CS3236 Semester 2 2019/20:
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 needed.

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

   (a) **(10 Marks)** Suppose that $X \sim \text{Bernoulli}(p)$ (i.e., $P_X(1) = p$ and $P_X(0) = 1 - p$), and that the conditional distribution of Y given X is as follows: $P_{Y|X}(0|0) = 1$, $P_{Y|X}(1|0) = 0$, $P_{Y|X}(0|1) = \delta$, and $P_{Y|X}(1|1) = 1 - \delta$.

   Compute $H(Y)$, $H(Y|X)$, and $I(X;Y)$, expressing your answers in terms of p and δ, and using the definition $H_2(q) = q \log_2 \frac{1}{q} + (1-q)\log_2 \frac{1}{1-q}$ in your answers if you wish.

   Solution. *(i) Since $Y = 1$ only occurs when $X = 1$ and a flip does not occur, we have $P_Y(1) = p(1-\delta)$, and hence $H(Y) = H_2(p(1-\delta))$.*

   (ii) Since Y is deterministic given $X = 0$, we have $H(Y|X=0) = 0$. In addition, $H(Y|X=1) = H_2(\delta)$, and hence $H(Y|X) = \sum_x P_X(x) H(Y|X=x) = pH_2(\delta)$.

   (iii) $I(X;Y) = H(Y) - H(Y|X) = H_2(p(1-\delta)) - pH_2(\delta)$.

   (b) **(6 Marks)** Let X and Y be discrete random variables with alphabets $\mathcal{X}$ and $\mathcal{Y}$, and suppose that $P_X(x) > 0$ for all $x \in \mathcal{X}$. Prove that if $H(Y|X) = 0$, then it must be the case that Y is a deterministic function of X (i.e., for all $x \in \mathcal{X}$, some value of $y \in \mathcal{Y}$ has $P_{Y|X}(y|x) = 1$ and the rest have $P_{Y|X}(y|x) = 0$).

   Solution. *By the definition of conditional entropy, $H(Y|X) = \sum_x P_X(x) H(Y|X=x)$. Since entropy is non-negative and $P_X(x) > 0$ for all x, the only way for the sum to be zero is for $H(Y|X=x)$ to be zero for all $x \in \mathcal{X}$. But entropy is only zero for deterministic random variables, so we conclude that Y must be deterministic given X.*

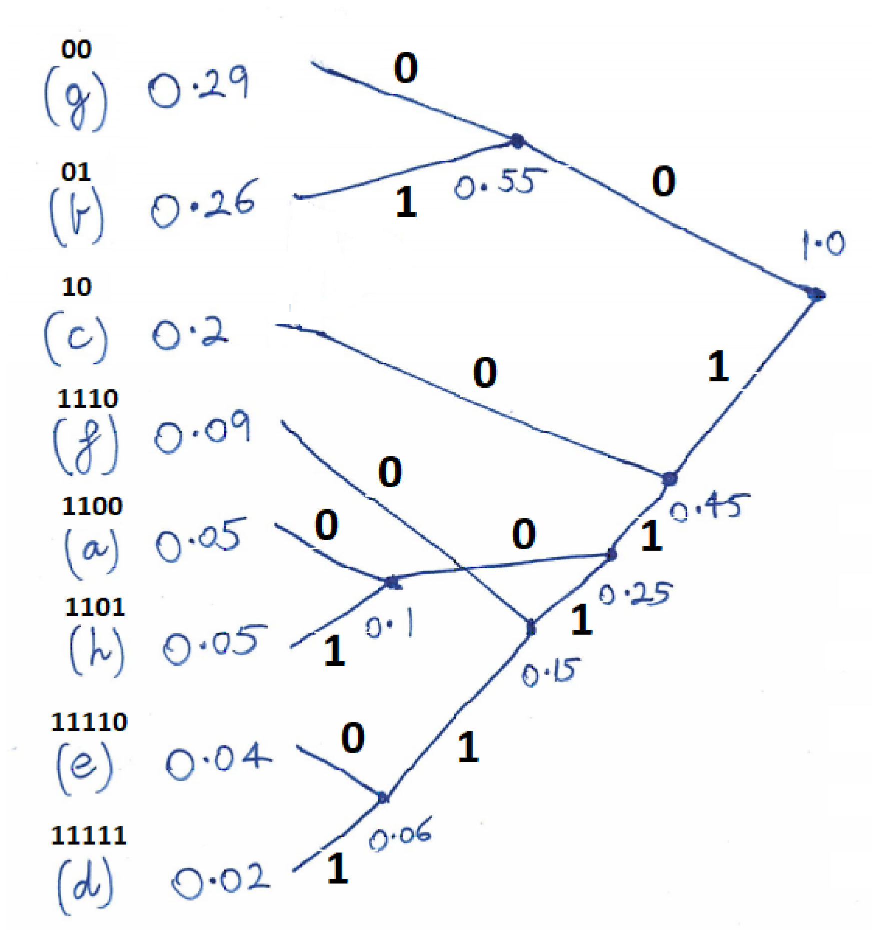
   (c) **(6 Marks)** Give an example of a collection of four random variables (X_1, X_2, Y_1, Y_2) such that $I(X_1, X_2; Y_1, Y_2) > I(X_1; Y_1) + I(X_2; Y_2)$. Explain your answer.

   Solution. *Let U and V be independent* $\text{Bernoulli}\big(\frac{1}{2}\big)$ *random variables, and $X_1 = Y_2 = U$ and $X_2 = Y_1 = V$. Then $I(X_1;Y_1) = I(U;V) = 0$ (by independence) and $I(X_2;Y_2) = I(U;V) = 0$. But $I(X_1, X_2; Y_1, Y_2) = I(U, V; U, V) = H(U, V) = 2$.*

2. [**Source Coding Algorithms**]

(a) **(8 Marks)** Apply the Huffman algorithm to the source with symbols (a, b, c, d, e, f, g, h) and corresponding probabilities $(0.05, 0.26, 0.2, 0.02, 0.04, 0.09, 0.29, 0.05)$. Furthermore, write down the codeword associated with each symbol.

Solution. *Arranging the symbols in descending order of probability and applying the algorithm gives the following (the codewords are shown on the left):*



(b) **(10 Marks)** Let X be a discrete random variable on the alphabet $\mathcal{X} = \{1, 2, 3, 4, 5, 6, 7, 8\}$ with probability mass function P_X such that $\min_{x \in \mathcal{X}} P_X(x) > 0$. For a symbol code $C(\cdot)$, let $\ell_{\min}(C) = \min_{x \in \mathcal{X}} \ell(x)$ be the length of its shortest codeword.

(i) Explain why the Huffman code may give $\ell_{\min}(C) = 1$ for some P_X.

(ii) Explain why the Huffman code cannot give $\ell_{\min}(C) > 3$ for any P_X.

(iii) Could the Shannon-Fano code give $\ell_{\min}(C) > 3$ for some P_X? Explain.

Solution. *(i) If $P_X(1) > 0.5$, then symbol 1 will be merged last in the Huffman tree, and will be assigned a length of one.*

(ii) The Huffman code is optimal. If all codeword lengths had length 4 or above, the average length would be at least 4, which is clearly suboptimal because we can assign each symbol a unique codeword of length 3.

(iii) No. The most probable codeword must have probability at least $\frac{1}{8}$, so it is assigned a length of at most 3.

(c) **(10 Marks)** Now consider the same setup as part (b) (i.e., $\mathcal{X} = \{1, 2, 3, 4, 5, 6, 7, 8\}$ and $\min_{x \in \mathcal{X}} P_X(x) > 0$), but instead consider $\ell_{\max}(C) = \max_{x \in \mathcal{X}} \ell(x)$, the length of the longest codeword.

(i) There exists an integer A such that the Huffman code may give $\ell_{\max}(C) = A$ for some P_X, but can never give $\ell_{\max}(C) < A$ for any P_X. Find a suitable choice of A and explain why these claims are true.

(ii) There exists an integer B such that the Huffman code may give $\ell_{\max}(C) = B$ for some P_X, but can never give $\ell_{\max}(C) > B$ for any P_X. Find a suitable choice of B and explain why these claims are true.

(iii) Could the Shannon-Fano code give $\ell_{\max}(C) > 100$ for some P_X? Explain.

Solution. *(iii) $A = 3$ is possible, e.g., for a uniform source. $A < 3$ is not possible, because if all lengths were are most 2, they would satisfy $\sum_{x=1}^{8} 2^{-\ell(x)} \geq \sum_{x=1}^{8} 2^{-2} = 2$, violating Kraft's inequality.*

(iv) $B = 7$. With 8 symbols, the Huffman algorithm performs 7 merging steps, and each one of these can add 0 or 1 to the longest codeword length. For instance, the longest length is indeed 7 when the probabilities are $\left\{\frac{1}{2}, \frac{1}{4}, \frac{1}{8}, \frac{1}{16}, \frac{1}{32}, \frac{1}{64}, \frac{1}{64}\right\}$.

(v) Yes. If $P_X(1) = 2^{-101}$, then symbol 1 is assigned a codeword of length $\left\lceil \log_2 \frac{1}{P_X(1)} \right\rceil = 101$.