ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

School of Computer and Communication Sciences

Handout 18 Information Theory and Coding
Midterm December 17, 2002

Three problems with a total of 100 points.
Duration 2 hours and 45 minutes.
5 sheets of notes allowed.

PROBLEM 1. (25 points) Consider an information source X with alphabet $\mathcal{X}$.

(a) (10 points) Suppose we are given M binary prefix-free codes for the alphabet $\mathcal{X}$. (Each code is an assignment from $\mathcal{X}$ to finite binary strings.) Let $l_m(x)$ be the length of the binary sequence that the m'th code assigns to the symbol $x \in \mathcal{X}$, $m = 1, \ldots, M$. Show that the function

$$l(x) = \lceil \log_2 M \rceil + \min_{1 \le m \le M} l_m(x)$$

satisfies the Kraft inequality.

(b) (5 points) Show that for every m,

$$\sum_x p_m(x) l(x) \le \lceil \log_2 M \rceil + \sum_x p_m(x) l_m(x).$$

(c) (10 points) Suppose we know that a source X has alphabet $\mathcal{X}$ and we know that the distribution of the source is one of M distributions $p_1(x), \ldots, p_M(x)$, but we don't know which of these M distributions is the true one. Show that there is a prefix free code which uses no more than

$$\lceil \log_2 M \rceil + H(X) + 1 \quad \text{bits/symbol}$$

to represent this source no matter which of the p_m's is the true distribution. [Hint: make a careful choice of the l_m's in parts (a) and (b).]

PROBLEM 2. (35 points) Suppose, after careful observation we have learned to partially predict the outcome of a fair coin toss before the coin lands: if $X = \{0, 1\}$ denotes the actual outcome of the coin toss and Y our guess,

$$Pr(X = 0|Y = 0) = P(X = 1|Y = 1) = 3/4.$$

(a) (6 points) Find $I(X;Y)$.

We now go to a casino in which we can bet on the outcome of a coin flip. The casino is fair: we double the money we wager if we guess correctly, we lose our money if we are wrong.

We start the game with an initial capital of C_0. In each successive bet we wager a fraction $(1 - q)$ of our current capital, holding a fraction q of our capital in reserve, $0 \le q \le 1$.

(b) (7 points) Show that after betting on n successive coin flips, our capital C_n is given by

$$C_n = C_0 \prod_{i=1}^{n} (2 - q)^{Z_i} q^{1-Z_i}$$

where $Z_i = 1$ if our guess is correct on the ith coin flip and $Z_i = 0$ if our guess is wrong on the ith coin flip.

(c) (7 points) Find the value of q that maximizes $E[C_n]$.

(d) (8 points) Let

$$R_n = \frac{1}{n} \log_2 \frac{C_n}{C_0}$$

be our 'rate of return on investment.' Find the value of q that maximizes $E[R_n]$. Compare the value of $E[R_n]$ for this q with $I(X;Y)$.

(e) (7 points) In playing a long betting game, which of the values we found for q should we use and why? [Hint: to which quantity does the law of large numbers apply?]

Problem 3. (40 points)

Suppose that we have source X with finite alphabet $\mathcal{X} = \{1, \dots, K\}$, and probability distribution p. Consider a *non-singular* code $C : \mathcal{X} \to \{0,1\}^*$ that encodes this source into finite binary sequences. [Reminder: Non-singular codes are those codes for which $C(x) \neq C(y)$ whenever $x \neq y$. The notation $\{0,1\}^*$ denotes the set of all finite binary sequences, including the null-sequence λ, i.e., $\{0,1\}^* = \{\lambda, 0, 1, 00, 01, 10, 11, 000, 001, \dots\}$.]

Let $l(x) = \text{length}(C(x))$ denote the length of the binary encoding of the source letter x with this code. Define $L = \sum_{x \in \mathcal{X}} p(x) l(x)$ as the average length of the encoding.

(a) (8 points) Show that a non-singular code C satisfies

$$|\{x : l(x) = k\}| \leq 2^k, \quad \text{for all } k = 0, 1, 2, \dots, \tag{1}$$

and that conversely, given a non-negative integer valued function l for which (1) holds, then there is a non-singular code C with these codeword lengths.

(b) (5 points) Show that if C is a non-singular code with least average length L, then

$$l(x) \leq l(y) \quad \text{whenever } p(x) > p(y).$$

(c) (8 points) Suppose that $p(1) \geq p(2) \geq \cdots \geq p(K)$. Show that for a non-singular code with the least average length L,

$$l(i) = \lfloor \log_2(i) \rfloor, \quad i = 1, \dots, K.$$

[Hint: $\lfloor \log_2(i) \rfloor$ is the length of the ith element of the sequence λ, 0, 1, 00, 01, 10, 11, 000, 001, ...]

(d) (7 points) Still supposing $p(1) \geq p(2) \geq \cdots \geq p(K)$, show that the average length L of any non-singular code satisfies

$$L \geq H(X) - 1 - \sum_{i=1}^{K} p(i) \log_2 \frac{1}{ip(i)}.$$

(e) (7 points) Let $S_K = \sum_{i=1}^{K} 1/i$. Show that

$$\sum_{i=1}^{K} p(i) \log \frac{1}{ip(i)} \leq \log S_K.$$

(f) (5 points) Use the fact $S_K \leq 1 + \ln K$, and conclude that for any non-singular code

$$L \geq H(X) - 1 - \log_2(1 + \ln K).$$