COMPSCI H311 Spring 2020 - Homework 2

Released: Wednesday 5 February 2020. Due: Tuesday 11 February in class

List your collaborators if any. The writeup you present must be your own work in presentation, and you must acknowledge all sources of aid other than course staff and course material.

Huffman vs. Shannon-Fano Coding

Shannon-Fano coding (slides p.15) employs a divide-and-conquer approach: it splits the symbols into two sets with close to equal frequencies, recursively finds codes for both sets, and prefixes them with 0 or 1. Clearly, for up to three symbols, we get the same result as for Huffman codes.

a) For the case of four symbols, find frequencies for which the difference in average bitlength between Shannon-Fano and Huffman codings is highest.

b) Can you establish a bound on the difference in the general case?