Information Theory Midsemseter Examination

Write your roll number in the space provided on the top of each page. *Write your solutions clearly in the space provided after each problem. You may use additional sheets for working out your solutions; attach those sheets at the end of the question paper.* **Attempt all problems.**

Name and Roll Number: ______________________________

Problem	Points	Score
1	15	
2	15	
3	15	
4	15	
5	15	
Total:	75	

1. We are given the following specifications for designing a prefix-free code over the alphabet $\{0, 1\}$. The first row has the symbol that we wish to encode. Under the symbol, in second row, is the length of the codeword we wish to assign to the symbol. Finally, the third row has the weight of the symbol (the higher the weight the more important the symbol). Let $A = \{a_1, a_2, a_3, a_4, a_5, a_6, a_7, a_8\}$ be the set of symbols. For $a \in A$, let $\ell(a)$ denote the length specified for a and let $\mathsf{wt}(a)$ denote its weight.

symbol a	a_1	a_2	a_3	a_4	a_5	a_6	a_7	a_8
length $\ell(a)$	1	2	2	3	3	3	3	2
weight $\mathsf{wt}(a)$	7	7	8	5	3	10	5	6

(a) State why there is no *prefix-free* encoding of symbols in A so that each symbol receives a codeword of the length specified for it in the table. [5]

(b) Ignoring the lengths specified in the table, design a prefix-free binary encoding $\mathsf{enc} : A \to \{0, 1\}^*$ of minimum cost $\sum_{a \in A} \mathsf{wt}(a) \cdot |\mathsf{enc}(a)|$. Do not list the codewords. Just present the code tree with eight leaves labelled, $a_1, \ldots, a_8$ that accurately describes your encoding; separately write down the cost of your encoding. [10]

2. There are four javelin throwers T_1, T_2, T_3, T_4 training for an athletic meet. In each round they each take turns to throw in the following order: $T_1, T_2, T_3, T_1, T_4, T_3, T_1$. (Note that in each round, T_1 throws three times, T_3 throws twice, and T_2 and T_1 throw only once.) The four throwers are known to throw the javelin according to the following distributions, which are known in advance to the coach.

	A foul	B (0, 82]	C (82, 84]	D (84, 86]	E (86, ∞]
T_1	0.1	0.2	0.4	0.2	0.1
T_2	0.2	0.5	0.1	0.1	0.1
T_3	0.1	0.4	0.3	0.1	0.1
T_4	0.0	0.1	0.2	0.5	0.2

Assume the throws are independent. A large number N of rounds are completed during the day, and the data from the throws needs to be recorded on the coach's computer. Once the javelin lands, the coach takes down the distance the javelin has travelled. So the coach's record for the day is a sequence of $7N$ letters, e.g., $A, B, A, C, D, A, A, B, \ldots$, which indicates that the first throw was foul, the second throw was in the range $(0, 82]$ (metres), the third throw was a foul, the fourth was in the range $(82, 84]$, and so on. At the end of the day, the coach encodes the entire data for the day as a single long sequence of bits and transmits it to the computer. On receiving the sequence of bits, the computer decodes it back into a sequence of $7N$ letters, and saves it as the data for that day.

(a) What is the minimum number of bits *per round* that the coach must send if we require that with probability at least 0.99, the data for the day is accurately recorded on the computer? You may use the shorthand row_i for T_i's distribution; e.g., $H(\mathsf{row}_2) = -(0.2 \log 0.2 + 0.5 \log 0.5 + 0.3 \log 0.1)$. [7]

Answer: approx. ______________________________ per round.

(b) Explain, in say five sentences, how you arrived at your answer. (Note that you must state why the answer is both a lower and an upper bound.) [8]

3. Let $\mathsf{enc} : \{0,1\}^n \to \{0,1\}^k$ be an encoding function and $\mathsf{dec} : \{0,1\}^k \to \{0,1\}^n$ a decoding function. Let $\mathbf{X}$ be distributed uniformly in $\{0,1\}^n$. Let $\mathbf{M} = \mathsf{enc}(\mathbf{X})$ and $\mathbf{Y} = \mathsf{dec}(\mathsf{enc}(\mathbf{X}))$. Suppose $\Pr[X_i = Y_i] = \eta \geq \frac{1}{2}$, where the probability is taken over $\mathbf{X}$ chosen uniformly at random from $\{0,1\}^n$, and i chosen uniformly in $\{1,2,\ldots,n\}$ independently of $\mathbf{X}$. Then,

$$\begin{aligned} k &\geq I[X_1 X_2 \ldots X_n : \mathbf{M}] && (1)\\ &\geq \sum_i I[X_i : \mathbf{M}] && (2)\\ &\geq \sum_i I[X_i : Y_i] && (3)\\ &\geq \sum_i (1 - h(\Pr[X_i = Y_i])) && (4)\\ &\geq (1 - h(\eta))n. && (5) \end{aligned}$$

Justify each inequality formally. You must say in which steps you use the independence of random variables.

(a) Justify (1). [2]

(b) Justify (2). [6]

(c) Justify (4). E.g., you might start by defining $Z_i = X_i + Y_i \bmod 2$. [5]

(d) Justify (5). [2]

4. Suppose $P = (p_1, p_2, \ldots, p_n)$ and $Q = (q_1, q_2, \ldots, q_n)$ are probability distributions on the set $[n] = \{1, 2, \ldots, n\}$. We say that P majorizes Q if for $k = 1, 2, \ldots, n$,

$$\max_{S \subseteq [n], |S|=k} \sum_{i \in S} p_i \geq \max_{S \subseteq [n], |S|=k} \sum_{i \in S} q_i.$$

(That is, the maximum probability in P is at least the maximum probability in Q; the sum of the maximum and second maximum probabilities in P is at least the sum of the maximum and the second-maximum probabilities in Q; and so on.) [For both parts, you may use the fact that every doubly stochastic matrix is a convex combination of permutation matrices, that is, the $n \times n$ doubly stochastic matrix M can be written in the form

$$M = \sum_{\sigma} \alpha_\sigma M_\sigma,$$

where σ ranges over the set of permutations of $[n]$, the coefficients α_σ are non-negative ($\sum_\sigma \alpha_\sigma = 1$), and M_σ is the $n \times n$ matrix of zeros and ones, whose (i, j)-th entry is 1 iff $\sigma(i) = j$.]

(a) Regard P and Q as row vectors. Suppose $Q = PM$, where M is an $n \times n$ doubly stochastic matrix. Show that P majorises Q. 5

(b) The converse of the above statement is also true (you don't have to provide a proof): 10

if P majorises Q, then there is a doubly stochastic matrix M such that $Q = PM$;

Using the above statement (or otherwise) show that if P majorises Q, then $H(Q) \geq H(P)$.

5. We represent a probability distribution on $\{0,1\}$ by a pair (p_0, p_1), where p_0 is the probability of 0 and p_1 is the probability of 1. In a homework problem, we derived the inequality

$$D((p+t, 1-p-t)\|(p, 1-p)) \geq (2 \log e) \cdot t^2. \tag{6}$$

Generalize this and show the following. Let P and Q be distributions on a finite set S. Then,

$$D(P\|Q) \geq \frac{1}{2 \ln 2} \|P - Q\|_1^2, \tag{7}$$

where $\|P - Q\|_1 = \sum_{i \in S} |P(i) - Q(i)|$ is the ℓ_1 distance between P and Q. Proceed as follows. Let $X \sim P$ and $Y \sim Q$. Argue that there is a function $f : [n] \to \{0,1\}$, such that the distributions P' and Q' of $f(X)$ and $f(Y)$ satisfy $\|P - Q\|_1 = \|P' - Q'\|_1$ and $D(P\|Q) \geq D(P'\|Q')$.

(a) Define the function f. [3]

(b) Why does $\|P - Q\|_1 = \|P' - Q'\|_1$ hold? [2]

(c) Why does $D(P\|Q) \geq D(P'\|Q')$ hold? You may use without proof properties of relative entropy discussed in class, but you must clearly state which property you are using. [7]

(d) Why does (7) follow from this? (You may assume that (6) holds for distributions on $\{0,1\}$.) [3]