Arithmetic coding of Markov source, example

Problem

A Markov source X_i of order 2 with alphabet $\mathcal{A} = \{1, \ 2\}$ is given by the transition probabilities $p(x_i|x_{i-1}, x_{i-2})$ below (note the order).

$$
\begin{aligned}
p(1|1,1) &= 0.95 & p(2|1,1) &= 0.05 \\
p(1|1,2) &= 0.9 & p(2|1,2) &= 0.1 \\
p(1|2,1) &= 0.2 & p(2|2,1) &= 0.8 \\
p(1|2,2) &= 0.3 & p(2|2,2) &= 0.7
\end{aligned}
$$

Code the sequence

$$122111$$

using arithmetic coding. The memory of the source should be utilized. You can assume that the source is in state (1,1) when the coding starts. You can also assume that the coder can store all probabilities and interval limits exactly. Give both the resulting interval and the codeword.

Solution

We will have four cumulative distribution functions $F(x_i|x_{i-1}, x_{i-2})$, given by

$$
\begin{aligned}
F(0|1,1) &= 0 & F(1|1,1) &= 0.95 & F(2|1,1) &= 1 \\
F(0|1,2) &= 0 & F(1|1,2) &= 0.9 & F(2|1,2) &= 1 \\
F(0|2,1) &= 0 & F(1|2,1) &= 0.2 & F(2|2,1) &= 1 \\
F(0|2,2) &= 0 & F(1|2,2) &= 0.3 & F(2|2,2) &= 1
\end{aligned}
$$

Starting interval

$$
\begin{aligned}
l^{(0)} &= 0 \\
u^{(0)} &= 1
\end{aligned}
$$

First symbol is 1, state is (1,1).

$$
\begin{aligned}
l^{(1)} &= 0 + (1-0) \cdot F(0|1,1) = 0 \\
u^{(1)} &= 0 + (1-0) \cdot F(1|1,1) = 0.95
\end{aligned}
$$

Second symbol is 2, state is (1,1).

$$
\begin{aligned}
l^{(2)} &= 0 + (0.95-0) \cdot F(1|1,1) = 0.9025 \\
u^{(2)} &= 0 + (0.95-0) \cdot F(2|1,1) = 0.95
\end{aligned}
$$

Third symbol is 2, state is (2,1).

$$l^{(3)} = 0.9025 + (0.95 - 0.9025) \cdot F(1|2, 1) = 0.912$$
$$u^{(3)} = 0.9025 + (0.95 - 0.9025) \cdot F(2|2, 1) = 0.95$$

Fourth symbol is 1, state is (2,2).

$$l^{(4)} = 0.912 + (0.95 - 0.912) \cdot F(0|2, 2) = 0.912$$
$$u^{(4)} = 0.912 + (0.95 - 0.912) \cdot F(1|2, 2) = 0.9234$$

Fifth symbol is 1, state is (1,2).

$$l^{(5)} = 0.912 + (0.9234 - 0.912) \cdot F(0|1, 2) = 0.912$$
$$u^{(5)} = 0.912 + (0.9234 - 0.912) \cdot F(1|1, 2) = 0.92226$$

Sixth symbol is 1, state is (1,1).

$$l^{(6)} = 0.912 + (0.92226 - 0.912) \cdot F(0|1, 1) = 0.912$$
$$u^{(6)} = 0.912 + (0.92226 - 0.912) \cdot F(1|1, 1) = 0.921747$$

The interval corresponding to the sequence is [0.912 0.921747) with interval size 0.009747. We will thus need at least $\lceil -\log_2 0.009747 \rceil = 7$ bits in our codeword, maybe one more.

Write the two interval limits as binary numbers:

$$0.912 = 0.11101001011110\ldots$$
$$0.921747 = 0.11101011111101\ldots$$

From this it is clear that there are no seven bit numbers such that all numbers starting with these seven bits are inside the interval. Thus, we will have to use eight bits in our codeword. Our codeword becomes **11101010**