IV054 Coding, Cryptography and Cryptographic Protocols
2014 - Exercises I.

1. (a) Prove that for any binary Huffman code, if the most probable message symbol has the probability $p > 2/5$, then that symbol must be assigned a codeword of length 1.

   (b) Prove that for any binary Huffman code, if the most probable message symbol has probability $p < 1/3$, then that symbol must be assigned a codeword of length ≥ 2.

2. The Universal Product Code (UPC) is widely used by supermarkets and mass market retailers for cash register checkout.

   0 12345 67890 5

   The UPC is a 12 digit code. The last digit of the UPC code is a check sum calculated as:

   $$3a_1 + a_2 + 3a_3 + a_4 + 3a_5 + \cdots + 3a_{11} + a_{12} \equiv 0 \pmod{10},$$

   where $a_1 a_2 a_3, \ldots, a_{11}, a_{12}$ is the UPC.

   (a) Does the UPC code detect all single digit errors?

   (b) Does the UPC code detect all adjacent transposition errors?

   Give a proof for your answers.

3. (a) Prove that $A_q(n, d) \leq qA_q(n-1, d)$.

   (b) Prove that $A_q(qn, (q-1)n) \leq q^2 n$.

   Hint: Plotkin Bound:

   $$A_q(n, d) \leq \left\lfloor \frac{qd}{qd - (q-1)n} \right\rfloor$$

4. Compare the upper bounds obtained from the Sphere Packing Bound and the Plotkin Bound (see previous exercise) for $A_2(18, 10)$.

5. Let C be the binary code of blocklength 12 consisting of all sequences in which there are at least three 0s between any two 1s. Find the code rate of C.

6. Prove the following two important properties of the entropy function

   $$H(p_1, \ldots, p_n) = -\sum_{i=1}^{n} p_i \log p_i :$$

   (a) $H(p_1, \ldots, p_n) = H(p_1 + p_2, p_3, \ldots, p_n) + (p_1 + p_2) H(\frac{p_1}{p_1+p_2}, \frac{p_2}{p_1+p_2})$

   (b) $H(p_1, \ldots, p_m, q_1, \ldots, q_n) = H(p, q) + pH(\frac{p_1}{p}, \ldots, \frac{p_m}{p}) + qH(\frac{q_1}{q}, \ldots, \frac{q_n}{q})$, where $p = \sum_{i=1}^{m} p_i$ and $q = \sum_{i=1}^{n} q_i$.

7. Consider the q-ary Huffman code for the source with the following relative frequencies of n symbols: $1, q, q^2, q^3, \ldots, q^{n-1}$, where $n = 1 + k(q-1)$ for some positive integer k.
   Find the number of symbols required to encode the most and the least frequent symbol.