.

$$\{0, 1, 2, \cdots, 15\}.$$

By asking each YES/NO question, we will shorten the no. of possibilities for X by $1/2$.

For eg.: we ask whether $X \leq 7$? If YES, we ask whether $X \leq 3$? and so on ...



$\therefore$ X can be determined with <u>4 questions</u>.

<u>Connection to entropy</u>: Since all values are equally probable

$$P(X = i) = \frac{1}{16}, \quad i = 0, 1, \cdots, 15$$

$$\therefore H_2(X) \equiv \log_2 16 = \underline{4}\,!$$

<u>10.</u> (a) $$H_2(Y) = -\sum_{y \in \mathcal{Y}} q(y) \log_2 q(y)$$

$$= -\sum_{y \in \mathcal{Y}} \left(\sum_{\substack{x \in \mathcal{X}: \\ g(x) = y}} p(x) \right) \log_2 \left(\sum_{\substack{x \in \mathcal{X}: \\ g(x) = y}} p(x) \right) \quad \text{— (1)}$$

Observe that, in general, ~~for~~ $a_i > 0$, $i = 1, \ldots n$

$$\left(\sum_{i=1}^{n} a_i \right) \log_2 \left(\sum_{i=1}^{n} a_i \right) = \left(\sum_{i=1}^{n} a_i \log_2 \left(\sum_{i=1}^{n} a_i \right) \right)$$

$$\geq \sum_i a_i \log_2 a_i$$

$$\therefore -\left(\sum_{i=1}^{n} a_i \right) \log_2 \left(\sum_{i=1}^{n} a_i \right) \leq -\sum_{i=1}^{n} a_i \log_2 a_i$$

with equality only when the summation involves only one term.

i.e. $\underline{n = 1}$.

Using this in (1),

$$H_2(Y) \leq -\sum_{y \in \mathcal{Y}} \sum_{\substack{x \in \mathcal{X}: \\ g(x) = y}} p(x) \log_2 p(x) \quad \text{— (2)}$$

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

(the two summations above are essentially the same as this single summation).

$$= H_2(X).$$

(b) For equality to hold, i.e. $H_2(Y) = H_2(X)$,

we need equality in 2.

This happens when the summation on $x \in \mathcal{X} : g(x) = y$

involves only one term <u>for all</u> $y \in \mathcal{Y}$.

i.e. $\forall y$, there exists a unique $x \in \mathcal{X}$ s.t. that $g(x) = y$

In other words, the function $g : \mathcal{X} \to \mathcal{Y}$ is <u>one-to-one</u>!