MA3466 Tutorial Sheet 7, q1-2 solutions[1]

6 April 2010

1. (C& T 5.2) Let $p(S_i) = p_i$ for some set of outcomes $\{S_1, S_2, \ldots, S_n\}$. The S_i's are uniquely encoded into strings from a D-symbol alphabet in a uniquely decodable manner. If $n = 6$ and the code word lengths are $(l_1, l_2, \ldots, l_n) = (1, 1, 2, 3, 2, 3)$ find a good lower bound on D.

   *Solution:*Well this question is easy to do by checking the first few cases. $D = 2$ clearly doesn't work, since $l(S_1) = l(S_2) = 1$ then we need to have $c(S_1) = 0$ and $c(S_2) = 1$ or visa versa, and then the length two code word for S_3, which needs to be 00, 01, 10 or 11, is ambiguous, since it could also be decoded as a pair from S_1 and S_2. However, $D = 3$ does work, for example

S_1	S_2	S_3	S_4	S_5	S_6
0	1	20	220	21	221

   is a prefix code with the correct lengths.

2. (C& T 5.4) Slackness in the Kraft inequality. An instantaneous code has word lengths l_1 to l_m satisfying the strict inequality

   $$\sum_{i=1}^{m} D^{-l_i} < 1 \tag{1}$$

   Show there are arbitrarily long sequences of code symbols in $\mathcal{D}^*$ which cannot be decoded into sequences of codewords: that is, not all sequences of symbols in D form a sentence.

   *Solution:*So the trick here is that, if

   $$\sum_{i=1}^{m} D^{-l_i} < 1 \tag{2}$$

   then, with l_x with maximum length,

   $$\sum_{i=1}^{m} D^{l_x - l_i} < D^{l_x} \tag{3}$$

   and, we know, that D^{l_x} is the number of nodes at level l_x, which the sum on the right hand side gives the number of nodes at that level which are codewords or their

[1]Conor Houghton, `houghton@maths.tcd.ie`, see also `http://www.maths.tcd.ie/~houghton/MA3466`

descendents. Thus, the strict inequality means that there are nodes at level l_x which are not codewords and which are not the descendents of codewords. Taking one these nodes: the node will correspond to a codeword and could of been added to the existing prefix code to extend it to a further symbol, this code would still be uniquely decodable. This means that strings in the codeword corresponding to this node do not correspond to any of the existing symbols.