ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE
School of Computer and Communication Sciences

Handout 11 Introduction to Communication Systems
Homework 6 October 26, 2010, SG1 – 15:15-17:00

Problem 1. In class we showed that

$$\binom{n}{k} \le 2^{nh_2(\frac{k}{n})}.$$

(a) Prove that

$$\frac{2^{nh_2(\frac{k}{n})}}{n+1} \le \binom{n}{k}$$

(Hint: start by writing $\sum_{i=0}^{n} \binom{n}{i} p^i (1-p)^{n-i} = 1$).

(b) What do the above two inequalities imply about the dominant behavior of $\frac{1}{n} \log \binom{n}{k}$?

Problem 2. In this problem, our aim is to give an alternative proof that $H(W) = H(X)\mathbb{E}[\text{length}(W)]$ for variable to fixed length coding with a valid and prefix-free dictionary. We use induction on the number α of internal nodes. This was suggested by a (smart) student in the class. :-)

(a) What is a valid size of the dictionary in terms of $K = |\mathcal{X}|$ ($\mathcal{X}$ is the set of the underlying alphabets)?

(b) For $\alpha = 1$ do the computation explicitly and prove the formula.

(c) Now assume that the formula is correct for $\alpha \le k$ and show that it then is also correct for $\alpha = k+1$. Expand a dictionary with k internal nodes to one with $k+1$ internal nodes, and assume that the leaf node we extend has probability p. How does $H(W)$ and $\mathbb{E}[\text{length}(W)]$ change?

Problem 3. Decode the string

10010011

that was encoded using the Lempel-Ziv algorithm with alphabet set $\mathcal{X} = \{a, l\}$.

Problem 4. We consider the case of encoding a binary sequence $x^n \in \{0,1\}^n$. We assume that the members of the sequence $x_1, x_2, \ldots, x_n$ are generated independently from Bernoulli distribution with probability p, where p is unknown.

We will encode the sequence in two steps. In the first step, we estimate the distribution p. We first observe the entire sequence, count the number of ones (i.e. $k = \sum_{i=1}^{n} x_i$), and then describe this number.

(a) How many bits need to be reserved for the binary description of k? How many different sequences of length n exist with k ones? Label this number N.

(b) In the second stage of our algorithm, we encode one of the possible N sequences. How many bits are needed for this description?

(c) Find a good upper bound on the total length of the description $l(x^n)$ for our procedure. You may use the following bound:

$$\sqrt{\frac{n}{8k(n-k)}} \leq \binom{n}{k} 2^{-nH(k/n)} \leq \sqrt{\frac{n}{\pi k(n-k)}}.$$

(d) If the length of the optimal code for the Bernoulli distribution corresponding to $\frac{k}{n}$ is $l^*(x^n)$, what is the cost of describing the sequence statistics (i.e. calculate $\frac{l(x^n)-l^*(x^n)}{l^*(x^n)}$). How does this quantity behave as $n \to \infty$?