1999

p 84 11

JGD

Model Answer, Information Theory and Coding, Question 1.

1. The entropy of the source alphabet is

$$H = -\sum_{i=1}^{4} p_i \log_2 p_i = (1/4)(2) + (1/8)(3) + (1/2)(1) + (1/8)(3)$$

$= \underline{1.75 \text{ bits}}$. [4 marks]

2. Fixed length codes are inefficient for alphabets whose letters are not equiprobable because the cost of coding improbable letters is the same as that of coding more probable ones. It is more efficient to allocate fewer bits to coding the more probable letters, and to make up for the fact that this would cover only a few letters, by making longer codes for the less probable letters. This is exploited in Morse Code, in which (for example) the most probable English letter, e, is coded by a single dot. [4 marks]

3. A uniquely decodable prefix code for the letters of this alphabet:
Code for A: 10
Code for B: 110
Code for C: 0
Code for D: 111 (the codes for B and D could also be interchanged)

This is a uniquely decodable prefix code because even though it has variable length, each code corresponds to a unique letter rather than any possible combination of letters; and the code for no letter could be confused as the prefix for another letter. [4 marks]

4. Multiplying the bit length of the code for each letter times the probability of occurence of that letter, and summing this over all letters, gives us a coding rate of:
$R = (2 \text{ bits})(1/4)+(3 \text{ bits})(1/8)+(1 \text{ bit})(1/2)+(3 \text{ bits})(1/8) = \underline{1.75 \text{ bits}}$.

This code is optimally efficient because $R = H$: its coding rate equals the entropy of the source alphabet. Shannon's Source Coding Theorem tells us that this is the lower bound for the coding rate of all possible codes for this alphabet. [4 marks]

5. The maximum possible entropy of an alphabet consisting of N different letters is $H = \log_2 N$. This is only achieved if the probability of every letter is $1/N$. Thus $1/N$ is the probability of both the "most likely" and the "least likely" letter. [4 marks]