2002 Paper 13 Question 1

Data Structures and Algorithms

Arithmetic encoding compactly represents a string of characters by an enormously precise number in the range [0,1) represented in binary by a finite sequence of digits following the decimal point. What is remarkable is that this number can be processed efficiently using only fixed point arithmetic on reasonably small integers. As a demonstration, if the original text contained only the characters `A`, `B`, `C` and the end-of-file mark `w`, such text can be arithmetically encoded using only 3-bit arithmetic. Illustrate how it can be done by decoding the string `101101000010` on the assumption that the character frequencies are such that the decoding tables of size 8 and 6 are, respectively, `wAABBCCC` and `wABBCC`. The first few lines of your working could be as follows:

```
                    0   0   0   0   1   1   1   1
                    0   0   1   1   0   0   1   1
                    0   1   0   1   0   1   0   1
101 101000010     |-w---A---A---B---B-+(C)++C+++C+|       => C
```

Your answer should include a brief description of how the decoding algorithm works.

[20 marks]