ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE
School of Computer and Communication Sciences

Handout 10 Information Theory and Coding
Solutions to homework 5 Oct 26, 2007

PROBLEM 1.

(a) Let I be the set of intermediate nodes (including the root), let N be the set of nodes except the root and let L be the set of all leaves. For each $n \in L$ define $A(n) = \{m \in N : m \text{ is an ancestor of } n\}$ and for each $m \in N$ define $D(m) = \{n \in L : n \text{ is a descendant of } m\}$. We assume each leaf is an ancestor and a descendant of itself. Then

$$E[\text{distance to a leaf}] = \sum_{n \in L} P(n) \sum_{m \in A(n)} d(m)$$
$$= \sum_{m \in N} d(m) \sum_{n \in D(m)} P(n) = \sum_{m \in N} P(m) d(m).$$

(b) Consider any leaf node say n_j. Consider the unique path in the tree from the leaf node n_j to the root. Let us label the nodes, which we encounter along the path to the root, as $n_j^1, n_j^2, \ldots, n_j^l$ where n_j^l is the root of the tree. We observe that

$$P(n_j) = \frac{P(n_j)}{P(n_j^1)} \frac{P(n_j^1)}{P(n_j^2)} \cdots \frac{P(n_j^{l-1})}{P(n_j^l)} \tag{1}$$

where $P(n_j^i)$ are the probabilities assigned in the usual way to the intermediate nodes. Thus from the definition of $Q(n)$ we can say that

$$P(n_j) = Q(n_j) Q(n_j^1) \cdots Q(n_j^{l-1}) \tag{2}$$

Let $d(n) = -\log Q(n)$. We see that $-\log P(n_j)$ is the distance associated with a leaf. From part (a),

$$\begin{aligned}
H(\text{leaves}) &= E[\text{distance to a leaf}] \\
&= \sum_{n \in N} P(n) d(n) \\
&= -\sum_{n \in N} P(n) \log Q(n) \\
&= -\sum_{n \in N} P(\text{parent of } n) Q(n) \log Q(n) \\
&= -\sum_{m \in I} P(m) \sum_{n:\ n \text{ is a child of } m} Q(n) \log Q(n) \\
&= \sum_{m \in I} P(m) H_{m'}
\end{aligned}$$

(c) Let us assume that there are K symbols. Remember that for a valid dictionary we require all the paths in the tree to have atleast one word and prefix free means that the words should be leaves. Hence from every intermediate node there are K children and clearly $P(\text{child})/P(\text{parent}) = p(k)$ where $p(k)$ is the probability of the symbol k. As a result

$$H_n = -\sum_{n:n \text{ is a child of n}'} \frac{P(n)}{P(n')} \log \frac{P(n)}{P(n')} = -\sum_k p(k) \log p(k) = H$$

Thus each $H_n = H$. Thus $H(\text{leaves}) = H \sum_{n \in I} P(n) = HE[L]$.

PROBLEM 2. (a) Since the words of a valid and prefix condition dictionary reside in the leaves of a full tree, the Kraft inequality must be satisfied with equality: Consider climbing up the tree starting from the root, choosing one of the D branches that climb up from a node with equal probability. The probability of reaching a leaf at depth l_i is then D^{-l_i}. Since the climbing process will certainly end in a leaf, we have

$$1 = \Pr(\text{ending in a leaf}) = \sum_i D^{-l_i}.$$

(b) Multiplying both sides of the expression above by $D^{l_{max}}$, where l_{max} is the maximum length of a string, we have

$$D^{l_{max}} = \sum_i D^{l_{max}-l_i}$$

We also have that $\forall j \geq 0$, $D^j = 1 \mod (D-1)$. Taking $\mod (D-1)$ on both sides of the above expression, we have that

$$\begin{aligned} 1 &= (\sum_i D^{l_{max}-l_i}) \mod (D-1) \\ &= (\sum_i D^{l_{max}-l_i} \mod (D-1)) \mod (D-1) \\ &= (\sum_i 1) \mod (D-1) \\ &= (\text{Number of words}) \mod (D-1) \end{aligned}$$

(c) If the dictionary is valid but not prefix-free, by removing all words that already have a prefix in the dictionary we would obtain a valid prefix-free dictionary. Since this reduced dictionary would satisfy the Kraft inequality with equality, the extra words would cause the inequality to be violated.

PROBLEM 3. Upon noticing $0.9^6 > 0.1$, we obtain $\{1, 01, 001, 0001, 00001, 000001, 0000001, 0000000\}$ as the dictionary entries.

PROBLEM 4. Let $s(m) = 0 + 1 + \cdots + (m-1) = m(m-1)/2$. Suppose we have a string of length $n = s(m)$. Then, we can certainly parse it into m words of lengths 0, 1, ..., $(m-1)$, and since these words have different lengths, we are guaranteed to have a distinct parsing. Since a parsing with the maximal number of distinct words will have at least as many words as this particular parsing, we conclude that whenever $n = m(m-1)/2$, $c \geq m$.

An all zero string of length $s(m)$ can be parsed into at most m words: in this case distinct words must have distinct lengths.

Now, given n, we can find m such that $s(m-1) \le n < s(m)$. A string with n letters can be parsed into $m-1$ distinct words by parsing its initial segment of $s(m-1)$ letters with the above procedure, and concatenating the leftover letters to the last word. Thus, if a string can be parsed into $m-1$ distinct words, then $n < s(m)$, and in particular, $n < s(c+1) = c(c+1)/2$.

PROBLEM 5. We have to check that the length of the codewords satisfy Kraft's inequality. Taking $l(n) = \lceil \alpha \log n + \mathrm{const}(\alpha) \rceil - 1$ we get,

$$\sum_n 2^{-l(n)} \le \sum_n 2^{-\mathrm{const}(\alpha)} \frac{1}{n^\alpha}$$

Also, we know that $\forall \alpha > 1$, $\sum_n \frac{1}{n^\alpha} < \infty$. Thus there exists a $\mathrm{const}(\alpha)$ large enough so that the right hand side is less than or equal to 1.

PROBLEM 6.

(a) (i) Let set E_{ik} contain all binary sequences $\{Y\}$ such that

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

The set E of all binary sequences which contain a 1, both at a negative and a non-negative time index, is the disjoint union of the sets E_{ik} for all $i \ge 0, k \ge 1$, i.e., $E = \bigcup_{i \ge 0, k \ge 1} E_{ik}$. Since the complement of E has probability 0 (the probability that a pattern will not occur in the infinite future (or the infinite past) is zero), we have

$$1 = \Pr(E) = \Pr(\bigcup_{i \ge 0, k \ge 1} E_{ik}) = \sum_{k=1}^{\infty} \sum_{i=0}^{\infty} \Pr(E_{ik})$$

Equivalently,

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

(ii) Let L.H.S denote the expression on the left hand side of the above relation. Since the $\{X\}$ sequence is stationary, so is the $\{Y\}$ sequence. We therefore have

$$\begin{aligned}
\text{L.H.S} &= \sum_{k=1}^{\infty} \sum_{i=0}^{\infty} \Pr(Y_{-k} = 1, Y_j = 0 \;\text{ for }\; -k < j < i, Y_i = 1) \\
&= \sum_{k=1}^{\infty} \sum_{i=0}^{\infty} \Pr(Y_{-(i+k)} = 1, Y_j = 0 \;\text{ for }\; -(k+i) < j < 0, Y_0 = 1) \\
&= \sum_{k=1}^{\infty} \sum_{i=0}^{\infty} \Pr(Y_0 = 1) \Pr(Y_{-(k+i)} = 1, Y_j = 0 \;\text{ for }\; -(k+i) < j < 0 | Y_0 = 1) \\
&= \Pr(Y_0 = 1) \sum_{k=1}^{\infty} \sum_{i=0}^{\infty} Q(i+k)
\end{aligned}$$

(iii) Reducing the double summation to a single summation, we have

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

Parts (i),(ii) and (iii) give us that

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

(b) For a positive random variable N,

$$\begin{aligned}
\log N &= \log \frac{N}{E(N)} + \log E(N) \\
&\leq \left(\frac{N}{E(N)} - 1 \right) \log e \, + \log E(N)
\end{aligned}$$

since $\ln x \leq x - 1$. Taking expectation on both sides of the inequality yields $E(\log N) \leq \log E(N)$.

Using this upper bound we get,

$$\begin{aligned}
E(l(N_l)) &= E(E(\alpha \log N_l + \mathrm{const}(\alpha)|X_1^l) \\
&\leq E(\alpha \log E(N_l|X_1^l) + \mathrm{const}(\alpha)) \\
&= E(\alpha \log \frac{1}{p_l} + \mathrm{const}(\alpha)) \\
&= \alpha H(X_1^l) + \mathrm{const}(\alpha).
\end{aligned}$$

The encoding efficiency is $\frac{1}{l} E(l(N_l)) = \frac{1}{l}(\alpha H(X_1^l) + \mathrm{const}(\alpha))$ bits/letter. We take α arbitrarily close to 1 and l large enough so that $\frac{1}{l}\mathrm{const}(\alpha)$ is as small as desired. Thus, we can get arbitrarily close to the entropy rate of the source.