EE 121 - Introduction to Digital Communications
Homework 3 Solutions

February 10, 2008

1 (a)

$$\begin{aligned}
H(X_1, X_2) &= -\sum_{x_1}\sum_{x_2} \Pr(X_1 = x_1, X_2 = x_2) \log \Pr(X_1 = x_1, X_2 = x_2) \\
&= -\sum_{x_1}\sum_{x_2} \Pr(X_1 = x_1, X_2 = x_2) \log \Pr(X_1 = x_1) \Pr(X_2 = x_2 | X_1 = x_1) \\
&= -\sum_{x_1}\sum_{x_2} \Pr(X_1 = x_1, X_2 = x_2) \log \Pr(X_1 = x_1) \\
&\quad -\sum_{x_1}\sum_{x_2} \Pr(X_1 = x_1, X_2 = x_2) \log \Pr(X_2 = x_2 | X_1 = x_1) \\
&= -\sum_{x_1} \Pr(X_1 = x_1) \log \Pr(X_1 = x_1) \\
&\quad -\sum_{x_1}\sum_{x_2} \Pr(X_1 = x_1, X_2 = x_2) \log \Pr(X_2 = x_2 | X_1 = x_1) \\
&= H(X_1) + H(X_2 | X_1)
\end{aligned}$$

(b) By application of part (a) we have

$$\begin{aligned}
H(X_1, X_2, X_3) &= H(X_1) + H(X_2, X_3 | X_1) \\
&= H(X_1) + H(X_2 | X_1) + H(X_3 | X_2, X_1)
\end{aligned}$$

(c) If X_1 and X_3 are conditionally independent given X_2 then $\Pr(X_3 | X_1, X_2) = \Pr(X_3 | X_2)$.

Thus

$$
\begin{aligned}
H(X_3|X_2,X_1) &= \sum_{x_1}\sum_{x_2} \Pr(X_1=x_1,X_2=x_2) \\
&\quad \sum_{x_3} \Pr(X_3=x_3|X_2=x_2,X_1=x_1)\log \Pr(X_3=x_3|X_2=x_2,X_1=x_1) \\
&= \sum_{x_1}\sum_{x_2} \Pr(X_1=x_1,X_2=x_2) \\
&\quad \sum_{x_3} \Pr(X_3=x_3|X_2=x_2)\log \Pr(X_3=x_3|X_2=x_2) \\
&= \sum_{x_2} \Pr(X_2=x_2) \sum_{x_3} \Pr(X_3=x_3|X_2=x_2)\log \Pr(X_3=x_3|X_2=x_2) \\
&= H(X_3|X_2)
\end{aligned}
$$

2 (a) A probability distribution $\pi(x)$ is a stationary distribution for a Markov chain with states $x \in \mathcal{S}$ if

$$\sum_{x\in\mathcal{S}} \pi(x)\Pr(x,y) = \pi(y) \tag{1}$$

for all states $y \in \mathcal{S}$, where $\Pr(x,y)$ is the probability of transitioning from state x to state y. For the Mickey mouse chain the stationary distribution is $\pi(1) = 1/2$ and $\pi(2) = 1/2$ for $\alpha \in [0,1)$. For $\alpha = 1$ the Markov chain is reducible and the stationary distribution is not unique, in fact all distributions $\pi(x)$ are stationary in this case, as can be verified from equation (1).

(b) Let α denote the self-transition probability.

$$
\begin{aligned}
\Pr(X_n=0) &= \alpha\Pr(X_{n-1}=0) + (1-\alpha)(1-\Pr(X_{n-1}=0)) \\
&= 1-\alpha+(2\alpha-1)P(X_{n-1}=0)
\end{aligned}
$$

The solution to the difference equation $y_n = a + by_{n-1}$ is $y_n = a(1-b^{n-1})/(1-b) + b^{n-1}y_1$. Thus for $\alpha \neq 0$ the distribution is

$$\Pr(X_n=0) = \frac{(1-\alpha)(1-(2\alpha-1)^{n-1})}{2\alpha} + (2\alpha-1)^{n-1}p$$

and $\Pr(X_n=1) = 1-\Pr(X_n=0)$. For $\alpha=0$ we get $\Pr(X_n=0)=p$ if n is odd and $\Pr(X_n=0)=1-p$ if n is even. For all p and $\alpha \in (0,1)$ the distribution of X_n always converges to the stationary distribution but for $\alpha=0$ the distribution of X_n is always equal to the stationary distribution for $p=1/2$ but never equal to it for $p \neq 1/2$.

(c)

$$\begin{aligned} H &= \lim_{n\to\infty} \frac{H(X_1,\dots,X_n)}{n} \\ &= \lim_{n\to\infty} \frac{1}{n} H(X_1) + \frac{1}{n}\sum_{i=2}^{n} H(X_n|X_{n-1}) \\ &= \lim_{n\to\infty} \frac{1}{n}\sum_{i=2}^{n} H(X_n|X_{n-1}). \end{aligned}$$

For $\alpha \notin \{0,1\}$ we have

$$\begin{aligned} H &= H(X_2|X_1) \\ &= -\alpha\log\alpha - (1-\alpha)\log(1-\alpha) \end{aligned}$$

and for $\alpha \in \{0,1\}$ we have $H(X_n|X_{n-1}) = 0$ so $H = 0$. Thus the limit exists for all α but its value depends on whether $\alpha \in \{0,1\}$ or not. It does not depend on the initial distribution.

(d) Solving equation (1) for the stationary distribution we have

$$\pi(0) = \frac{\alpha_1 - 1}{\alpha_0 + \alpha_1 - 2}.$$

and $\pi(1) = 1 - \pi(0)$. The entropy rate exists for all $\alpha_0, \alpha_1 \in [0,1]$ but it's value is zero if either of the $\alpha_i = 1$. The entropy rate does not depend on the initial distribution.

3 (a) The Huffman coding algorithm groups the 0.1 and 0.4 probability symbols into a supersymbol of probability 0.5 and then groups this with the 0.5 probability symbol. The codeword assignments are thus $a = 0$, $b = 10$ and $c = 11$. The expected length is then $\overline{L}_{\min} = 0.5 * 1 + 0.4 * 2 + 0.1 * 2 = 1.5$ bits/symbol.

(b) Now $X^2 \in \{aa, ab, ac, ba, bb, bc, ca, cb, cc\}$ with probabilities

$$\{0.25, 0.2, 0.05, 0.2, 0.16, 0.04, 0.05, 0.04, 0.01\},$$

respectively. Creating a Huffman tree and assigning codewords results in an average length of 2.75 bits per X^2 symbol which is $\overline{L}_{\min,2} = 1.375$ bits/symbol.

(c) By concatenating two identical versions of the code for X from part (a) we can create a prefix-free code for X^2. The average length of this code is $2\overline{L}_{\min}$ bits per X^2 symbol or $\overline{L}_{\min}$ bits per symbol. As this average length must be either equal to or longer than the average length of the optimal code we have $\overline{L}_{\min} \geq \overline{L}_{\min,2}$.

4 (a) It is uniquely decodable because at each step if the decoder reads 0 then it will read the next 3 bits and convert them to integer and decode correctly. If it reads 1 then it will decode 8 a's and move to the next code.

(b)

$$
\begin{aligned}
E[\text{number of bits per B}] &= \sum_{i=0}^{\infty}(i+4)P\{\text{8i+k consecutive a's for some } 0 \le k \le 7\ \} \\
&= \sum_{i=0}^{\infty}(i+4)\left(0.9^{8i} + \ldots + 0.9^{8i+7}\right)0.1 \\
&= \sum_{i=0}^{\infty}(i+4)0.9^{8i}\left(1-0.9^{8}\right) \\
&= (1-0.9^8)\left(\frac{4}{1-0.9^8} + \sum_{i=0}^{\infty} i(0.9)^{8i}\right) \\
&= 4 + (1-0.9^8)\frac{0.9^8}{1-0.9^8}\frac{1}{1-0.9^8} \approx 4.75
\end{aligned}
$$

(c) Define the random variable Y_i to be equal to 1 if we have b at position i, otherwise zero. We are interested in

$$\frac{1}{n}\sum_{i=1}^{n} Y_i$$

By WLLN for any $\epsilon > 0$ we have,

$$P\{|\frac{1}{n}\sum_{i=1}^{n} Y_i - E[Y]| \ge \epsilon\} \to 0, \quad \text{as n increases}$$

Since $E[Y] = 0.1$ it gives the desired result.

(d) $0.1 \times 4.75 = 0.475$.

5 (a) Initially encode the window with 1024 bits then

Window Pointer	Encoded String (u,n)	log w	$2\lfloor \log n \rfloor + 1$	total bits
1024	(1,3976)	10	23	33
5000	(1,1)	0	2	2
5001	(1,3999)	10	23	33
9000	(1,1)	0	2	2
9001	(1,999)	10	19	29

(b) number of bits=1024+33+2+33+2+29=1123

(c) Initially encode the window with 1024 bits then

Window Pointer	Encoded String (u,n)	log w	$2\lfloor \log n \rfloor + 1$	total bits
8	(1,4992)	3	25	28
5000	(1,1)	0	2	2
5001	(1,3999)	3	23	26
9000	(1,1)	0	2	2
9001	(1,999)	3	19	22

Number of bits=88.

(d) Create the markov chain such that:

$$\begin{aligned} P(X_{i+1} = 1 | X_i = 0) &= \frac{P(X_{i+1} = 1, X_i = 0)}{P(X_i = 0)} \\ P(X_{i+1} = 0 | X_i = 1) &= \frac{P(X_{i+1} = 0, X_i = 1)}{P(X_i = 1)} \end{aligned}$$

Now the empirical average is a good estimate for these values:

$$\begin{aligned} P(X_i = 0) &\approx \frac{\text{number of zeros}}{10^4} = 0.6 \\ P(X_i = 1) &\approx \frac{\text{number of ones}}{10^4} = 0.4 \\ P(X_{i+1} = 0, X_i = 1) &\approx \frac{\text{number of (1-0)'s}}{10^4} = 10^-4 \\ P(X_{i+1} = 1, X_i = 0) &\approx \frac{\text{number of (0-1)'s}}{10^4} = 10^-4 \end{aligned}$$

Therefore

$$\begin{aligned} P(X_{i+1} = 1 | X_i = 0) &\approx \frac{1}{6000} \\ P(X_{i+1} = 0 | X_i = 1) &\approx \frac{1}{4000} \end{aligned}$$

(e) Assuming that the Markov chain is in stationary distribution,

$$\lim_{n \to \infty} \frac{H(X_1, \ldots, X_n)}{n} = \lim_{n \to \infty} \frac{H(X_1) + H(X_2|X_1) + H(X_3|X_2) + \ldots + H(X_n|X_{n-1})}{n} = H(X_2|X_1)$$

Now

$$H(X_2|X_1) = \frac{3}{5}H(10^{-4}) + \frac{2}{5}H(10^{-4}) = H(10^{-4})$$