Fundamentals of Information Science: Homework 4

March 12, 2025

Problem 1.
Entropy of a sum. Let X and Y be random variables that take on values $x_1, x_2, ..., x_r$ and $y_1, y_2, ..., y_s$, respectively. Let $Z = X + Y$.

(a) Show that $H(Z|X) = H(Y|X)$. Argue that if X, Y are independent, then $H(Y) \leq H(Z)$ and $H(X) \leq H(Z)$. Thus, the addition of independent random variables adds uncertainty.

(b) Give an example of (necessarily dependent) random variables in which $H(X) > H(Z)$ and $H(Y) > H(Z)$.

(c) Under what conditions does $H(Z) = H(X) + H(Y)$?

Problem 2.
AEP and source coding. A discrete memoryless source emits a sequence of statistically independent binary digits with probabilities $p(1) = 0.01$ and $p(0) = 0.99$. The digits are taken 100 at a time and a binary codeword is provided for every sequence of 100 digits containing three or fewer 1's.

(a) Assuming that all codewords are the same length, find the minimum length required to provide codewords for all sequences with three or fewer 1's.

(b) Calculate the probability of observing a source sequence for which no codeword has been assigned.

(c) Use Chebyshev's inequality (search online if you don't know Chebyshev's inequality) to bound the probability of observing a source sequence for which no codeword has been assigned. Compare this bound with the actual probability computed in part (b).

Problem 3.
Consider messages made up entirely of vowels (A,E, I,O,U). Here's a table of probabilities for each of the vowels:

l	p_l	$log_2(1/p_l)$	$p_l log_2(1/p_l)$
A	0.22	2.18	0.48
E	0.34	1.55	0.53
I	0.17	2.57	0.43
O	0.19	2.40	0.46
U	0.08	3.64	0.29
Totals	1.00	12.34	2.19

(a) After studying Huffman codes, use Huffman's algorithm to construct a variable-length code assuming that each vowel is encoded individually. Draw a diagram of the Huffman tree and give the encoding for each of the vowels.

(b) Using your code from part (a) above, give an expression for the expected length in bits of an encoded message transmitting 100 vowels.

(c) Ben spends all night working on a more complicated encoding algorithm and sends you email claiming that using his code the expected length in bits of an encoded message transmitting 100 vowels is 197 bits. Would you pay good money for his implementation?