Appendix B

Choice of Codeword Lengths in a Depth-constrained Alphabetic Tree

Lemma 1 A depth-constrained alphabetic tree with maximum depth D satisfies the characteristic inequality of Lemma 3.1 (Chapter 3), when the codeword lengths l_k of the k^{th} letter occurring with probability $q_k \left(q_k \geq 2^{-D} \forall k \right)$ are given by: $l_k^* = \begin{cases} min(\lceil -\log q_k \rceil, D) & k = 1, n \\ min(\lceil -\log q_k \rceil + 1, D) & 2 \leq k \leq n-1 \end{cases}$.

Proof: We need to prove that $s_n \leq 1$ where $s_k = c(s_{k-1}, 2^{-l_k}) + 2^{-l_k}$, $s_0 = 0$, and c is defined by $c(a,b) = \lceil a/b \rceil b$. We first prove by induction that

$$s_i \leq \sum_{k=1}^{i} q_k \qquad \forall 1 \leq i \leq n-1$$

For the base case, $s_1 = 2^{-l_1} \leq q_1$ by the definition of l_1. For the induction step, assume the hypothesis is true for $i-1$. By definition, $s_i = c(s_{i-1}, 2^{-l_i}) + 2^{-l_i}$. Now there are two possible cases:

1. $\lceil -\log q_i \rceil + 1 \le D$, and therefore $2^{-(l_i - 1)} \le q_i$. Using the fact that $\lceil a/b \rceil < a/b + 1$, i.e., $c(a, b) < a + b$ for all nonzero real numbers a and b, we get the following using inductive hypothesis:

$$s_i \le s_{i-1} + 2^{-l_i} + 2^{-l_i} \le 2^{-(l_i - 1)} + \sum_{k=1}^{i-1} q_k \le q_i + \sum_{k=1}^{i-1} q_k = \sum_{k=1}^{i} q_k$$

2. $\lceil -\log q_i \rceil + 1 > D$. This implies that $l_i = D$ and hence $q_i \ge 2^{-l_i}$. Also, as s_j is an integral multiple of $2^{-D} \forall j$, $c(s_{i-1}, 2^{-l_i}) = s_{i-1} \le \sum_{k=1}^{i-1} q_k$ and thus:

$$s_i = 2^{-l_i} + c(s_{i-1}, 2^{-l_i}) \le q_i + \sum_{k=1}^{i-1} q_k = \sum_{k=1}^{i} q_k$$

Therefore, $s_{n-1} \le \sum_{i=1}^{n-1} q_i = 1 - q_n \le 1 - 2^{-l_n}$. Also:

$s_n = 2^{-l_n} + c(s_{n-1}, 2^{-l_n}) \le 2^{-l_n} + c(1 - 2^{-l_n}, 2^{-l_n}) = 2^{-l_n} + 1 - 2^{-l_n} = 1$. This completes the proof that these codeword lengths satisfy the characteristic inequality.❑