Math 567: Introduction to Coding Theory

Homework 5: Due Wednesday, March 27

A **probability distribution** is a set of probabilities $P = \{p_1, \ldots, p_n\}$ with the property that each $p_i \geq 0$ and $\sum_{i=1}^{n} p_i = 1$.

The **entropy** of this probability distribution is defined to be

$$H(P) = \sum_{i=1}^{n} p_i \cdot \log(1/p_i).$$

Problem 1.1. Prove that entropy is a concave function of probability distributions. That is, if $0 \leq \alpha \leq 1$ and P, Q are probability distributions of length n then

$$H(\alpha P + (1-\alpha)Q)) \geq \alpha H(P) + (1-\alpha)H(Q).$$

Problem 1.2. Show that if P is a probability distribution with n elements then $H(P) \leq \log(n)$, with equality only when $p_i = 1/n$ for all i. When is $H(P) = 0$?

Problem 1.3. Let P be a probability distribution with n elements. Show that if $Q = \{q_1, \ldots, q_n\}$ is any other probability distribution with n elements then

$$\sum_{i=1}^{n} p_i \cdot \log(1/p_i) \geq \sum_{i=1}^{n} p_i \log(q_i).$$

Problem 1.4. Let C be a binary prefix code using the alphabet $\Sigma = \{x_1, \ldots, x_m\}$ and for each i let ℓ_i be the length of the codeword for x_i. Let $P = \{p_1, \ldots, p_m\}$ be the probability distribution associated with this code (i.e. p_i is the probability that the symbol x_i occurs). Show that

$$\sum_{i=1}^{m} p_i \ell_i \geq H(P).$$

This means that the expected length of a codeword is greater than the entropy.

Problem 1.5. Prove that there exists a binary prefix code C such that (in the notation of Problem 1.4)

$$\sum_{i=1}^{m} p_i \ell_i \leq H(P) + 1.$$

Problem 1.6. (Constructing Huffman Codes)

1. Construct a binary Huffman code for the following distribution on 5 symbols: $P = \{0.3, 0.3, 0.2, 0.1, 0.1\}$. What is the average length of this code?

2. Construct a probability distribution P' on 5 symbols for which the code you constructed in part 1 has an average length (under P') equal to its entropy $H(P')$.