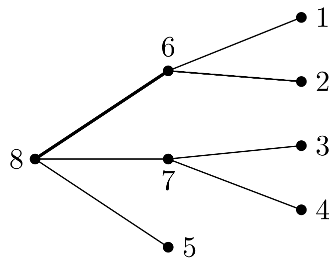
ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

School of Computer and Communication Sciences

Handout 9 — Information Theory and Coding

Homework 5 — October 26, 2007

PROBLEM 1. Consider a tree with M leaves $n_1, \dots, n_M$ with probabilities $P(n_1), \dots, P(n_M)$. Each intermediate node n of the tree is then assigned a probability $P(n)$ which is equal to the sum of the probabilities of the leaves that descend from it. Label each branch of the tree with the label of the node that is on that end of the branch further away from the root. Let $d(n)$ be a "distance" associated with the branch labelled n. The distance to a leaf is the sum of the branch distances on the path to from root to leaf.



For example, in the tree shown above, nodes 1, 2, 3, 4, 5 are leaves, the probability of node 6 is given by $P(1) + P(2)$, the probability of node 7 by $P(3) + P(4)$, of node 8 (root) by $P(1) + P(2) + P(3) + P(4) + P(5) = 1$. The branch indicated by the heavy line would be labelled 6. The distance to leaf 2 is given by $d(6) + d(2)$.

(a) Show that the expected distance to a leaf is given by $\sum_n P(n)d(n)$ where the sum is over all nodes other than the root. Recall that we showed this in the class for $d(n) = 1$.

(b) Let $Q(n) = P(n)/P(n')$ where n' is the parent of n, and define the entropy of an intermediate node n' as

$$H_{n'} = \sum_{n\colon n \text{ is a child of } n'} -Q_n \log Q_n.$$

Show that the entropy of the leaves

$$H(\text{leaves}) = -\sum_{j=1}^{M} P(n_j) \log P(n_j)$$

is equal to $\sum_{n \in I} P(n)H_n$ where the sum is over all intermediate nodes including the root. Hint: use part (a) with $d(n) = -\log Q(n)$.

(c) Let X be a memoryless source with entropy H. Consider some valid prefix-free dictionary for this source and consider the tree where leaf nodes corresponds to dictionary words. Show that $H_n = H$ for each intermediate node in the tree, and show that

$$H(\text{leaves}) = E[L]H$$

where $E[L]$ is the expected word length of the dictionary. Note that we proved this result in class by a different technique.

PROBLEM 2. Consider a valid, prefix-free dictionary of words from a source of alphabet size D.

(a) Show that the set of lengths $L_1, \dots, L_M$ of the dictionary words satisfy the Kraft inequality

$$\sum_j D^{-L_j} \leq 1$$

with equality.

(b) Show that equality can happen only if the number of words $\mod (D-1) = 1$. Hint: Show and use that $\forall j \geq 0$, $D^j = 1 \mod (D-1)$.

(c) Show that if the dictionary is valid, but not prefix-free, then the Kraft inequality is violated.

PROBLEM 3. Construct a Tunstall code with $M = 8$ words in the dictionary for a binary memoryless source with $P(0) = 0.9$, $P(1) = 0.1$.

PROBLEM 4. From the notes on the Lempel-Ziv algorithm, we know that the maximum number of distinct words c a string of length n can be parsed into satisfies

$$n > c \log_K (c/K^3)$$

where K is the size of the alphabet the letters of the string belong to. This inequality lower bounds n in terms of c. We will now show that n can also be upper bounded in terms of c.

(a) Show that, if $n \geq \frac{1}{2}m(m-1)$, then $c \geq m$.

(b) Find a sequence for which the bound in (a) is met with equality.

(c) Show now that $n < \frac{1}{2}c(c+1)$.

PROBLEM 5. Show that for all $\alpha > 1$ one can construct a prefix-free code $\mathcal{C}$ mapping positive integers to variable length binary strings,

$$\mathcal{C}: \ \mathcal{Z}^+ \to \{0,1\}^*$$

with $l(n)$, the length of the codeword corresponding to integer n, satisfying

$$l(n) \leq \alpha \log n + \text{const}(\alpha)$$

for some appropriate constant const(α). Hint: $\sum_n \frac{1}{n^\alpha} < \infty$ for $\alpha > 1$.

PROBLEM 6. Consider an infinite sequence $X = \{\dots, X_{-2}, X_{-1}, X_0, X_1, X_2, \dots\}$ generated by a stationary ergodic source (the only property we need of an ergodic source is that each pattern with a positive probability occurs infinitely many times in X). The symbol X_i takes values in a finite set $\mathcal{X}$ and is distributed as $p(x)$. We consider a scheme for the lossless encoding of X.

For $l = 1, 2, \dots$ define $N_l(X)$ as the *smallest* integer $N \geq 1$ such that $X_1^l = X_{-N+1}^{-N+l}$. As an example suppose that X_k is as shown in Figure. 1. Consider $l = 4$. The pattern $X_1^4 = (b\, b\, c\, a)$. As we slide this pattern to the left, observe that we get the first perfect match at the 5$^{\text{th}}$ slide (since $X_1^4 = X_{-4}^{-1}$). Therefore, $N_4(X) = 5$. Also observe that $N_2(X) = 1$.

Fix l, and assume that the sequence $\{\dots, X_{-2}, X_{-1}, X_0\}$ has been received losslessly at the decoder. The encoder encodes the sequence $\{X_1, X_2, \dots, X_l\}$ by looking at the past and identifying the *smallest* integer $N_l \geq 1$ such that $X_1^l = X_{-N_l+1}^{-N_l+l}$.

N_l is then described to the decoder using a carefully chosen code $\mathcal{C}$. We will show that with this scheme one can get arbitrarily close to the entropy rate of the source.

$k:$	-5	-4	-3	-2	-1	0	1	2	3	4	5
$X_k:$	a	b	b	c	a	b	b	b	c	a	c

Figure 1: Example of a pattern

(a) Consider a $l-$length sequence x_1^l whose probability of occurence p_l is given by $p_l = p(x_1, x_2, \ldots, x_l)$. Now define a binary random sequence $\{Y_i\}_{-\infty}^{\infty}$ by

$$Y_i = \begin{cases} 1 & X_{i+1}^{i+l} = x_1^l \\ 0 & \text{else} \end{cases}$$

Note that the Y_i's are not i.i.d. Further, we also define $Q(k)$ as

$$\begin{aligned} Q(k) &= \Pr(N_l(X) = k | X_1^l = x_1^l) \\ &= \Pr(Y_{-k} = 1, Y_{-j} = 0 \text{ for } 1 \leq j < k | Y_0 = 1) \end{aligned}$$

Through the following sequence of steps we will prove that

$$E(N_l(X)|X_1^l = x_1^l) = \frac{1}{p_l}$$

Show that

(i)

$$\sum_{k=1}^{\infty} \sum_{i=0}^{\infty} \Pr(Y_{-k} = 1, Y_j = 0 \text{ for } -k < j < i, Y_i = 1) = 1$$

Hint - Define E_{ik}, where E_{ik} contains all binary sequences Y such that

$$Y_j = \begin{cases} 1 & j = -k, i \\ 0 & -k < j < i \\ \text{arbitrary} & \text{else} \end{cases}$$

Using these justify that the expression on the left hand side is a probability.

(ii)

$$\sum_{k=1}^{\infty} \sum_{i=0}^{\infty} \Pr(Y_{-k} = 1, Y_j = 0 \text{ for } -k < j < i, Y_i = 1) = \Pr(Y_0 = 1) \sum_{k=1}^{\infty} \sum_{i=0}^{\infty} Q(i+k)$$

(iii)

$$\Pr(Y_0 = 1) \sum_{k=1}^{\infty} \sum_{i=0}^{\infty} Q(i+k) = p_l E(N_l(X)|X_1^l = x_1^l)$$

(b) N_l is described to the decoder using the code $\mathcal{C}$ described in Problem 5 (prefix free code). Justify the following steps

$$\begin{aligned} E(l(N_l)) &= E(E(\alpha \log N_l + \text{const}(\alpha)|X_1^l) \\ &\leq E(\alpha \log E(N_l|X_1^l) + \text{const}(\alpha)) \\ &= \alpha H(X_1^l) + \text{const}(\alpha) \end{aligned}$$

Show that one can get arbitrarily close to the entropy rate of the source by choosing l and α carefully.
Hint: In order to show the inequality, first show that for a positive random variable N,

$$E(\log N) \leq \log(E(N))$$