Arithmetic coding, lecture example

Coding

Alphabet $\mathcal{A} = \{1, 2, 3\}$. Let $m = k = 6$. Cumulative distribution function

$$F(0) = 0,\ F(1) = 38,\ F(2) = 58,\ F(3) = 64$$

Code the sequence 1,3,2,1

$$\begin{aligned} l^{(0)} &= 0 = (000000)_2 \\ u^{(0)} &= 63 = (111111)_2 \end{aligned}$$

$$\begin{aligned} l^{(1)} &= 0 + \lfloor \frac{(63-0+1)\cdot 0}{64} \rfloor = 0 = (000000)_2 \\ u^{(1)} &= 0 + \lfloor \frac{(63-0+1)\cdot 38}{64} \rfloor - 1 = 37 = (100101)_2 \\ l^{(2)} &= 0 + \lfloor \frac{(37-0+1)\cdot 58}{64} \rfloor = 34 = (100010)_2 \\ u^{(2)} &= 0 + \lfloor \frac{(37-0+1)\cdot 64}{64} \rfloor - 1 = 37 = (100101)_2 \end{aligned}$$

The first bit is the same in l and u. Shift out 1 to the codeword, shift a 0 into l and a 1 into u. The codeword so far is 1.

$$\begin{aligned} l^{(2)} &= (000100)_2 = 4 \\ u^{(2)} &= (001011)_2 = 11 \end{aligned}$$

The first bit is the same in l and u. Shift out 0 to the codeword, shift a 0 into l and a 1 into u. The codeword so far is 10.

$$\begin{aligned} l^{(2)} &= (001000)_2 = 8 \\ u^{(2)} &= (010111)_2 = 23 \end{aligned}$$

The first bit is the same in l and u. Shift out 0 to the codeword, shift a 0 into l and a 1 into u. The codeword so far is 100.

$$\begin{aligned} l^{(2)} &= (010000)_2 = 16 \\ u^{(2)} &= (101111)_2 = 47 \end{aligned}$$

Case 3. Shift both l and u, but don't put any bits in the codeword yet. Shift a 0 into l and a 1 into u, and invert the new most significant bit in both l and u.

$$\begin{aligned} l^{(2)} &= (000000)_2 = 0 \\ u^{(2)} &= (111111)_2 = 63 \end{aligned}$$

$$\begin{aligned} l^{(3)} &= 0 + \lfloor \frac{(63-0+1)\cdot 38}{64} \rfloor = 38 = (100110)_2 \\ u^{(3)} &= 0 + \lfloor \frac{(63-12+1)\cdot 58}{64} \rfloor - 1 = 57 = (111001)_2 \end{aligned}$$

The first bit is the same in l and u. Shift out 1 plus an extra 0 from the previous case 3 shift to the codeword, shift a 0 into l and a 1 into u. The codeword so far is 10010.

$$\begin{aligned} l^{(3)} &= (001100)_2 = 12 \\ u^{(3)} &= (110011)_2 = 51 \end{aligned}$$

$$\begin{aligned} l^{(4)} &= 12 + \lfloor \frac{(51-12+1)\cdot 0}{64} \rfloor = 12 = (001100)_2 \\ u^{(4)} &= 12 + \lfloor \frac{(51-12+1)\cdot 38}{64} \rfloor - 1 = 34 = (100010)_2 \end{aligned}$$

Since there are no more symbols we don't need to do any more shift operations. The codeword is the bits that have been shifted out before, plus all of $l^{(4)}$, ie **10010001100**.

Decoding

The decoder has to know the precision ($m = k = 6$), the number of symbols coded ($n = 4$) and the cumulative distribution function:

$$F(0) = 0, \; F(1) = 38, \; F(2) = 58, \; F(3) = 64$$

Decode the first codeword in the bitstream starting 100100011001010...
For the given F, the interval 0-37 belongs to symbol 1, the interval 38-57 to symbol 2 and the interval 58-63 to symbol 3. Start the tag t as the first six bits from the bitstream.

$$\begin{aligned} l^{(0)} &= (000000)_2 = 0 \\ u^{(0)} &= (111111)_2 = 63 \\ t &= (100100)_2 = 36 \end{aligned}$$

$$\lfloor \frac{(36-0+1)\cdot 64 - 1}{63-0+1} \rfloor = 36 \Rightarrow \text{the first symbol is } 1$$

$$\begin{aligned} l^{(1)} &= 0 + \lfloor \frac{(63-0+1)\cdot 0}{64} \rfloor = 0 = (000000)_2 \\ u^{(1)} &= 0 + \lfloor \frac{(63-0+1)\cdot 38}{64} \rfloor - 1 = 37 = (100101)_2 \end{aligned}$$

$$\lfloor\frac{(36-0+1)\cdot 64-1}{37-0+1}\rfloor = 62 \Rightarrow \text{the second symbol is 3}$$

$$\begin{aligned} l^{(2)} &= 0+\lfloor\frac{(37-0+1)\cdot 58}{64}\rfloor = 34 = (100010)_2 \\ u^{(2)} &= 0+\lfloor\frac{(37-0+1)\cdot 64}{64}\rfloor - 1 = 37 = (100101)_2 \end{aligned}$$

The first three bits are the same in l and u. Shift them out, shift zeros into l, ones into u and three new bits from the bitstream into t.

$$\begin{aligned} l^{(2)} &= (010000)_2 = 16 \\ u^{(2)} &= (101111)_2 = 47 \\ t &= (100011)_2 = 35 \end{aligned}$$

Case 3. Shift l, u and t. 0 into l, 1 into u and a new bit from the bitstream into t. Invert the new most signficant bit in each.

$$\begin{aligned} l^{(2)} &= (000000)_2 = 0 \\ u^{(2)} &= (111111)_2 = 63 \\ t &= (100110)_2 = 38 \end{aligned}$$

$$\lfloor\frac{(38-0+1)\cdot 64-1}{63-0+1}\rfloor = 38 \Rightarrow \text{the third symbol is 2}$$

$$\begin{aligned} l^{(3)} &= 0+\lfloor\frac{(63-0+1)\cdot 38}{64}\rfloor = 38 = (100110)_2 \\ u^{(3)} &= 0+\lfloor\frac{(63-0+1)\cdot 58}{64}\rfloor - 1 = 57 = (111001)_2 \end{aligned}$$

The first bit is the same in l and u. Shift them out, shift 0 into l, 1 into u and a new bit from the bitstream into t.

$$\begin{aligned} l^{(3)} &= (001100)_2 = 12 \\ u^{(3)} &= (110011)_2 = 51 \\ t &= (001100)_2 = 12 \end{aligned}$$

$$\lfloor\frac{(12-12+1)\cdot 64-1}{51-12+1}\rfloor = 1 \Rightarrow \text{the fourth symbol is 1}$$

Since we have now decoded four symbols we don't have to do any more calculations. The decoded sequence is 1,3,2,1 which is exactly what we coded. The rest of the bits in the stream belong the next codeword.