Arithmetic Coding: Example 1

Suppose we have an alphabet $\{1, 2\}$ and a sequence $X_1X_2X_3X_4$ of four independent variables drawn from the alphabet. Assume $p(X_j = 1) = \frac{3}{4}$ and $p(X_j = 2) = \frac{1}{4}$ for all j. That is, we assume a stationary, zeroth order Markov model. For this case, the arithmetic coding induction equations are:

$$\begin{aligned} l_{k+1} &= l_k + (u_k - l_k)\ F(pred(i_{k+1})) \\ u_{k+1} &= l_k + (u_k - l_k)\ F(i_{i+1}) \end{aligned}$$

Question: What is the codeword of the sequence $(1, 2, 1, 2)$

Answer: Since $n = 4$, we first need to calculate l_4 and u_4 by applying the induction equations.

k	X_k	l_k	u_k
1	1	0	$\frac{3}{4}$
2	2	$\frac{9}{16}$	$\frac{12}{16}$
3	1	$\frac{36}{64}$	$\frac{45}{64}$
4	2	$\frac{171}{256}$	$\frac{180}{256}$

Next we compute the tag:

$$T(\vec{x}) = \frac{l_4 + u_4}{2} = (\frac{171}{256} + \frac{180}{256})/2 = \frac{351}{512}$$

which in binary is .101011111, that is

$$\frac{351}{512} = \frac{1}{512}(256 + 64 + 16 + 8 + 4 + 2 + 1) = \frac{1}{2} + \frac{1}{8} + \frac{1}{32} + \frac{1}{64} + \frac{1}{128} + \frac{1}{256} + \frac{1}{512}$$

[For those of you who do not know how to quickly convert an decimal integer to binary, see by COMP 273 lecture notes (lecture 1 p. 1,2, and lecture 2 p. 1,2).]

How many of these bits do we use for the code?

$$p(\vec{x}) = p(1, 2, 1, 2) = u_4 - l_4 = \frac{9}{256}$$

thus, by an earlier definition of the arithmetic code, we use

$$\lambda(\vec{x}) = \lceil \log \frac{2}{p(\vec{x})} \rceil = \lceil \log \frac{512}{9} \rceil = \lceil \log 56.* \rceil = 6 \text{ bits}$$

Hence,

$$C(\vec{x}) = 101011$$

Decoder

Let's now consider how the decoder decodes this sequence. (I did not do it in class, but I'll cover it here for completeness.)

Suppose the decoder is given that $n = 4$ and it is given that the alphabet is $\{1, 2\}$ and that the probabilities are as previously stated. The decoder needs to infer what is the sequence $\vec{x} = (i_1, i_2, i_3, i_4)$ such that $C(\vec{x}) = 101011$.

Since $n = 4$, the decoder needs to calculate l_k and u_k for each l_k, and figure out which symbol (1 or 2) is consistent with $C(\vec{x}) = 101011$. If we write .101011 as a fraction, we get $\frac{43}{64}$. So we need to choose an l_k and u_k sequence such that

$$l_4 \leq \frac{43}{64} < u_4$$

k	X_k	l_k	u_k	
1	1	0	$\frac{48}{64}$	since $0 < \frac{43}{64} < \frac{48}{64}$
2	2	$\frac{36}{64}$	$\frac{48}{64}$	since $\frac{36}{64} < \frac{43}{64} < \frac{48}{64}$
3	1	$\frac{36}{64}$	$\frac{45}{64}$	since $\frac{36}{64} < \frac{43}{64} < \frac{45}{64}$
4	2	$\frac{171}{256}$	$\frac{180}{256}$	since $\frac{171}{256} < \frac{172}{256}(\textit{ i.e. } \frac{43}{64}) < \frac{180}{256}$