CS3236 Semester 2 2022/23:
Midterm **(Solutions)** (Total 50 Marks)

Name:____________________________

Matriculation Number:______________

Score:____________________________

You are given 1 hour and 30 minutes for this assessment. You are allowed one sheet of A4 paper, printed or written on both sides. Calculators are not permitted.

1. [**Entropy and Mutual Information**]

   (a) **(18 Marks)** Consider the pair (Z, X) described as follows for some $p \in (0, 1)$:

   - $Z \sim \text{Bernoulli}(p)$ (i.e., $P_Z(1) = p$ and $P_Z(0) = 1 - p$)
   - Given $Z = 0$, the conditional distribution of X is uniform on $\{1, 2, 3, 4\}$
   - Given $Z = 1$, the conditional distribution of X is uniform on $\{5, 6, 7, 8, 9, 10, 11, 12\}$

   Recall the definition $H_2(q) = q \log_2 \frac{1}{q} + (1 - q) \log_2 \frac{1}{1-q}$, which you may use in your answers if you wish.

   (i) Compute $H(Z)$, $H(X|Z = 0)$, $H(X|Z = 1)$, $H(X|Z)$, $H(Z|X)$, $H(X)$, $H(X, Z)$ and $I(X; Z)$, expressing your answers in terms of p when needed.

   (ii) Show that there exists at least one choice of p such that $H(X|Z = 1) > H(X)$, and another choice of p such that $H(X|Z = 1) < H(X)$.

   (iii) Do there exist values of $p \in [0, 1]$ such that $H(X|Z = 0) > H(X)$? Explain briefly.

   Solution. *(i) Since Z is Bernoulli distributed, we simply have $H(Z) = H_2(p)$.*

   Since the uniform distribution on M symbols has entropy $\log_2 M$, we have $H(X|Z = 0) = \log_2 4 = 2$ and $H(X|Z = 1) = \log_2 8 = 3$. Then the definition of conditional entropy gives $H(X|Z) = 2(1 - p) + 3p = 2 + p$.

   Next, X has 4 symbols with probability $\frac{1-p}{4}$ each, and 8 symbols with probability $\frac{p}{8}$ each, so $H(X) = 4 \cdot \frac{1-p}{4} \cdot \log_2 \frac{4}{1-p} + 8 \cdot \frac{p}{8} \cdot \log_2 \frac{8}{p} = (1-p) \log_2 \frac{4}{1-p} + p \log_2 \frac{8}{p}$. **(Alternatively, use the paragraph after this one to get $H(X) = H(X, Z)$, then expand $H(X, Z) = H(Z) + H(X|Z) = H_2(p) + 2 + p$ (by substituting the previous answers). It's straightforward to show that this is the same as the more explicit form just stated.)**

   $H(Z|X) = 0$, since X uniquely determines Z. Then the chain rule gives $H(X, Z) = H(X) + H(Z|X) = H(X) = (1 - p) \log_2 \frac{4}{1-p} + p \log_2 \frac{8}{p}$.

   Finally, $I(X; Z) = H(Z) - H(Z|X) = H(Z) = H_2(p)$.

(ii) Recall that $H(X|Z=1)=3$. If $p=\frac{1}{2}$, then we have $H(X)=\frac{1}{2}\log_2 8+\frac{1}{2}\log_2 16 > \log_2 8 = 3$, so $H(X|Z=1) < H(X)$.

On the other hand, as p approaches zero, $H(X)$ approaches $\log_2 4 = 2$; hence, we have for sufficiently small p that $H(X|Z=1) > H(X)$.

(iii) We have $H(X)=(1-p)\log_2\frac{4}{1-p}+p\log_2\frac{8}{p}$, and since $p\le 1$ and $1-p\le 1$ this gives $H(X)\ge(1-p)\log_2 4+p\log_2 8\ge\log_2 4=2$. Since $H(X|Z=0)=2$, we conclude that there are no values of p that satisfy $H(X|Z=0)>H(X)$. **(Alternatively, use the form $H_2(p)+p+2$ and simply note that p and $H_2(p)$ are both non-negative.)**

(b) **(6 Marks)** Suppose that for some triplet of random variables (X,Y,Z), we know that $I(Y;Z)-I(X;Z)=1$. Then can we conclude that $I(Y;Z|X)-I(X;Z|Y)$ is less than 1, equal to 1, or greater than 1, or is it the case that two or more of these are possible based on the information given? Explain your answer.

Solution. *Equal to 1. This is because using the chain rule in two different ways, $I(X,Y;Z)$ is equal to both $I(X;Z)+I(Y;Z|X)$ and $I(Y;Z)+I(X;Z|Y)$. Equating these two expressions and re-arranging gives $I(Y;Z)-I(X;Z)=I(Y;Z|X)-I(X;Z|Y)$.*

2. **[Source Coding Algorithms]**

In both parts (a) and (b) below, you should assume that $P_X(x)>0$ for all x under consideration, i.e., there are no zero-probability symbols.

(a) **(12 Marks)** This question concerns Shannon-Fano coding. We use letters (a, b, c, ...) to denote the symbols in the source.

<u>Notes</u>:

- *The following are 4 separate questions, and none of the information from one should be assumed in any of the other 3.*
- *When the question mentions "potentially taking values", this means that the corresponding quantity may (or may not) take different values depending on the alphabet size, the unspecified P_X values, etc., and your task is to identify all values that may occur when those vary.*

Let $\ell(a),\ell(b),\dots$ denote the codeword lengths of the Shannon-Fano code corresponding to source P_X, and answer the following:

(i) If we (only) know that $P_X(a)=8P_X(b)$ and $\ell(a)=2$, then what integer values could the length $\ell(b)$ potentially take?

(ii) If we (only) know that $P_X(a)=6P_X(b)$ and $\ell(a)=2$, then what integer values could $\ell(b)$ potentially take?

(iii) If there are 10 symbols (i.e., a–j) and symbol a is the **least** likely among them, what values could the length $\ell(a)$ potentially take?

(iv) If there are 10 symbols (i.e., a–j) and symbol a is the **most** likely among them, what values could the length $\ell(a)$ potentially take?

Solution.

(i) We have $\ell(b)=\lceil\log_2\frac{1}{P_X(b)}\rceil=\lceil\log_2\frac{8}{P_X(a)}\rceil=\lceil\log_2\frac{1}{P_X(a)}+3\rceil=\ell(a)+3=5$.

(ii) Similar to the first part, we have $\ell(b)=\lceil\log_2\frac{6}{P_X(a)}\rceil=\lceil\log_2\frac{1}{P_X(a)}+\log_2 6\rceil$. Note that $\log_2 6=2+\Delta$ for some fixed $\Delta\in(0,1)$ [it's $\Delta\approx 0.585$, but that's not important],

and $\log_2 \frac{1}{P_X(a)}$ *could be anything in the range* $(1, 2]$ *in order to get rounded up to 2. Hence,* $\ell(b)$ *could be either be 4 or 5 depending on whether* $\log_2 \frac{1}{P_X(a)}$ *is above or below* $1 + (1 - \Delta)$.

(iii) The smallest of the 10 probabilities could be anything in $(0, \frac{1}{10}]$ *(any higher would give a total probability exceeding one). Since a probability of* $\frac{1}{10}$ *corresponds to a length of* 4, *the value of* $\ell(a)$ *could be any integer with value 4 or higher.*

(iv) The largest of the 10 probabilities could be anything in $[\frac{1}{10}, 1)$ *(any lower would give a total probability less than one). Since a probability of* $\frac{1}{10}$ *corresponds to a length of* 4, *the value of* $\ell(a)$ *could be any of* $\{1, 2, 3, 4\}$.

(b) **(14 Marks)** This question concerns Huffman coding. Recall that the Huffman algorithm repeatedly merges two nodes to create a new node whose value sums those of the two being merged. Let $P_X(\cdot)$ be the source distribution (with an unspecified alphabet size). *(Note: The following are 4 separate questions, and none of the information from one should be assumed in any of the other 3)*

(i) If $P_X(a) = \frac{1}{2}$, explain why the Huffman algorithm will only ever merge symbol a's node with another node (possibly formed from merging) whose value is also $\frac{1}{2}$.

(ii) If we (only) know that $P_X(a) = 0.2$, then there are many possible values of the node that a is eventually merged with (depending on the alphabet size and the other P_X values). Find the entire set of such possible values, explaining your answer.

(iii) Can the Huffman code lengths potentially satisfy $\sum_x 2^{-\ell(x)} > 1$? Explain.

(iv) Can the Huffman code lengths potentially satisfy $\sum_x 2^{-\ell(x)} < 1$? Explain.

Solution.

(i) Suppose for contradiction that a gets merged with same value $v < 0.5$. *Then since all remaining nodes must sum to one, we conclude that one or more other nodes exist totaling* $1 - 0.5 - v = 0.5 - v$. *But this means that there are at least two nodes with value less than* 0.5, *contradicting the fact that a (with value* 0.5*) was chosen for merging.*

(ii) We first argue that anything in $(0, 0.4]$ *is possible. This is because if there are 3 probabilities* $(0.2, \alpha, 0.8 - \alpha)$ *with* $\alpha \in (0, 0.4]$, *then the values* 0.2 *and* α *will get merged. On the other hand, if the alphabet size is at least 3, we claim that* 0.2 *can't get merged with any node exceeding* 0.4. *This is because the values sum to one, so whenever at least 3 values remain, if one of them is above* 0.4 *(and one of them is* 0.2*) then the others must be below* 0.4. *This means that if* 0.2 *gets merged on such a step, it can only be with one of the ones below* 0.4.

In addition, if the alphabet size is 2, then the source probabilities can only be 0.2 *and* 0.8. *In this case, merging with* 0.8 *is possible.*

Overall answer: The set of possible values is $(0, 0.4] \cup \{0.8\}$.

(iii) No, because Huffman codes are prefix-free and cannot violate Kraft's inequality.

(iv) No, because if $\sum_x 2^{-\ell(x)} < 1$ *then we can shorten one or more of the lengths to create a new code such that* $\sum_x 2^{-\ell(x)} = 1$. *This contradicts the fact that the Huffman code has the smallest possible average length.*

(Alternatively, use the fact that if we start at the final merged node of the Huffman tree and branch randomly with probability $\frac{1}{2}$ each, then some codeword will definitely be reached, and if its length is $\ell(x)$ then the associated probability is $2^{-\ell(x)}$. These probabilities must sum to one.)