ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE
Communication Systems Department

Handout 20 Information Theory and Coding
Midterm Solutions December 18, 2001

PROBLEM 1.

(a) Given $V = v$, the MAP decoder will make an error if U is not the same as the $\hat{u}(v)$ that the decoder chooses. Thus

$$\Pr(\text{error} \mid |V = v) = \Pr(U \neq \hat{u}(v) \mid V = v) = 1 - \Pr(U = \hat{u}(v) \mid V = v)$$

But $\hat{u}(v)$ satisfies $\Pr(U = \hat{u}(v) \mid V = v) = \max_u p_{U|V}(u|v)$, so the conclusion follows.

(b) This follows from part (a) by multiplying both sides by $p_V(v)$ and summing over v.

(c) (15 points) For any random variable W,

$$\begin{aligned} H(W) &= -\sum_w p_W(w) \log p_W(w) \\ &= -\log e \sum_w p_W(w) \ln p_W(w) \\ &\geq \sum_w p_W(w)[1 - p_W(w)](\log e). \end{aligned}$$

where the inequality follows from $\ln p_W(w) \leq p_W(w) - 1$ which is equivalent to $-\ln p_W(w) \geq 1 - p_W(w)$.

Since $1 - p_W(w) \geq q \stackrel{\triangle}{=} 1 - \max_w p_W(w)$, we see that

$$\sum_w p_W(w)[1 - p_W(w)] \geq \sum_w p_W(w) q = q$$

from which we conclude that

$$H(W) \geq q = 1 - \max_w p_W(w).$$

(d) This follows exactly as in (c) if replace $p_W(w)$ by $p_{U|V}(u|v)$.

(e) Multiplying both sides of the inequality in (d) with $p_V(v)$ and summing over v we get

$$H(U|V) \geq \sum_v p_V(v)[1 - \max_u p_{U|V}(u|v)] \log e,$$

but, by part (b) the right hand side is $\Pr(U \neq \hat{U}) \log e$.

PROBLEM 2.

(a) We know that for a stationary Markov source the entropy rate is given by

$$H(X_2|X_1) = \Pr(X_1 = 0)H(X_2|X_1 = 0) + \Pr(X_1 = 1)H(X_2|X_1 = 1).$$

But $H(X_2|X_1 = 0) = H(X_2|X_1 = 1) = -\alpha \log \alpha - (1-\alpha)\log(1-\alpha)$ and thus the entropy rate is given by $\mathcal{H}(X) = h(\alpha) = -\alpha \log \alpha - (1-\alpha)\log(1-\alpha)$.

(b) Now, given the values of $R_1, R_2, \ldots, R_k$, we know that the source has emitted R_1 repetitions of the same letter, followed by R_2 repetitions of the other letter, R_3 repetitions of the first letter, etc., and that the $n+1$st letter where $n = R_1 + R_2 + \cdots + R_k$ is different from the nth letter. Since the number of times the source will repeat this $n+1$st letter is does not depend on (i) the past symbols [Markov] and (ii) what the $n+1$st letter is [the source statistics don't change if we replace 0 with 1], we see that R_{k+1} is independent of $R_1, \ldots, R_k$. This shows that $R_1, R_2, \ldots$ form an independent sequence. The condition (ii) above also shows that R_k are identically distributed.

(c) Without loss of generality, let us consider R_1. Since R_1 is the number of times the source repeats itself. So, $R_1 = r$ if and only if

$$X_{k+1} = X_k, \quad k = 1, \ldots, r-1, \quad \text{and} \quad X_{r+1} \neq X_r.$$

But this has probability

$$\Pr(X_{r+1} \neq X_r \mid X_r) \prod_{k=1}^{r-1} \Pr(X_{k+1} = X_k \mid X_k) = (1-\alpha)\alpha^{r-1}, \quad r = 1, 2, \ldots.$$

(d) We compute $E[R_1] = \sum_{r=1}^{\infty} r(1-\alpha)\alpha^{r-1} = 1/(1-\alpha)$, and

$$\begin{aligned}
H(R_1) &= -\sum_{r=1}^{\infty}(1-\alpha)\alpha^{r-1}\log[(1-\alpha)\alpha^{r-1}] \\
&= -\sum_{r=1}^{\infty}(1-\alpha)\alpha^{r-1}\log(1-\alpha) - \sum_{r=1}^{\infty}(r-1)(1-\alpha)\alpha^{r-1}\log\alpha \\
&= -\log(1-\alpha) - \alpha/(1-\alpha)\log\alpha \\
&= h(\alpha)/(1-\alpha).
\end{aligned}$$

(e) A long sequence $X_1, X_2, \ldots, X_N$, will correspond to long sequence $R_1, \ldots, R_M$ of run lengths. Since each R_i encodes R_i of the source letters, by the law of large numbers $N/M \to E[R_1]$. By the efficient encoding, the number of encoded bits L satisfies

$$(L-1)/M \to \mathcal{H}(R).$$

Putting these together we see that

$$L/N \to \mathcal{H}(R)/E[R] = h(\alpha) = \mathcal{H}(X).$$

PROBLEM 3.

(a) Let $B_i = H(X_i|X_1, \ldots, X_{i-1})$. Since the source is stationary

$$B_i = H(X_{i+1}|X_2, \ldots, X_i) \geq H(X_{i+1}|X_1, \ldots, X_i) = B_{i+1},$$

so that B_i is a non-increasing sequence. Now

$$\begin{aligned} A_{n+1} &= \frac{1}{n+1}[B_{n+2} + \cdots + B_{2n+1} + B_{2n+2}] && \text{chain rule} \\ &\leq \frac{1}{n+1}[B_{n+1} + \cdots + B_{2n} + B_{2n+1}] && \text{since } B_{i+1} \leq B_i \end{aligned}$$

Now, $B_{2n+1} \leq B_i$ for $i = n+1, \ldots, 2n$. If we sum these n inequalities, we get $nB_{2n+1} \leq B_{n+1} + \cdots + B_{2n}$, or $B_{2n+1} \leq \frac{1}{n}[B_{n+1} + \cdots + B_{2n}]$. Using this in our bound for A_{n+1} we get

$$\begin{aligned} A_{n+1} &\leq \frac{1}{n+1}(1 + 1/n)[B_{n+1} + \cdots + B_{2n}] \\ &= \frac{1}{n}[B_{n+1} + \cdots + B_{2n}] \\ &= A_n. \end{aligned}$$

(b) From $A_n = \frac{1}{n}[B_{n+1} + \cdots + B_{2n}]$ and $B_{n+1} \geq \cdots \geq B_{2n}$ we see that

$$B_{n+1} \geq A_n \geq B_{2n}.$$

(c) Since $\lim_{n\to\infty} B_{n+1} = \lim_{n\to\infty} B_{2n} = \mathcal{H}(X)$, from part (b) we see that $\lim_{n\to\infty} A_n(X) = \mathcal{H}(X)$.

(d) Since the source is stationary we can, without loss of generality, consider the encoding of Y_2 by C_{Y_1}. Now, conditional of $Y_1 = v$, Y_2 has distribution $p_{Y_2|Y_1=v}$, and since C_v is uniquely decodable the average number of bits used to encode Y_2 when $Y_1 = v$ is at least $H(Y_2|Y_1 = v)$. Since $Y_1 = v$ with probability $p_{Y_1}(v)$, the average number of bits used to encode Y_2 is at least $H(Y_2|Y_1)$. But since the encoding of Y_2 encodes n source letters, we see that

$$L_n \geq \frac{1}{n}H(Y_2|Y_1) = \frac{1}{n}H(X_{n+1}, \ldots, X_{2n}|X_1, \ldots, X_n) = A_n(X).$$

(e) We can choose the code C_v to have codeword lengths $l(y) = \lceil -\log_2 p_{Y_2|Y_1}(y|v) \rceil$. Note that the codeword lengths satisfy the Kraft inequality:

$$\sum_y 2^{-l(y)} \leq \sum_y 2^{\log_2 p_{Y_2|Y_1}(y|v)} = \sum_y p_{Y_2|Y_1}(y|v) = 1,$$

and the average codeword length

$$\sum_y p_{Y_2|Y_1}(y|v) l(y) < \sum_y p_{Y_2|Y_1}(y|v)[-\log_2 p_{Y_2|Y_1}(y|v) + 1] = H(Y_2|Y_1 = v) + 1.$$

Since $Y_1 = v$ with probability $p_{Y_1}(v)$, the average number of bits used to encode Y_2 with these encoders is at most $H(Y_2|Y_1) + 1$. Since the encoding of Y_2 encodes n source letters and the best encoder must to at least as well, we see that

$$\min L_n < \frac{1}{n}H(Y_2|Y_1) + \frac{1}{n} = A_n(X) + \frac{1}{n}.$$