Information Theory Semester Examination (3 hours, 14 Nov 2024)

Write your roll number in the space provided on the top of each page. *Write your solutions clearly in the space provided after each question; if the space is insufficient, write your answers on additional sheets and mark them clearly. You may also use additional sheets for working out your solutions; attach those sheets at the end of the question paper.* **Attempt all problems.**

Name: ______________________________

Roll number: ______________

Problem	Points	Score
1	20	
2	20	
3	20	
4	20	
5	5	
6	5	
7	10	
Total:	100	

1. (a) Suppose X is a random variable that takes values in the set $A = \{a_1, a_2, \ldots, a_n\}$. Assume that $p_i = \Pr[X = a_i] > 0$ for all i. Suppose $e : A \to \{0,1\}^*$ is a prefix-free encoding of the elements of A obtained using the Huffman coding algorithm. Show that [7]

$$H[X] \leq \mathbb{E}[\mathsf{len}(e(X))] < H[X] + 1,$$

where $\mathsf{len}(x)$ is the number of bits in x. You may use facts discussed in class, but you should state them clearly in your answer.

(b) Which of the two inequalities above might fail to hold if there are symbols with probability 0, and they are not excluded before running the Huffman coding algorithm? For each inequality that you think might fail to hold, provide a counter-example. [3]

(c) Write down the parsing produced in the first phase of the Lempel-Ziv algorithm for the following string: [10]

01011010000001011

[E.g., the parsing of 1011010100010 is (1), (0), (11), (01), (010), (00), (10).]

Answer:____

2. (a) Let X, Y and Z be random variables taking values in a set A. Let f be a function with domain A. For a random variable X, let P_X denote its probability distribution. In the following, fill-in the blanks with $=$, $\leq$, or $\geq$ as appropriate. (You will not get credit if your fill-in an inequality when, in fact, equality holds. ☺) [6]

$$H[(X, f(X))] \text{ \_\_\_\_\_\_\_\_\_\_ } H[X]$$

$$H[Y|X] \text{ \_\_\_\_\_\_\_\_\_\_ } H[Y]$$

$$I[(X,Z):Y] \text{ \_\_\_\_\_\_\_\_\_\_ } I[X:Y]$$

$$I[(X,Z):Y] \text{ \_\_\_\_\_\_\_\_\_\_ } I[X:Y|Z]$$

$$D(P_{(X,f(X))} \| P_{(Y,f(Y))}) \text{ \_\_\_\_\_\_\_\_\_\_ } D(P_X \| P_Y)$$

$$D(P_{(X,f(X))} \| P_{(Y,f(Y))}) \text{ \_\_\_\_\_\_\_\_\_\_ } D(P_{f(X)} \| P_{f(Y)}).$$

(b) For a real number λ, let X_λ be the random variable taking values in the set $\{1, 2, 3, 4\}$ such that $\Pr[X = i] = N^{-1}2^{i\lambda}$, where N is the normalizing constant, that is, $N = 2^\lambda + 2^{2\lambda} + 2^{3\lambda} + 2^{4\lambda}$. Show that for each real number $\mu \in (1, 4)$, there is a choice $\lambda(\mu)$ for λ such that $\mathbb{E}[X_\lambda(\mu)] = \mu$. [6]

(c) Assume the notation of the previous part. Suppose Y is random variable taking values in $\{1, 2, 3, 4\}$ such that $\mathbb{E}[Y] = 3$. Show that [8]

$$H[Y] \leq H[X_{\lambda(3)}],$$

for $\lambda(3)$ as defined in part (b). Hint: Consider $D(P_Y \| P_{X_{\lambda(3)}})$.

3. Consider the following memoryless channel W with input alphabet $A = \{0, 1\}$ and output alphabet $B = \{0, 1, \lambda\}$: $W(0|0) = \frac{3}{4}$, $W(\lambda|0) = \frac{1}{4}$, $W(1|1) = \frac{2}{3}$, $W(\lambda|1) = \frac{1}{3}$. Let $P = \mathsf{Ber}(\frac{1}{4})$, and for n divisible by 4, let $\mathcal{T}_P^n \subseteq \{0, 1\}^n$ be the set of strings of type P, that is, strings with exactly $3n/4$ zeros and $n/4$ ones. We say that a subset $C \subseteq \{0, 1\}^n$ is ϵ-decodable if there is a decoder $g : B^n \to A^n$, such that such that for all $\overline{x} \in C$,

$$\Pr[g(\bar{y}) = \bar{x}] \geq 1 - \epsilon,$$

where $\bar{y}$ is the received word when $\bar{x}$ is sent over the channel. Let

$$\widetilde{\mathsf{cap}} = \limsup_{n \to \infty} \max_{C} \frac{1}{n} \log |C|,$$

where n ranges over multiples of 4, and the inner maximum is taken over all $(\frac{1}{n})$-decodable codes $C \subseteq \mathcal{T}_P^n$. For parts (a) and (b), write your answers using the binary entropy function $h_2(p) = -p \log p - (1 - p) \log(1 - p)$.

(a) Suppose the random bit $X \sim \mathsf{Ber}(\frac{1}{4})$ (that is, $\Pr[X = 1] = \frac{1}{4} = 1 - \Pr[X = 0]$) is sent through the channel and the bit Y is received. What is $H[X|Y]$? [2]

Answer:____

(b) What is the value of $\widetilde{\mathsf{cap}}$? [4]

Answer:____

(c) State briefly why $\widetilde{\mathsf{cap}}$ cannot exceed your answer for part (b). A rough calculation based on estimates of the sizes of *typical sets* and the *almost equipartition property* (or alternatively using Fano's inequality) will be sufficient. [7]

(d) For each n (a multiple of 4), suppose $C_n \subseteq A^n$ is an $\frac{1}{n}$-decodable code for the channel W. Let $\bar{x} \in C_n$. We wish to determine how many received words must necessarily be decoded to $\bar{x}$ in order to ensure that the error probability is at most $\frac{1}{n}$. Consider decoding functions $\tilde{g}_n : B^n \to A^n$ for C_n such that $\Pr[\tilde{g}_n(\bar{y}) = \bar{x}] \geq 1 - \frac{1}{n}$, where $\bar{y}$ is the random received word when $\bar{x}$ is sent through the channel. Determine **[7]**

$$\lim_{n\to\infty} \min_{\tilde{g}_n} \frac{1}{n} \log |\tilde{g}_n^{-1}(\bar{x})|$$

(As before, here n ranges over multiples of 4.)

Answer:____

Briefly justify your answer.

4. Consider the following distortion measure d for $\{0,1\}\times\{a,b\}$: $d(0,a), d(1,b) = \frac{1}{2}$ and $d(0,b), d(1,a) = 1$. Suppose we wish to efficiently compress a sequence of n iid bits each taking values in $\{0,1\}$ with distribution $\mathsf{Ber}(p)$ ($p \leq \frac{1}{2}$), and then uncompress it as a sequence in $\{a,b\}^n$ such that the distortion is at most D. Let $R(D)$ be the rate-distortion function for the above distortion measure d. [Recall that this (informally) corresponds to encoding the source strings $\bar{x} = (x_1, x_2, \ldots, x_n)$ as a number $i_{\bar{x}} \in \{1, 2, \ldots, 2^{Rn}\}$ and then decoding $i_{\bar{x}}$ as a string $\bar{y} = (y_1, y_2, \ldots, y_n) \in \{a,b\}^n$ such that the average distortion between x_i and y_i is asymptotically not much more than D.]

(a) Argue directly that $R(\frac{1+p}{2}) = 0$; what are the corresponding encoding and decoding functions? **[5]**

(b) Argue directly that $R(\frac{1}{2}) = h_2(p)$. 5

(c) In general, what is $R(D)$ as a function of p and D? You may use the following fact derived in class for the distortion function d': $d'(0,a), d'(1,b) = 0$, $d'(0,b), d'(1,a) = 1$. 10

$$R_{d'}(D) = \begin{cases} h(p) - h(D) & 0 \leq D \leq \min\{p, 1-p\}; \\ 0 & \text{otherwise.} \end{cases}$$

5. Recall that for a continuous real-valued random variable X with density $f(x)$, its differential entropy is given by

$$h[X] = -\int_0^\infty f(x) \log_2 f(x) \,\mathrm{d}x.$$

What is the minimum that $h[X]$ can be if X has variance 20? Justify your answer. 5

6. Let P and Q be distributions on a finite set A. Let $S \subseteq A^n$ be such that $P^n(A) = \frac{1}{2}$. What is 5

$$\lim_{n\to\infty} \frac{1}{n} \log Q^n(S)?$$

(You just need to write the expression in terms of P and Q.)

7. Recall that the capacity of a channel with input alphabet A, output alphabet B and characteristics represented as a matrix of conditional probabilities $W = (W(b|a) : a \in A, b \in B)$ is given by 10

$$\mathsf{cap}(W) = \max_P \sum_a P(a) D(W_a \| PW),$$

where we maximize over all probability distributions on A (regarding them as row vectors indexed by $a \in A$); W_a is a-th row of the matrix W whose b-th entry is $W(b|a)$, and PW represents a probability distribution on B (again represented as a row vector). Show that

$$\mathsf{cap}(W) = \max_P \max_{P'} \left[\left(\sum_{a\in A} P(a) D(W_a \| P'W) \right) - D(P\|P') \right].$$