Albert-Ludwigs-Universität Freiburg
Physikalisches Institut

Johannes Berg

Information Theory: From Statistical Physics to Quantitative Biology

3. exercise class – 19. November 2008

1. Coding in the presence of costs

Frequent occurrence is one reason linked with *short code* words in natural languages ("the", "a", "is"). However, in the case of words like "help!", "fire!", etc., the reason for their shortness is not that these words are frequently used. Instead, in the cases where they are used, time is precious.

Consider a random variable $X = i$ with probabilities p_i, $i = 1, \ldots, m$. Let l_i be the number of binary symbols in the codeword for i, and let c_i be the *cost* per binary letter of the codeword. The average cost C of the code is $\sum_i p_i c_i l_i$.

Minimize C over all l_i subject to the constraint given by the Kraft inequality. Compute the minimum value of C. (Constraints due to the integer nature of l_i can be ignored.)

2. Lempel-Ziv coding

Give the Lempel-Ziv coding of 000000110101000001101 01.