Solutions: Homework Set # 2

Problem 1

(a) X_i's are independent and identically distrtibuted, so

$$H(X_1, X_2, ..., X_n) = \sum_{i=1}^{n} H(X_i) = \sum_{i=1}^{n} H(p) = nH(p) \tag{1}$$

(b) $f(X_1, X_2, ..., X_n) = (Z_1, Z_2, ..., Z_K)$, where K is a function of $X_1, X_2, ..., X_n$ as well. So $(Z_1, Z_2, \cdots, Z_K, K)$ is a function of $(X_1, X_2, ..., X_n)$. We show that in general,

$$H(X) \geq H(f(X)), \tag{2}$$

for any function f.

$$H(X) = H(X, f(X) = H(f(X)) + \underbrace{H(X|f(X))}_{\geq 0 \text{(for discrete entropy)}} \tag{3}$$

$$\text{therefore } H(X) \geq H(f(X)). \tag{4}$$

This concludes that $H(X_1, X_2, ..., X_n) \geq H(Z_1, Z_2, \cdots, Z_K, K)$

(c) Chain rule for entropy.

(d)

$$H(Z_1, Z_2, ..., Z_K|K) = \sum p(K = k) \underbrace{H(Z_1, Z_2, ..., Z_K|K = k)}_{H(Z_1, Z_2, ..., Z_k)} \tag{5}$$

As $\{Zi\}$s are i. i. d. Bernoulli(1/2)

$$H(Z_1, Z_2, ..., Z_k) = \sum_{i=1}^{k} H(Z_i) = k \tag{6}$$

Thus,

$$H(Z_1, Z_2, ..., Z_K|K) = \sum k \cdot p(K = k) = E(K) \tag{7}$$

(e) $H(K) \geq 0$ for discrete entropy.

(f) Generate a good map f on sequences of length 4: we know that all sequences with the same numbre of ones are equally likely.

An example of a map f to generate random bits is:

$0000 \rightarrow \Lambda$

$0001 \rightarrow 00 \quad 0010 \rightarrow 01 \quad 0100 \rightarrow 10 \quad 1000 \rightarrow 11$

$0011 \rightarrow 00 \quad 0110 \rightarrow 01 \quad 1100 \rightarrow 10 \quad 1001 \rightarrow 11$

$1010 \rightarrow 0 \quad 0101 \rightarrow 1$

$1110 \rightarrow 00 \quad 1101 \rightarrow 01 \quad 1011 \rightarrow 10 \quad 0111 \rightarrow 11$

$1111 \rightarrow \Lambda$

Note that K is not fixed and we have both $K = 1$ and $K = 2$ in the above example.

Note further that the above map f has the property that

$\Pr\{Z_1Z_2 = 00|K = 2\} = \Pr\{Z_1Z_2 = 01|K = 2\} = \Pr\{Z_1Z_2 = 10|K = 2\} = \Pr\{Z_1Z_2 = 11|K = 2\}$ and $\Pr\{Z_1 = 0|K = 1\} = \Pr\{Z_1 = 1|K = 1\}$.

Problem 2

(a)

1. Left hand side:

$$H(X, Y, Z) - H(X, Y) = H(Z|X, Y) \tag{8}$$

Right hand side:

$$H(X, Z) - H(X) = H(Z|X) \tag{9}$$

Conditioning reduces entropy, thus, $H(Z|X) \geq H(Z|X, Y)$

2.

$$I(X; Z|Y) \geq I(Z; Y|X) - I(Z; Y) + I(X; Z) \tag{10}$$

$$\Leftrightarrow I(X; Z|Y) + I(Z; Y) \geq I(Z; Y|X) + I(X; Z) \tag{11}$$

$$\Leftrightarrow I(X, Y; Z) \geq I(Z; X, Y) \tag{12}$$

and indeed equality holds.

(b)

1.

$$I(X; Y|Z) < I(X; Y) \tag{13}$$

If $X \longrightarrow Y \longrightarrow Z$ forms a Markov chain, then

$$H(X|YZ) = H(X|Y) \tag{14}$$

Thus,

$$I(X; Y|Z) = H(X|Z) - H(X|YZ) = H(X|Z) - H(X|Y) \tag{15}$$

$$\underbrace{\geq}_{\text{conditioning reduces entropy}} H(X) - H(X|Y) = I(X; Y) \tag{16}$$

A trivial example is $Y = X$ and $Z = Y$.

2. Let X, Y be independent random variables and $Z = X + Y$

$$I(X; Y) = 0 \tag{17}$$

$$I(X; Y|Z) = H(X|Z) - H(X|YZ) = H(X|Z) \geq 0 = I(X; Y) \tag{18}$$

since $Z = X + Y$, $H(X|YZ) = 0$.

Problem 3

(a)

$$L(\mathcal{P}) = \sum_{i=1}^{n} p_i \, \ell_i \tag{19}$$

$$H(\mathcal{P}) = \sum_{i=1}^{n} -p_i \log p_i \tag{20}$$

(b) We will discuss both parts (b) and (d) here.

Why isn't the nature of the problem simply a part of the Huffman procedure as most of you argued in your homeworks?
In Huffman procedure, we may not construct the whole sub-tree $\mathcal{T}_u$ before continuing to the other nodes of $\mathcal{T}^u$. What we should do to check the validity of the Huffman trees, is to see whether $\mathcal{T}^u$ and $\mathcal{T}_u$ satisfy properties of optimal Huffman trees or not.

i) If $\ell_i < \ell_j$ then $p_i \geq p_j$;

- For $\mathcal{T}^u$: If $i \& j \neq u$, as $\mathcal{T}$ is a valid Huffman tree the property still holds from first property of $\mathcal{T}$. Now consider that $i = u$: In this case by construction of $\mathcal{T}$ probability of u is smaller than the nodes with less depth and is larger than the nodes with more depth, otherwise $\mathcal{T}$ wouldn't be a valid Huffman tree. For $j = u$ it is similar.
- For $\mathcal{T}_u$: This property obviously holds because $\mathcal{T}_u$ is part of $\mathcal{T}$ and we previously had:

$$\ell_i < \ell_j \Longleftrightarrow p_i \geq p_j$$

And now:

$$\ell_i - \ell < \ell_j - \ell \Longleftrightarrow \frac{p_i}{q} \geq \frac{p_j}{q}.$$

ii) The two least probable codewords have the largest length.

iii) The two least probable codewords differ only in one bit.
These two properties are simply guaranteed by Huffman procedure.

(c)

$$L(\mathcal{P}^u) = \sum_{i=1}^{k} p_i \, \ell_i + q \, \ell \tag{21}$$

$$H(\mathcal{P}^u) = \sum_{i=1}^{k} -p_i \log p_i - q \log q \tag{22}$$

(d) Refer to explanation in part (b).

(e)

$$L(\mathcal{P}_u) = \sum_{k+1}^{n} \frac{p_i}{q} (\ell_i - \ell) \tag{23}$$

$$H(\mathcal{P}_u) = \sum_{i=k+1}^{n} -\frac{p_i}{q} \log \frac{p_i}{q} \tag{24}$$

(f)

$$\begin{aligned} L(\mathcal{P}^u) + qL(\mathcal{P}_u) &= \sum_{i=1}^{k} p_i\, \ell_i + q\,\ell + q \sum_{i=k+1}^{n} \frac{p_i}{q}(\ell_i - \ell) \\ &= \sum_{i=1}^{n} p_i\, \ell_i + q\,\ell - \ell \sum_{i=k+1}^{n} p_i \\ &= \sum_{i=1}^{n} p_i\, \ell_i \\ &= L(\mathcal{P}) \end{aligned} \tag{25}$$

$$\begin{aligned} H(\mathcal{P}^u) + qH(\mathcal{P}_u) &= \sum_{i=1}^{k} -p_i \log p_i - q \log q - q \sum_{i=k+1}^{n} \frac{p_i}{q} \underbrace{\log \frac{p_i}{q}}_{\log p_i - \log q} \\ &= H(\mathcal{P}) \end{aligned} \tag{26}$$

Problem 4

(a) $\theta \to X \to T(X)$ forms a Markov chain. Applying data processing inequality,

$$I(\theta; T(X)) \le I(\theta; X). \tag{27}$$

If further $\theta \to T(X) \to X$ forms a Markov chain, then

$$I(\theta; X) \le I(\theta; T(X)) \tag{28}$$

And then equality holds.

(b) Because all sequences with the same number of ones are equally likely,

$$\Pr\{(X_1, X_2, \cdots, X_n) = (x_1, x_2, \cdots, x_n) \big| \sum_{i=1}^{n} X_i = k\} = \begin{cases} \frac{1}{\binom{n}{k}} & \text{if } \sum x_i = k \\ 0 & \text{otherwise} \end{cases} \tag{29}$$

Which is independent of θ.

Thus $\theta \to \sum X_i \to (X_1, X_2, \cdots, X_n)$ forms a Markov chain. Therefore,

$$I(\theta; X) = I(\theta; T(X)) \tag{30}$$

And this implies that $T(X)$ is a sufficient statistic for θ.