1. For a r.v. X with state space $\mathcal{X} = \{x_1, \dots, x_7\}$ 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. Prove that

   (a) the Shannon 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) the Elias code is a prefix code and calculate bounds on its expected length. Is it an optimal code?

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

   (a) Let $c : \mathcal{X} \to \{0, \dots, d-1\}^\star$ be a prefix code. Consider its code-tree and argue that $\sum d^{-|c(x)|} \leq 1$. [Note that the assumption prefix code is crucial here or 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 $\sum_{x \in \mathcal{X}} d^{-\ell_x} \leq 1$ 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.

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

5. Let X be uniformly distributed over a finite set $\mathcal{X}$, $|\mathcal{X}| = 2^n$ for some $n \in \mathbb{N}$. Given a sequence $A_1, A_2, \dots$ 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 (i.i.d. and uniform) draw a sequence of sets $A_1, A_2, \dots$ 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?

6. Let $|\mathcal{X}| = 100$ and p the uniform distribution on $\mathcal{X}$. How many codewords are there of length $l = 1, 2, \dots$ in an optimal binary code? here are 28 codewords of length 6 and 72 of length 7. To see it, you can do the Huffman procedure by hand, or notice the following: since the distribution is uniform, the leaves of the Huffman tree can only occupy at most 2 levels. Since $2^6 = 64 < 100 < 128 = 2^7$, these two levels must be 6 and 7. Call x the number of leaves at the level 6, then the other level 6 nodes (of which there are $64 - x$) have two branches and so the total number of leaves is $100 = x + 2(64 - x)$, giving $x = 28$.

7. * (**Optional**) Let X be a Bernoulli r.v. with $\mathbb{P}(X = 0) = 0.995$, $\mathbb{P}(X = 1) = 0.005$ and consider a sequence $X_1, \dots, X_{100}$ consisting of i.i.d. 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 such that there exists c such that its restriction 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.

8. * (**Optional**) Let X be a $\mathcal{X} = \{1, 2, 3, 4\}$-valued r.v. with pmf $p(1) = 0.5$, $p(2) = 0.25$, $p(3) = 0.125$, $p(4) = 0.125$ and a code $c(1) = 0$, $c(2) = 10$, $c(3) = 110$, $c(4) = 111$. For $n \in \mathbb{N}$, we generate a sequence in $\mathcal{X}^n$ by sampling i.i.d. from 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?