ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE
School of Computer and Communication Sciences

Handout 16 Information Theory and Coding
Midterm Exam Oct. 28, 2014

3 problems, 95 points
3 hours
2 sheets (4 pages) of notes allowed

Good Luck!

PLEASE WRITE YOUR NAME ON EACH SHEET OF YOUR ANSWERS

PLEASE WRITE THE SOLUTION OF EACH PROBLEM ON A SEPARATE SHEET

PROBLEM 1. (30 points) Consider an encryption system, where the encoder and the decoder share a secret key $K \in \mathcal{K}$. The plaintext $T \in \mathcal{T}$ is encrypted by the encoder f to obtain the ciphertext $C \in \mathcal{C}$ (i.e., $C = f(T, K)$). The plaintext T can be recovered from the ciphertext C by using the decoder g (i.e., $T = g(C, K)$). Let $\mathcal{T}$, $\mathcal{C}$ and $\mathcal{K}$ denote the alphabets of T, C and K. The mappings $f : \mathcal{T} \times \mathcal{K} \longrightarrow \mathcal{C}$ and $g : \mathcal{C} \times \mathcal{K} \longrightarrow \mathcal{T}$ are deterministic functions.

a) (5 pts) What are the values of $H(C|T, K)$ and $H(T|C, K)$?

b) (5 pts) Show that $H(C|K) = H(T|K)$.

c) (5 pts) Suppose that the key K is independent of T. Show that $H(C) \geq H(T)$.

d) (10 pts) Under the same assumption show that $H(C|T) \leq H(K)$.

e) (5 pts) Assume that K is independent of T and that C is independent of T. Show that $H(K) \geq H(T)$.

PROBLEM 2. (30 points) Let $\mathcal{U} = \{a, b\}$ and $\{U_i : i \geq 0\}$ be an i.i.d. process with

$$U_i = \begin{cases} a & \text{with probability } \; p, \\ b & \text{with probability } \; 1-p, \end{cases}$$

where $p > 0$. Consider the dictionary $\mathcal{D}_n = \{a, ba, bba, \ \ldots \ , \underbrace{b \ldots b}_{n-1}a, \ \underbrace{b \ldots b}_{n}\}$. Suppose that $U_1, U_2, \ldots$ is parsed into a sequence of words $W_1, W_2, \ldots$ from $\mathcal{D}_n$.

a) (5 pts) Show that $\mathcal{D}_n$ is a valid prefix-free dictionary.

b) (10 pts) Show that $\mathbb{E}[\text{length}(W_i)] = \frac{1-(1-p)^n}{p}$. Hint: $\displaystyle\sum_{j=0}^{n-1} x^j = \frac{1-x^n}{1-x}$.

c) (5 pts) Calculate $H(W_i)$.

d) (5 pts) Show that there exists a uniquely decodable code $\mathcal{C}_n : \ \mathcal{D}_n \to \{0,1\}^*$ such that $\mathbb{E}[\text{length}(\mathcal{C}_n(W_i))] \leq H(W) + 1$.

e) (5 pts) Show that for any $\epsilon > 0$ there exists n and a coding scheme based on $\mathcal{D}_n$ and $\mathcal{C}_n$ which can encode $U_1, U_2, \ldots$ using on average at most $H(U) + p + \epsilon$ bits/letter.

PROBLEM 3. (35 pts) Recall that s is a prefix of t if t is of the form $t = sv$, the concatenation of s and v for some string v. Similarly we say s is a *suffix* of t if $t = vs$. E.g., the suffixes of "banana" are "a", "na", "ana", "nana, "anana" and "banana".

A code $\mathcal{C}$ is said to be a *fix-free code* if and only if no codeword is the prefix or the suffix of any other codeword. Let $l_1, \ldots, l_k$ be k integers satisfying $l_1 \leq \ldots \leq l_k$. Consider the following algorithm:

- Initialize $A_i = \{0,1\}^{l_i}$ as the set of available codewords of length l_i for every $1 \leq i \leq k$.

for $i = 1 \ldots k$ **do**
  if $A_i \neq \emptyset$ **then**
    - Pick $\mathcal{C}(i) \in A_i$.
    for $j = i \ldots k$ **do**
      - (*) Remove from A_j all the words which start with $\mathcal{C}(i)$.
      - (**) Remove from A_j all the words which end with $\mathcal{C}(i)$.
    end
  else
    - Algorithm failure.
  end
end

- Return $\mathcal{C} = \{\mathcal{C}(i) : \ 1 \leq i \leq k\}$.

(a) (10 pts) For every $1 \leq i \leq k$ and every $i \leq j \leq k$, show that the number of words in A_j that start with $\mathcal{C}(i)$ is $2^{l_j - l_i}$, and that the number of words in A_j that end with $\mathcal{C}(i)$ is $2^{l_j - l_i}$.

(b) (5 pts) Show that the number of words that are removed from A_j in $(*)$ and $(**)$ is at most $2^{l_j - l_i + 1}$.

(c) (5 pts) Show that if $\sum_{i=1}^{k} 2^{-l_i} \leq \frac{1}{2}$, then the algorithm does not fail.

(d) (5 pts) Show that if $\sum_{i=1}^{k} 2^{-l_i} \leq \frac{1}{2}$, then the returned code $\mathcal{C}$ is fix-free and that the lengths of its codewords are $l_1, \ldots, l_k$.

(e) (10 pts) Let U be a random variable taking values in an alphabet $\mathcal{U}$. Show that there exists a fix-free code $\mathcal{C} : \mathcal{U} \longrightarrow \{0,1\}^*$ such that $H(U) \leq \mathbb{E}[\text{length}(\mathcal{C}(U))] \leq H(U)+2$.