Math 5251 Error-correcting codes and finite fields, Fall 2024, Vic Reiner
Midterm exam 1
Due Wednesday October 9 by 11:59pm, via Canvas

Instructions: This is an open book, open notes, open web, take-home exam, but you may *not* collaborate. The instructor is the only human source that you are allowed to consult. You must **clearly indicate** any such sources used, including AI sources such as ChatGPT.

1. (40 points total; 5 points each) **True or False**. Your answers must be justified either by counterexamples or proofs to receive full credit.

(a) There exists a memoryless source $W = \{w_1, w_2, w_3, w_4, w_5\}$ and choice of word probabilities having a binary Huffman encoding $f : W \to \{0,1\}^*$ with codewords of lengths $(2, 3, 4, 5, 6)$.

(b) There exists a prefix ternary encoding $f : W \to \{0,1,2\}^*$ for some memoryless source W with $\#W = 7$ words codewords have lengths

$$(\ell_1, \ell_2, \ell_3, \ell_4, \ell_5, \ell_6, \ell_7) = (1, 1, 2, 2, 3, 3, 5).$$

(c) If one has a uniquely decipherable n-ary encoding $f : W \to \Sigma^*$ of a source W, so $\#\Sigma = n$, and if every codeword $f(w_i)$ has length at most ℓ, then $\#W \leq n^\ell$.

(d) Assume a memoryless source $W = \{w_1, \ldots, w_m\}$ with probabilities $\{p_1, \ldots, p_m\}$ with $0 < p_i < 1$ has some $p_{i_0} \geq \frac{1}{2}$. Then in any binary Huffman encoding $h : W \to \{0,1\}^*$, the length of $h(w_{i_0})$ is 1.

(e) If memoryless source $W = \{w_1, \ldots, w_m\}$ with word probabilities $(p_1, \ldots, p_m)$ has a word $h(w_{i_0})$ of length 1 in one of its binary Huffman encodings $h : W \to \{0,1\}^*$, then $p_{i_0} \geq \frac{1}{2}$.

(f) Assume the encoding $f : W \to \Sigma^*$ has the property that whenever $w_i \neq w_j$ then $f(w_i)$ is not a suffix (final segment) of $f(w_j)$. Then f is uniquely decipherable.

(g) The encoding $f : W = \{A, B, C, D, E\} \to \{0,1\}^*$ sending A, B, C, D, E to $10, 01, 010, 000, 101$ is not uniquely decipherable.

(h) Two binary Huffman encodings $h : W \to \{0,1\}^*$ and $h' : W \to \{0,1\}^*$ of the same source W always have the same average length.

2. Let the memoryless source $W = \{w_1, w_2, \ldots, w_8\}$ have word proba-biities

$$\left(\frac{1}{4}, \frac{1}{4}, \frac{1}{8}, \frac{1}{8}, \frac{1}{16}, \frac{1}{16}, \frac{1}{16}, \frac{1}{16}\right)$$

(a) (10 points) Compute the (binary) entropy $H(W)$. You must explain how it was computed, not just give the final numerical answer.

(b) (10 points) Compute the minimum among all uniquely decipherable binary encodings $f : W \to \{0,1\}^*$ of the average length of the code words $f(w_i)$.

3. (20 points) Let W be a memoryless source with m words, and $h : W \to \{0,1\}^*$ any binary Huffman encoding for W. How many nodes total will there be in the Huffman tree used to compute h, including the root? (For example, when $m = 3$, there are 5 nodes in total.) Your answer should be a function of m, and must be proven to receive full credit.

4. Suppose that we are sending length 5 binary words $w = b_1b_2b_3b_4b_5$ with $b_i \in \{0,1\} = \mathbb{F}_2$ through a noisy binary symmetric channel (BSC) having error probability p for each bit sent.

(a) (5 points) Compute the probability of at least one error during transmission, as a function of p.

Now we choose to send w with two extra parity check bits b_6, b_7 as follows:

$$f(w) = b_1b_2b_3b_4b_5b_6b_7$$

where

$$b_6 := b_1 + b_2,$$
$$b_7 := b_3 + b_4 + b_5.$$

(b) (10 points) Compute the probability of at least one undetected error when w is sent as $f(w)$, again as a function of p.

(c) (5 points) Considering the image of f as a set of codewords $\mathcal{C}$ inside $\{0,1\}^*$ of length 7, what is the (binary) rate of the code $\mathcal{C}$?