Information Theory MT18 **Sheet 3**

1. For a rv X with state space $\mathcal{X} = \{x_1, \ldots, x_7\}$ 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.

2. Let X be a Bernoulli rv with $\mathbb{P}(X = 0) = 0.995$, $\mathbb{P}(X = 1) = 0.005$ and consider a sequences $X_1, \ldots, X_{100}$ consisting of iid copies of X. We study a block code of the form $c : \{0, 1\}^{100} \to \{0, 1\}^m$ for a fixed $m \in \mathbb{N}$.

   (a) What is the minimal m if we just require that the restriction of c to sequences $\{0, 1\}^{100}$ that contain three or fewer 1's is injective?

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

3. Prove that

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

   (b) the Elias code is an prefix code and calculate bounds on its expected length. Give an example to demonstrate that it is not an optimal code.

4. Prove a weaker version of the Kraft–McMillian theorem (called Kraft's theorem) using rooted trees:

   (a) Let $c : \mathcal{X} \to \mathcal{Y}^\star$ be a prefix code. Consider its codetree and if $\ell_{\max}$ denotes the length of the longest codeword, argue that $\sum |\mathcal{Y}|^{\ell_{\max} - |c(x)|} \leq |\mathcal{Y}|^{\ell_{\max}}$, hence $\sum |\mathcal{Y}|^{-|c(x)|} \leq 1$. follows [Note: the assumption of a prefix code is crucial here, in the Kraft–McMillan theorem as presented in the lecture we only require c to be uniquely decodebable].

   (b) Assume $\sum_{x \in \mathcal{X}} |\mathcal{Y}|^{-\ell_x} \leq 1$ with $\ell_x \in \mathbb{N}$. Show there exists a prefix code with codeword lengths $(\ell_x)_{\in \mathcal{X}}$ by constructing a rooted tree.

5. Give yet another proof for $\sum_x |\mathcal{Y}|^{-|c(x)|} \leq 1$ if c is a prefix code by using the "probabilistic method": randomly generate elements of $\mathcal{Y}^\star$ by sampling iid from $\mathcal{Y}$ and consider the probability of writing a codeword of c.

6. Let X be uniformly distributed over a finite set $\mathcal{X}$, $|\mathcal{X}| = 2^n$ for some $n \in \mathbb{N}$. Given a sequence $A_1, A_2, \ldots$ 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?

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

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

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

7. Let $\mathcal{X}$ be a finite set and X a $\mathcal{X}$-valued random variable with pmf p. Let $r : \mathcal{X} \to (0, \infty)$.

(a) Find non-negative numbers $(\ell_x)_{x\in\mathcal{X}}$ that minimize $\sum_x p(x)\, r(x)\, \ell_x$ and such that $\sum_{x\in\mathcal{X}} 2^{-\ell_x} = 1$. Calculate this minimum and denote it by $L^\star$.
[Hint: consider $q_1(x) = \frac{p(x)r(x)}{\sum p(x)r(x)}$ and $q_2(x) = 2^{-\ell_x}$, and the divergence between q_1 and q_2.]

(b) Describe how a Huffman code construction be used to find a prefix code $c : \mathcal{X} \longrightarrow \{0,1\}^\star$ such that

$$L^\star \le \mathbb{E}[|c(X)|\, r(X)] < L^\star + \mathbb{E}[|r(X)|].$$

8. Let X be a $\mathcal{X} = \{1,2,3,4\}$-valued rv with pmf $p(1) = 0.5$, $p(2) = 0.25$, $p(3) = 0.125$, $p(4) = 0.125$ and a code $c(1) = 0$, $c(2) = 10$, $c(3) = 110$, $c(4) = 111$. We generate a sequence in $\mathcal{X}^n$ by sampling iid from p. We then pick one bit uniformly at random from the binary encoded sequence. What is the probability that this bit equals 1?