ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

School of Computer and Communication Sciences

Handout 16 Information Theory and Coding

Midterm Solutions November 9, 2007

PROBLEM 1.

(a) Expanding $I(U,T;V)$ by the chain rule:

$$\begin{aligned} I(U,T;V) &= I(U;V) + I(T;V|U) \\ &= I(U;V) && \text{since } T \text{ and } V \text{ are independent conditional on } U \end{aligned}$$

Using the chain rule again

$$\begin{aligned} I(U,T;V) &= I(T;V) + I(U;V|T) \\ &\geq I(U;V|T) && \text{since mutual information is non-negative} \end{aligned}$$

Putting the two together we see that $I(U;V) \geq I(U;V|T)$.

(b) $I(X;Y|W) = \Pr(W=1)I(X;Y|W=1) + \Pr(W=2)I(X;Y|W=2)$. But conditional on $W=k$, the probability distribution of (X,Y) is $p_k(x)p(y|x)$ thus,

$$I(X;Y|W) = \lambda I_1 + (1-\lambda)I_2.$$

(c) We obtain $p(x)$ by summing $p(w,x,y)$ over y and w. This gives

$$p(x) = \lambda p_1(x) + (1-\lambda)p_2(x).$$

(d) Note that $p(w,x,y)$ is of the form $p(w)p(x|w)p(y|x)$, that is Y is independent of W when X is given. Thus by part (a)

$$I(X;Y) \geq I(X;Y|W). \tag{1}$$

Letting $f(p_X)$ denote the value of $I(X;Y)$ as a function the distribution of X we can rewrite (1) as

$$f(\lambda p_1 + (1-\lambda)p_2) \geq \lambda f(p_1) + (1-\lambda)f(p_2)$$

which says that f is concave.

PROBLEM 2.

(a) Note that with $l(u) = \lceil \log_2(1/q(u)) \rceil$ we have $2^{-l(u)} \leq q(u)$, and thus

$$\sum_u 2^{-l(u)} \leq \sum_u q(u).$$

As $q(u) = \sum_{k=1}^K \alpha_k p_k(u)$, we see that $\sum_u q(u) = \sum_k \alpha_k = 1$. Thus $l(u)$ satisfies Kraft's inequality and so a prefix-free code $\mathcal{C}$ with codewords lengths $l(u)$ exist.

(b) Since $\mathcal{C}$ is a prefix free code, its expected codeword length L_k is at least H_k and we get $0 \le L_k - H_k$. To upper bound $L_k - H_k$, note that since $\lceil x \rceil < x + 1$,

$$\begin{aligned} L_k(\mathcal{C}) &= \sum_u p_k(u) \operatorname{length}(\mathcal{C}(u)) \\ &< \sum_u p_k(u)[1 + \log(1/q(u))] \\ &= 1 + \sum_u p_k(u) \log \frac{1}{q(u)}. \end{aligned}$$

Thus, $L_k - H_k < 1 + \sum_u p_k(u) \log[p_k(u)/q(u)]$. Observe now that $q(u) \ge \alpha_k p_k(u)$, thus $p_k(u)/q(u) \le 1/\alpha_k$, and

$$L_k - H_k < 1 + \sum_u p_k(u) \log(1/\alpha_k) = 1 + \log(1/\alpha_k).$$

(c) Choosing $\alpha_k = 1/K$ for each k, we get the desired conclusion.

(d) We can view the source as producing a sequence of 'supersymbols' each consisting of a block of L letters. Applying part (c) to this 'supersource', and noticing that the entropy of the supersymbols is $H(U_1, \dots, U_L) = LH(U)$, we see that there is a prefix-free code for which

$$E_k[\text{number of bits to describe a supersymbol}] - LH_k \le 1 + \log_2 K.$$

for each k. Dividing the above by L we get the desired conclusion.

Problem 3.

(a) The intermediate nodes of a tree have the property that if w is an intermediate node, then so are its ancestors. Conversely, as we remark on the notes on Tunstall coding, if a set of nodes has this property, it is the intermediate nodes of some tree. Thus, all we need to show is that $w \in \mathcal{S}$ implies that its prefixes are also in $\mathcal{S}$.

Suppose v is a prefix of w, and $v \ne w$. Then $p_j(v) > p_j(w)$. Thus, $\hat{p}(v) > \hat{p}(w)$. Since $\mathcal{S}$ is constructed by picking nodes with highest possible values of $\hat{p}$, we see that if $w \in \mathcal{S}$, then $v \in \mathcal{S}$.

From class, we know that if a K-ary tree has α intermediate nodes and the tree has $1 + (K-1)\alpha$ leaves.

(b) Since $\mathcal{S}$ contains the α nodes with the highest value of $\hat{p}$, no node outside of $\mathcal{S}$ can have a strictly larger $\hat{p}$ than any node in $\mathcal{S}$. Thus, $\hat{p}(w) \le Q$.

(c) From part (b) $p_j(w) \le \hat{p}(w) \le Q$. Thus, $\log(1/p_j(w)) \ge \log(1/Q)$. Multipling both sides by $p_j(w)$ and summing over all w we get

$$H_j(W) \ge \log(1/Q).$$

(d) For any leaf w in $\mathcal{W}$ we have

$$\begin{aligned} p_1(w) &= p_1(\text{parent of } w) p_1(\text{last letter of } w) \\ &\ge p_1(\text{parent of } w) p_{1,\min} \end{aligned}$$

For a leaf w in $\mathcal{W}_1$, $p_1(\text{parent of } w) = \hat{p}(\text{parent of } w) \geq Q$. Thus, all leaves in $\mathcal{W}_1$ have $p_1(w) \geq Q p_{1,\min}$. Now

$$1 = \sum_{w \in \mathcal{W}} p_1(w) \geq \sum_{w \in \mathcal{W}_1} p_1(w) \geq |\mathcal{W}_1| Q p_{1,\min}.$$

(e) The same argument as in (d) establishes that $|\mathcal{W}_2| Q p_{2,\min} \leq 1$. Thus

$$|\mathcal{W}| = |\mathcal{W}_1 \cup \mathcal{W}_2| \leq |\mathcal{W}_1| + |\mathcal{W}_2| \leq \frac{1}{Q}[1/p_{1,\min} + 1/p_{2,\min}].$$

(f) By part (e) $\log|\mathcal{W}| \leq \log(1/Q) + \log(1/p_{1,\min} + 1/p_{2,\min})$. By part (c) $\log(1/Q) \leq H_j(W)$, we also know $H_j(W) = H_j(U) E_j[\text{length}(W)]$.

Thus, using $\lceil x \rceil < x + 1$,

$$\begin{aligned} \rho_j &= \frac{\lceil \log|\mathcal{W}| \rceil}{E_j[\text{length}(W)]} \\ &< \frac{1 + H_j(U) E_j[\text{length}[W] + \log(1/p_{1,\min} + 1/p_{2,\min}))]}{E_j[\text{length}(W)]} \\ &= H_j(U) + \frac{1 + \log(1/p_{1,\min} + 1/p_{2,\min})}{E_j[\text{length}(W)]}. \end{aligned} \tag{2}$$

(g) As α gets larger, since $|\mathcal{W}| = 1 + (K-1)\alpha$, $\log|\mathcal{W}|$ get larger. As we saw in part (f), $H_j(W)$ is lower bounded by $\log|\mathcal{W}| - \log(1/p_{1,\min} + 1/p_{2,\min})$, so $H_j(W)$ get larger too. Furthermore, $E_j[\text{length}(W)] = H_j(W)/H_j(U)$, and thus as α gets large $E_j[\text{length}(W)]$ gets larger also. Thus, as α gets large we see that the right hand side of (2) approaches $H_j(U)$.