EE606: Midterm Exam - Solutions

(Total 100 points + 10 bonus, Allowed time: 2 hours)

Given: October 26, 2004

1. *(10 pts)* The weatherman's record in a given city is shown in the table below; the numbers indicate the relative frequencies of the joint prediction/actual events. For example, the event that the weatherman predicted rain and it actually rained happens 1/8 of the time. A clever student notices that the weatherman is right only 12/16 of the time, but could be right 13/16 of the time by always predicting no rain. The student explains the situation and applies for the weatherman's job, but the weatherman's boss, who is an information theorist, turns him down. Why?

		Actual	
		Rain	No Rain
Prediction	Rain	1/8	3/16
	No Rain	1/16	10/16

Solution

The weatherman's boss, being an information theorist, knows that the weatherman's prediction is better because it reduces the uncertainty about the actual weather at a higher rate than the student's prediction. To see that, consider $I(X;Y)$ and $I(X;Z)$, where X, Y, and Z are random variables corresponding to the actual weather, weatherman's prediction, and student's prediction. Distribution $p(x,y)$ is given by the weatherman's prediction table above. Thus we can compute $I(X;Y) = 0.091$. Since the student always predicts rain, his predictions are independent of the actual weather and we have $I(X;Z) = 0$. To convince yourself, you can compute $I(X;Z)$ based on $p(x,z)$ given by the following student's prediction table:

		Actual	
		Rain	No Rain
Prediction	Rain	0	0
	No Rain	3/16	13/16

2. *(Total 15 pts)* Prove the following inequalities and say when the equality holds:

   (a) *(5 pts)* $\log x \leq x - 1$ for all $x > 0$.

   (b) *(10 pts)* For two different probability distributions $\{p_i\}$ and $\{q_i\}$, $q_i \neq 0 \; \forall i$, we have

$$\sum_i \frac{p_i^2}{q_i} \geq 1.$$

Solution

(a) Consider $f(x) = \log x - (x-1)$. We have

$$f'(x) = \frac{1}{x} - 1; \quad \text{note that} \quad f'(1) = 0,$$

and

$$f'' = -\frac{1}{x^2} > 0 \quad \forall x.$$

Therefore function f has a unique maximum at $x = 1$, and thus

$$f(x) \le f(1) = 0 \text{ giving } \boxed{\log x \le x - 1}$$

(b) By (a) we have $\log x \le x - 1$ and thus $x \ge \log x + 1$. Therefore,

$$\begin{aligned}\sum_i \frac{p_i^2}{q_i} &= \sum_i p_i \frac{p_i}{q_i} \\ &\ge \sum_i p_i \Big[\log\Big(\frac{p_i}{q_i}\Big) + 1\Big] \\ &= D(p\|q) + 1 \ge 1.\end{aligned}$$

with equality iff $p_i = q_i$ for all i.

3. *(Total 25 pts)* Let X, Y and Z be three random variables that take values in the alphabet $\{0, 1, \ldots, M-1\}$. We assume X and Z are independent and $Y = X + Z(\mod M)$. The distribution of Z is given as $P(Z = 0) = 1 - p$ and $P(Z = i) = p/(M - 1)$, for $i = 1, \ldots, M - 1$. For parts (a)-(c), we will assume that X is uniform on $\{0, 1, \ldots, M - 1\}$.

   (a) *(5 pts)* Find $H(X)$ and $H(Z)$.

   (b) *(5 pts)* Find $H(Y)$.

   (c) *(7 pts)* Calculate $I(X;Y)$ and $I(X;Y|Z)$ and compare. Which one is larger? Explain?

   (d) *(8 pts)* Is there a non-uniform distribution $p(x)$ for X such that $I(X;Y)$ is higher than that of part (c)? If yes, find one such distribution; if no, prove that the uniform distribution maximizes $I(X;Y)$.

Solution

(a) $H(X) = \log M$, $H(Z) = -(1-p)\log(1-p) - p\log\frac{p}{M-1} = h(p) + p\log(M-1)$

(b) We first calculate the probability distribution of Y:

$$P(Y = i) = P(X = i) \cdot (1 - p) + \sum_{\substack{j=0 \\ j \ne i}}^{M} P(X = j) \cdot \frac{p}{M-1} = \frac{1}{M} \tag{1}$$

(c) Note that $H(Y|X) = H(Z)$ and $H(Y|Z) = H(X)$. We thus have

$$I(X;Y) = H(Y) - H(Y|X) = H(Y) - H(Z) = \log M - h(p) - p\log(M-1)$$

and

$$I(X;Y|Z) = H(Y|Z) - H(Y|X,Z) = H(X) - 0 = \log M.$$

We have $I(X;Y|Z) > I(X;Y)$. In fact $I(X;Y|Z) = H(Y)$, because once when we know Z, knowledge about X completely removes the uncertainty about Y.

(d) We have

$$I(X;Y) = H(Y) - H(Z) \leq \log M - H(Z),$$

with equality iff Y is uniform on $\{0, 1, \ldots, M-1\}$, which is achieved with uniform X. No other distribution $p(x)$ would give higher $I(X;Y)$.

4. (*Total 50 pts + bonus)* A memoryless source produces a sequence of binary digits with the probabilities $P(0) = 0.9$ and $P(1) = 0.1$. We encode this sequence in two stages. In the first stage, we count the number of 0s between successive 1s as illustrated by the following example:

1	0	0	1	0	0	0	0	0	0	0	0	0	0	1	1	0	0	0	0	1
0			2								8			2	0					4

Note that when the counter reaches 8, we start anew. A binary source sequence is thus mapped into a sequence of *intermediate* digits taking values in the set $S = \{0, 1, 2, 3, 4, 5, 6, 7, 8\}$. In the second stage, we assign codewords to these intermediate digits as follows: a codeword of one binary digit is assigned to 8 and codewords of four binary digits are assigned to the other integers in S.

(a) (*10 pts)* Specify a codeword assignment to the elements of set S so that this source code is uniquely decodable.

(b) (*10 pts)* Find the expected number $\overline{n}_1$ of source digits per intermediate digit.

(c) (*10 pts)* Find the average number $\overline{n}_2$ of encoded binary digits per intermediate digit.

(d) (*10 pts, bonus)* Use the weak low of large numbers to show that for a very long sequence of source digits, the ratio of the number of encoded binary digits to the number of source binary digits will with high probability be close to $\overline{n}_2/\overline{n}_1$.

(e) (*10 pts)* Demonstrate a Huffman code which encodes four source digits at a time.

(f) (*10 pts)* Find the expected number of encoded binary digits per source binary digit for your Huffman code. Compare the obtained number with $\overline{n}_2/\overline{n}_1$. Which code would you prefer to use? Why? In which sense is the Huffman coding technique optimal?

Solution

(a) A codeword assignment is given in the following table:

symbol	0	1	2	3	4	5	6	7	8
codeword	0000	0001	0010	0011	0100	0101	0110	0111	1

(b)

$$\overline{n}_1 = 8 \cdot P(0)^8 + \sum_{k=1}^{8} k \cdot P(0)^{k-1} P(1) = 5.695$$

(c)

$$\overline{n}_2 = 1 \cdot P(0)^8 + 4 \cdot [1 - P(0)^8] = 2.708$$

(d) $\overline{n}_2/\overline{n}_1 = 0.475$

(e) The 16 input symbols are listed according to their probabilities, together with the resulting Huffman codeword lengths and average length $\overline{l}$:

l_i	x_i	p_i	
1	0000	.6561	
3	0001	.0729	
3	0010	.0729	
3	0100	.0729	
4	1000	.0729	
5	0101	.0081	
7	0110	.0081	
7	0011	.0081	$\Rightarrow \bar{l} = \sum_i p_i l_i = 0.505$
9	1100	.0081	
9	1001	.0081	
9	1010	.0081	
11	1011	.0009	
11	0111	.0009	
11	1101	.0009	
12	1110	.0009	
12	1111	.0001	

(f) The former code achieves better compression. The Huffman coding technique is in general asymptotically optimal.