Problem sheet 3, Information Theory, HT 2022

Designed for the third tutorial class

Question 1 For a random variable X with state space $X = \{x_1, \cdots, x_7\}$ and distribution $p_i = \mathbb{P}(X = x_i)$ given by

p_1	p_2	p_3	p_4	p_5	p_6	p_7
0.49	0.26	0.12	0.04	0.04	0.03	0.02

(a) Find a binary Huffman code for X and its expected length.

(b) Find a ternary Huffman code for X and its expected length.

Question 2 (a) Prove that the Shannon's code is a prefix code and calculate bounds on its expected length. Give an example to demonstrate that it is not an optimal code.

(b) Prove that the Elias code is a prefix code and calculate bounds on its expected length. Is it an optimal code?
`Hint:` Suppose $\mathcal{Y} = \{0, 1, \cdots, d\}$. For any $i = 1, \cdots, |\mathcal{X}|$, suppose $c(x_i) = a_1 \cdots a_k$ with $k = |c(x)|$. Denote $v_i = \sum_{j=1}^{|c(x_i)|} a_j d^{-j}$, $r(i) = \sum_{j=1}^{i-1} p_j + p_i/2$ and $\hat{r}(i) = r(i) + p_i/2$. Try to show that the interval $[v_i, v_i + d^{-|c(x_i)|})$ is contained in the interval $[\hat{r}_{i-1}, \hat{r}_i)$. Hence the intervals $[v_i, v_i + d^{-|c(x_i)|})$ are disjoint to each other.

Question 3 Prove the following weaker version of the Kraft-McMillan theorem (called Krafts theorem) using rooted trees

(a) Let $c : \mathcal{X} \mapsto \{0, \cdots, d-1\}^*$ be a prefix code. Consider its code-tree and argue that $\sum_{x \in \mathcal{X}} d^{-|c(x)|} \leq 1$. [Note that the assumption that c is a prefix code is crucial here, otherwise the code-tree cannot be defined to begin with. In the Kraft-McMillan theorem from the lecture we only require c to be uniquely decodable].

(b) Assume that $\sum_{x \in \mathcal{X}} d^{-l_x} \leq 1$ with $l_x \in \mathbb{N}$. Show that there exists a prefix code c with codeword lengths $|c(x)| = l_x$ for $x \in \mathcal{X}$ by constructing a rooted tree.

Question 4 Give yet another proof for $\sum_{x\in\mathcal{X}} d^{-|c(x)|} \leq 1$ if c is a prefix code by using the "probabilistic method": randomly generate elements of $\{0,\cdots,d-1\}^*$ by sampling i.i.d. from $\{0,\cdots,d-1\}$ and consider the probability of writing a codeword of c.

Question 5 Let X be uniformly distributed over a finite set $\mathcal{X}$ with $|\mathcal{X}| = 2^n$ for some $n \in \mathbb{N}$. Given a sequence $A_1, A_2, \cdots$ of subsets of $\mathcal{X}$ we ask a sequence of questions of the form $X \in A_1, X \in A_2$, etc.

(a) We can choose the sequence of subsets. How many such questions do we need to determine the value of X? What is the most efficient way to do so?
`[Note: If we regard all questions as a mapping from` $\mathcal{X}$ `to` $\{Yes, No\}^*$`, we can even think about how to design the sequence of subsets to minimise the expected number of questions to ask to get the value of a random variable` X `with any given distribution.]`

(b) We now randomly (i.i.d. and uniform) draw a sequence of sets $A_1, A_2, \cdots$ from the set of all subset of $\mathcal{X}$. Fix $x, y \in \mathcal{X}$. Conditional on $\{X = x\}$:

i. What is the probability that x and y are indistinguishable after the first k random questions?

ii. What is the expected number of elements in $\mathcal{X}\backslash\{x\}$ that are indistinguishable from x after the first k questions?

Question 6 Let $|\mathcal{X}| = 100$ and p the uniform distribution on $\mathcal{X}$. How many codewords are there of length $l = 1, 2, \cdots$ in an Huffman binary code?

Question 7 (Optional) Let X be a Bernoulli random variable with $\mathbb{P}(X = 0) = 0.995, \mathbb{P}(X = 1) = 0.005$ and consider a sequence $X_1, \cdots, X_{100}$ consisting of i.i.d. copies of X. We study a block code of the form $c : \{0,1\}^{100} \mapsto \{0,1\}^m$ for a fixed $m \in \mathbb{N}$.

(a) What is the minimal m such that there exists c such that its restriction to sequences $\{0,1\}^{100}$ that contain three or fewer 1s is injective?

(b) What is the probability of observing a sequence that contains four or more 1s? Compare the bound given by the Chebyshev inequality with the actual probabiltiy of this event.

x=	1	2	3	4
p=	0.5	0.25	0.125	0.125
c=	0	10	110	111

Table 1: Data for Question 8

.

Question 8 (Optional) Let X be a $\mathcal{X} = \{1, 2, 3, 4\}$-valued random variable with pmf p and binary code c as in the Table 3.

For $n \in \mathbb{N}$, we generate a sequence in $\mathcal{X}^n$ by sampling i.i.d. from the distribution p. We then pick one bit uniformly at random from the binary encoded sequence. What is the asymptotic (as $n \to +\infty$) probability that this bit equals 1?