1. Arithmetic Coding

(a) Build the Shannon-Fano-Elias code for triplets of bits, when the probability of 0 is 0.1 and the probability of 1 is 0.9. Find the average code length per source bit, and compare to the entropy of the source.

(b) Perform arithmetic coding of the sequence 10111 with the probabilities above. In each step (for each source bit), who which bits are added to the codeword.

(c) Decode your encoded sequence in (b), and show for each step which bits can be decoded.

(d) Write a (short) Matlab function that encodes a binary vector with given probabilities of 1 and 0 using a binary arithmetic code, and returns the complete codeword. Do not worry about precision issues (assume only short sequences). Test your program on the sequences coded in (a) and (b), and the sequence 01110 with Pr(1) = 0.7 and Pr(0) = 0.3, and the sequence 1110101111 with the same probabilities. Output the codeword for the last three sequences.

(e) Write a Matlab function for decoding a binary sequence representing an arithmetic coded binary source vector into the original vector. The function should get the code vector, the probabilities of 1 and 0, and the length of the source vector as input. Try your function on all the code vectors you obtained in the previous parts of the problem. Also decode the vector 0111101 into 12 bit sequence, where Pr(1) = 0.8. Output the source vector.

(f) We showed that for arithmetic coding over a binary alphabet, we can recursively compute the cumulative distribution function by

\[ F(u_1^{N+1}) = F(u_1^{N}) - P(u_1^{N}) + 1_{u_{N+1}} \cdot P(u_1^{N}, 0) + P(u_1^{N+1}), \]

where \(1_{u_{N+1}}\) is an indicator function that is 1 if \(u_{N+1} = 1\) and 0 otherwise. Derive a similar recursive relation for arithmetic coding over a general alphabet \(\{0, 1, \cdots, r - 1\}\).