EE 376A/Stat 376A
Information Theory
Prof. T. Cover

Handout #20
Saturday, February 14, 2009

Solutions to Sample Midterm Examination (Originally given in 2001)

1. The stationary distribution is $[\frac{2}{22}, \frac{2}{22}, \frac{3}{22}, \frac{3}{22}, \frac{2}{22}, \frac{2}{22}, \frac{5}{22}, \frac{1}{22}, \frac{1}{22}, \frac{1}{22}]$. Hence, the entropy rate is

$$\begin{aligned}
H &= H(X_2|X_1) \\
&= \sum_i \mu_i H(X_2|X_1 = i) \\
&= \sum_i \frac{E_i}{2E} \log E_i \\
&= \frac{2}{22}\log 2 + \frac{2}{22}\log 2 + \frac{3}{22}\log 3 + \frac{3}{22}\log 3 + \frac{2}{22}\log 2 + \frac{2}{22}\log 2 + \frac{5}{22}\log 5 \\
&= \frac{4}{11} + \frac{3}{11}\log 3 + \frac{5}{22}\log 5.
\end{aligned}$$

2. (a) D=2

(00)	8	8	8	10	16	26
(01)	8	8	8	8	10	
(10)	7	7	7	8		
(111)	1	2	3			
(1100)	1	1				
(1101)	1					

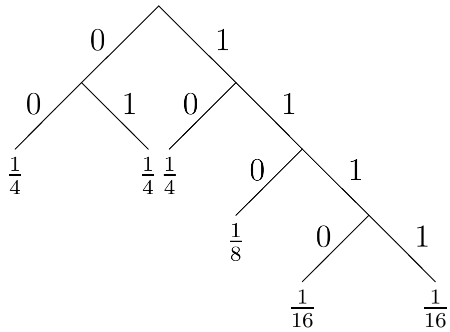
(b) D=4

(0)	8	8	26
(1)	8	8	
(2)	7	7	
(30)	1	3	
(31)	1		
(32)	1		
(dummy)	0		

(c) The first question implied by the binary code is "Is the first bit equal to 1 in the object's codeword?" Hence, the entropy of the first question is equal to the entropy of the first bit of the binary code, $H(\frac{16}{26}) = H(\frac{8}{13}, \frac{5}{13}) = \log 13 - \frac{5}{13}\log 5 - \frac{14}{13}$ bits.

3. (a) Since the distribution is dyadic, by Theorem 5.12.2 in the text, using the Huffman code tree is optimal. The Huffman code for the given distribution is as follows:

$$
\begin{array}{llllll}
(00) & 4 - 4 - 4 - 8 - 8 & 16 \\
(01) & 4 - 4 - 4 - 4 & 8 \\
(10) & 4 - 4 - 4 & 4 \\
(110) & 2 - 2 & 4 \\
(1110) & 1 & 2 \\
(1111) & 1 &
\end{array}
$$



Any tree with same lengths of codewords would be also optimal.

(b)

$$ET = 2 \times \frac{1}{4} + 2 \times \frac{1}{4} + 2 \times \frac{1}{4} + 3 \times \frac{1}{8} + 4 \times \frac{1}{16} + 4 \times \frac{1}{16} = \frac{19}{8}.$$

(c) Again by Theorem 5.12.2, $H(X) = ET = \frac{19}{8}$ bits.

4. Fix $b_1 = a$. We maximize $\sum_{k=1}^{m} p_k \log b_k o_k = \sum_{k=1}^{m} p_k \log m b_k$ as a function of $\mathbf{b}$ subject to the constraint $\sum_{k=1}^{m} b_k = 1$ and $b_1 = a$, i.e. $\sum_{k=2}^{m} b_k = 1 - a$. Writing the functional with a Lagrange multiplier, we have

$$J(\mathbf{b}) = \sum p_i \log m b_i + \lambda \sum b_i.$$

Differentiating this with respect to b_i yields

$$\frac{\partial J}{\partial b_i} = \frac{p_i}{b_i} + \lambda, \qquad i = 2, \ldots, m.$$

Setting the partial derivative equal to 0 for a maximum, we have

$$b_i = -\frac{p_i}{\lambda}.$$

Substituting this in the constraint $\sum_{k=2}^{m} b_k = 1 - a$ yields

$$\begin{aligned} 1 - p_1 &= \sum_{k=2}^{m} p_k \\ &= -\lambda \sum_{k=2}^{m} b_k \\ &= -\lambda(1-a), \end{aligned}$$

that is,

$$\lambda = -\frac{1-p_1}{1-a}$$

and

$$b_i = (1-a)\frac{p_i}{1-p_1}.$$

5. (a) Let μ_n denote the probability mass function at time n. Since $\mu_1 = (\frac{1}{3}, \frac{1}{3}, \frac{1}{3})$ and $\mu_2 = \mu_1 P = \mu_1$, $\mu_n = \mu_1 = (\frac{1}{3}, \frac{1}{3}, \frac{1}{3})$ for all n and $\{X_n\}$ is stationary.
Alternatively, the observation P is doubly stochastic will lead the same conclusion.

(b) Since $\{X_n\}$ is stationary Markov,

$$\begin{aligned} \lim_{n\to\infty} H(X_1, \ldots, X_n) &= H(X_2|X_1) \\ &= \sum_{k=0}^{2} P(X_1 = k) H(X_2|X_1 = k) \\ &= 3 \times \frac{1}{3} \times H(\frac{1}{2}, \frac{1}{4}, \frac{1}{4}) \\ &= \frac{3}{2}. \end{aligned}$$

(c) Since $(X_1, \ldots, X_n)$ and $(Z_1, \ldots, Z_n)$ are one-to-one, by the chain rule of entropy and the Markovity,

$$\begin{aligned} H(Z_1, \ldots, Z_n) &= H(X_1, \ldots, X_n) \\ &= \sum_{k=1}^{n} H(X_k|X_1, \ldots, X_{k-1}) \\ &= H(X_1) + \sum_{k=2}^{n} H(X_k|X_{k-1}) \\ &= H(X_1) + (n-1)H(X_2|X_1) \\ &= \log 3 + \frac{3}{2}(n-1). \end{aligned}$$

Alternatively, we can use the results of parts (d), (e), and (f). Since $Z_1, \dots, Z_n$ are independent and $Z_2, \dots, Z_n$ are identically distributed with the probability distribution $(\frac{1}{2}, \frac{1}{4}, \frac{1}{4})$,

$$\begin{aligned} H(Z_1, \dots, Z_n) &= H(Z_1) + H(Z_2) + \dots + H(Z_n) \\ &= H(Z_1) + (n-1)H(Z_2) \\ &= \log 3 + \frac{3}{2}(n-1). \end{aligned}$$

(d) Since $\{X_n\}$ is stationary with $\mu_n = (\frac{1}{3}, \frac{1}{3}, \frac{1}{3})$,

$$H(X_n) = H(X_1) = H(\frac{1}{3}, \frac{1}{3}, \frac{1}{3}) = \log 3.$$

For $n \geq 2$, $Z_n = \begin{cases} 0, & \frac{1}{2}, \\ 1, & \frac{1}{4}, \\ 2, & \frac{1}{4}. \end{cases}$

Hence, $H(Z_n) = H(\frac{1}{2}, \frac{1}{4}, \frac{1}{4}) = \frac{3}{2}$.

(e) Due to the symmetry of P, $P(Z_n|Z_{n-1}) = P(Z_n)$ for $n \geq 2$. Hence, $H(Z_n|Z_{n-1}) = H(Z_n) = \frac{3}{2}$.

Alternatively, using the result of part (f), we can trivially reach the same conclusion.

(f) Let $k \geq 2$. First observe that by the symmetry of P, $Z_{k+1} = X_{k+1} - X_k$ is independent of X_k. Now that

$$\begin{aligned} P(Z_{k+1}|X_k, X_{k-1}) &= P(X_{k+1} - X_k|X_k, X_{k-1}) \\ &= P(X_{k+1} - X_k|X_k) \\ &= P(X_{k+1} - X_k) \\ &= P(Z_{k+1}), \end{aligned}$$

Z_{k+1} is independent of (X_k, X_{k-1}) and hence independent of $Z_k = X_k - X_{k-1}$.

For $k = 1$, again by the symmetry of P, Z_2 is independent of $Z_1 = X_1$ trivially.

6. (a)

$$\begin{aligned} 1 &= \sum_{x^n \in \mathcal{X}^n} p(x^n) \\ &\geq \sum_{x^n \in C_n(t)} p(x^n) \\ &\geq \sum_{x^n \in C_n(t)} 2^{-nt} \\ &= |C_n(t)| 2^{-nt}. \end{aligned}$$

Hence, $|C_n(t)| \leq 2^{nt}$.

(b) By AEP, for any $\epsilon > 0$, $P(A_\epsilon^{(n)}) \to 1$ where

$$A_\epsilon^{(n)} = \{x^n \in \mathcal{X}^n : 2^{-n(H+\epsilon)} \leq p(x^n) \leq 2^{-n(H-\epsilon)}\}.$$

Hence, the sufficient condition for $P(\{X^n \in C_n(t)\}) \to 1$ is
$C_n(t) \supset A_\epsilon^{(n)}$ for some $\epsilon > 0$,
i.e. $\exists \epsilon > 0$ such that $2^{-nt} \leq 2^{-n(H+\epsilon)}$,
i.e. $\exists \epsilon > 0$ such that $t \geq H + \epsilon$,
i.e. $t > H$.
Therefore, $t > H$ implies $P(\{X^n \in C_n(t)\}) \to 1$.

Note that for $t < H$, $P(C_n(t)) \to 0$ since $-\frac{1}{n}\log p(X^n) \to H$ in probability by the weak law of large numbers.