Homework #1 Solutions

ECE 745 Spring, 2009

1)(a) $\mathcal{C} = \{01, 00, 101, 100, 110, 1110, 1111\}$ is prefix-free and has lower rate $\Rightarrow$ code is not optimal $\Rightarrow$ code is not a Huffman code.

(b)

$$H(x) \le 2 \le H(x) + 1/2$$
$$H(x) \le 1.6 \le H(x) + 1/3$$

$\Rightarrow$

$$2 \ge H(x) \ge 1.5$$
$$1.6 \ge H(x) \ge 1.6 - 1/3$$

$\Rightarrow$

$$1.5 \le H(x) \le 1.6$$

- We know $0 \le H(x) \le \log_2 Q \Rightarrow \log_2 Q \ge 1.5$

$$\Rightarrow Q \ge 2^{1.5} = 2.83$$
$$\Rightarrow Q \ge 3.$$

(Actually, you can show $Q \ge 4$ as follows. Suppose $Q = 3$. Take blocks $N=2$ at a time and employ a code with codeword lengths $\{3,3,3,3,3,3,3,4,4\}$. The maximal rate of this code is

$$\frac{1}{2} \cdot \frac{1}{9}(3 \cdot 7 + 2 \cdot 4) = 1.61 < 2$$

This implies $R_{Huff}^{(2)} \le 1.61$, which violates our assumption. Thus, $Q \ge 4$.)

(c) No. For any m, N and IID source, suppose I build a (suboptimal) lossless code on blocks of length mN by concatenating m Huffman codes of length N. Then

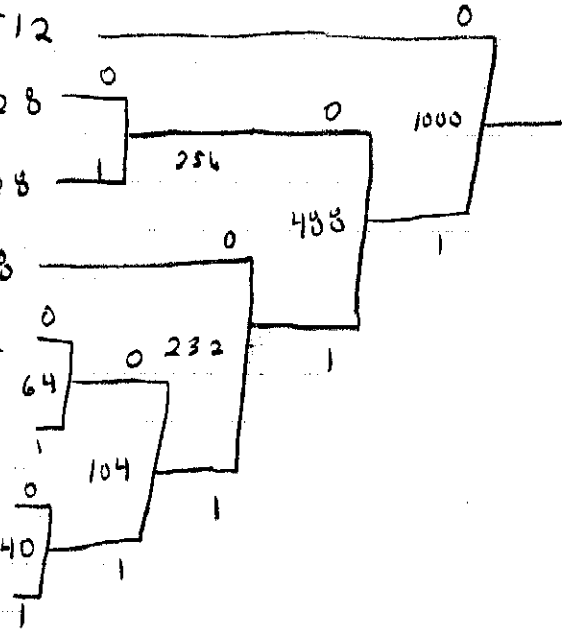
$$R^{(Nm)}_{subopt} = R^{(N)}_{Huff}$$

Since Huffman coding is optimal,

$$R^{(Nm)}_{Huff} \le R^{(Nm)}_{subopt} = R^{(N)}_{Huff}.$$

2)(a) • Form blocks

	$p(\cdot) * 1000$	Codewords
000	512	0
001	128	100
010	128	101
100	128	110
011	32	11100
110	32	11101
101	32	11110
111	8	11111



• $R = \frac{1}{1000}\left(1\cdot 512 + 3\cdot 128\cdot 3 + 5\cdot 32\cdot 3 + 5\cdot 8\cdot 1\right)$

$= 2.184$ output bits/block

$= 0.728$ output bits/input bit

$H(X) = H_2(0.8) = 0.72193$

Know $H(X) \le R_{Huff} \le H(X) + 1/N$

$0.72193 \le 0.728 \le 1.05$

- Form blocks

	$p(\cdot)*1000$	$\ell(\cdot) = \lceil -\log_2 p(\cdot) \rceil$	code
000	512	1	0
001	128	3	100
010	128	3	101
100	128	3	110
011	32	5	11100
110	32	5	11101
101	32	5	11110
111	8	7	1111111

start with Huffman - adjust last codeword

① correct lengths

② prefix-free

$\Longrightarrow$ Shannon-Fano code

$$R_{S-F} = \frac{1}{1000}\left(1 \cdot 512 + 3 \cdot 128 \cdot 3 + 5 \cdot 32 \cdot 3 + 7 \cdot 8 \cdot 1\right)$$

$= 2.2$ output bits/block

$= 0.733$ output bits/input bit

Know $\quad R_{HUFF} \le R_{S-F} \le H(X) + 1/N$

$0.728 \le 0.733 \le 1.05$

(b)

$$H((X_i)) = \lim_{n\to\infty} H(X_n \mid X_1, X_2, \ldots, X_{n-1})$$

$$= \lim_{n\to\infty} H(X_2 \mid X_1) \qquad (\text{this case})$$

$$= H(X_2 \mid X_1)$$

$$= E_{X_1}\left[H(X_2 \mid X_1 = x_1)\right]$$

$$= \mathcal{H}_2(0.9)\cdot 0.8 + \mathcal{H}_2(0.6)\cdot 0.2$$

$$= 0.469\cdot 0.8 + 0.97095\cdot 0.2$$

$$= 0.569$$

$$\lim_{n\to\infty} R_{Huff} = \lim_{n\to\infty} R_{s-f} = H((X_i)) = 0.569 \ \text{output bits}/_{\text{input bit}}$$

$$0.569 < 0.72193$$

Correlated sources are easier to code.

3)(a) $I(X;Y|Z)=0$

<u>False</u>

Suppose $H(X)>0$, $Y=X$, and $Z=0$ (always)

Then $I(X;Y|Z)=I(X;Y)=H(X)>0$.

(b) $H(X|Z)=H(X)$

<u>False</u>

Let $Z=Y=X$ and $H(X)>0$

Then $H(X|Z)=0 \neq H(X)$

(c) $H(X|Y,Z)=H(X|Y)$

<u>True</u>

$$0=I(X;Z|Y)=H(X|Y)-H(X|Y,Z)$$

(d) $I(X;Z)\leq H(Y)$

<u>True</u>

$$I(X;Z)\overset{\text{Data Proc. Theorem}}{\leq} I(X;Y)=H(Y)-H(Y|X)\leq H(Y)$$

(e) $H(Y|Z) \le H(X|Z)$

<u>False</u>

Suppose Y is independent of X and Z is independent of Y and $H(Y) > H(X)$

Then $H(Y|Z) = H(Y) > H(X) = H(X|Z)$

(f)

$$I(X;(Y,Z)) = I(X;Y)$$

<u>True</u>

$$\begin{aligned} I(X;(Y,Z)) &= H(X) - H(X|Y,Z) \\ &= H(X) - H(X|Y) \qquad \text{(from (c))} \\ &= I(X;Y) \end{aligned}$$

4) <u>Rough Idea:</u>

Getting the rough idea is pretty easy: since $A_\varepsilon^{(N)}$ contains essentially all of the probability as $N \to \infty$.

$$A_\varepsilon^{(N)} \subseteq B_N \quad \text{(roughly)}$$

and thus

$$|A_\varepsilon^{(N)}| \le |B_N| \le 2^{6\varepsilon N + 10N}$$

$$\Rightarrow 2^{N(H(x)-\varepsilon)} \le 2^{N(6\varepsilon+10)}$$

Since ε is arbitrary, $H(x) \le 10$.

<u>Rigorize:</u>

Suppose $H(x) > 10$. Now consider sequences in B_N that are also in $A_\varepsilon^{(N)}$ for any $\varepsilon > 0$.

$$P(B_N \cap A_\varepsilon^{(N)}) \le \underbrace{2^{N(6\varepsilon+10)}}_{\substack{\text{number}\\ \text{of}\\ \text{sequences}}} \underbrace{2^{-N(H(x)-\varepsilon)}}_{\substack{\text{max prob of}\\ \text{a sequence in } A_\varepsilon^{(N)}}}$$

$$= 2^{N(7\varepsilon + 10 - H(x))}$$

$$< \delta_1, \quad N \text{ large} \qquad (\text{because } 10 - H(x) < 0)$$

But $P(A_\varepsilon^{(N)}) > 1 - \delta_2$ for any $\delta_2 > 0$, N large and thus

$$P(B_N) = \underbrace{P(B_N \cap A_\varepsilon^{(N)})}_{<\delta_1} + \underbrace{P(B_N \cap A_\varepsilon^{(N)c})}_{<\delta_2}$$

Since δ_1, δ_2 are arbitrary, $P(B_N) \not\to 1$. Contradiction.

5)

(a) Use construction from class. Pick any $R > H(x)$ and $\delta > 0$. Pick $\varepsilon = \frac{R - H(x)}{2}$.

Method

1. Encode sequences in $A_\epsilon^{(n)}$ with less than $\lceil N(H(x)+\varepsilon)\rceil$ bits. (possible by AEP, part 3)

2. Encode sequences in $A_\epsilon^{(n)c}$ with $(0, 0, \dots, 0)$.

Thus, code length $\leq N(H(x)+\varepsilon)+1$

$$\text{rate} \leq H(X)+\varepsilon + 1/N < R \quad \text{for } N \geq N_1 = \frac{2}{R - H(x_i)}$$

and, by the AEP, $\exists N_2$ s.th. $\forall N \geq N_2$,

$$P_e \leq P\left(A_\epsilon^{(n)c}\right) < \delta.$$

Thus, choose $N = \max(N_1, N_2)$ and we have the code.

(b) Suppose $R < H(x)$ and pick any $\varepsilon > 0$. Lower bound: consider only errors in the typical set, which has probability

$$P(A_\epsilon^{(n)}) > 1-\varepsilon \quad \text{for } N \text{ large enough}$$

We can only code 2^{NR} typical sequences each of maximal probability $2^{-N(H(x)-\varepsilon)}$

Thus,

$$P_e > 1-\varepsilon - 2^{NR}2^{-N(H(x)-\varepsilon)}$$

$$= 1-\varepsilon - 2^{N(R-H(x)+\varepsilon)}$$

(c)

Choose $\varepsilon = \frac{H(x)-R}{2}$. For any $\delta > 0$, $\exists N_1$ s.t. $\forall N \geq N_1$,

$$P_e > 1 - \frac{\delta}{2} - 2^{N(R-H(x)+\varepsilon)}$$

and $\exists N_2$ s.t.

$$2^{N(R-H(x)+\varepsilon)} < \delta/2$$

$\Rightarrow$ $P_e > 1-\delta$ for $\forall \delta > 0$ and N large enough

$\Rightarrow$ $\lim_{N\to\infty} P_e = 1$.

6) (a) From class, we know this is simply given by

$$H(X) = -\sum_{x \in \mathcal{X}} p_X(x) \log_2 p_X(x)$$

$$= 1.75$$

(b) We need to encode the entire typical set of size

$$2^{NH(X)} = 2^{1.75N}$$

How many sequences of length k have $\le k/4$ 1's? Assume we have an IID binary source (Y_i) with $P(Y_i=0) = 3/4$, $P(Y_i=1) = 1/4$.
Let $A_Y^{(k)}$: set of sequences with $\le k/4$ 1's.

$$1 \ge P(A_Y^{(k)}) = \sum_{\underline{x} \in A_Y^{(k)}} P_{\underline{X}}(\underline{x})$$

$$\ge |A_Y^{(k)}| \, 2^{-k \mathcal{H}_2(1/4)}$$

(since all sequences in $A_Y^{(k)}$ have higher probability than those with approximately $k/4$ 1's).

$$\Rightarrow |A_Y^{(k)}| \le 2^{k \mathcal{H}_2(1/4)}$$

But $|A_Y^{(k)}| \ge 2^{k \mathcal{H}_2(1/4)}$ (it includes the typical set)

$\Rightarrow |A_Y^{(k)}| \approx 2^{k \mathcal{H}_2(1/4)}$. We must have $2^{k \mathcal{H}_2(1/4)} \ge 2^{1.75N}$

$$\Rightarrow k/N \ge \frac{1.75}{\mathcal{H}_2(1/4)} = 1.75/0.81 = 2.16.$$

7)

Let C = capacity of the cascade

For any $p_X(x)$,

$$I(X;Z) \le I(X;Y) \quad \text{(Data processing inequality)}$$

$$\Rightarrow \max_{p_X(x)} I(X;Z) \le \max_{p_X(x)} I(X;Y)$$

$$\Rightarrow C \le C_1$$

Suppose $C > C_2$, then $\exists\, \hat{p}_X(x)$ (with associated $\hat{p}_Y(y)$) s. th.

$$C = I_{\hat{p}}(X;Z) > C_2 = \max_{p_Y(y)} I(Y;Z)$$

But, by the data processing inequality,
$I_{\hat{p}}(Y;Z) \ge I_{\hat{p}}(X;Z)$

$$\Rightarrow I_{\hat{p}}(Y;Z) > \max_{p_Y(y)} I(Y;Z)$$

and thus our supposition is false $\Rightarrow C \le C_2$.

$$\Rightarrow C \le \min(C_1, C_2)$$

8) The key here is rigor. Since the channel is a DMC, I know:

$$C = \max_{p_X(x)} I(X;Y)$$

$$I(X;Y) = H(Y) - H(Y|X)$$

$$\leq 1 - E_X\left[H(Y|X=x)\right] \qquad \text{(since } Y \text{ is binary)}$$

$$= 1 - \left[p_X(0)\mathcal{H}_2(p) + p_X(1) + p_X(2)\mathcal{H}_2(p)\right]$$

$$\leq 1 - \mathcal{H}_2(p) \qquad \text{(since } \mathcal{H}_2(p) \leq 1 \text{ and } p_X(0)+p_X(1)+p_X(2)=1\text{)}$$

which can be achieved with

$p_X(0) = 1/2$

$p_X(1) = 0$

$p_X(2) = 1/2$

Thus,

$$C = 1 - \mathcal{H}_2(p)$$

9)

(a)

$$H(Y \mid X = x) = \mathcal{H}_2(p), \quad \forall x$$

Thus, $I(X;Y) = H(Y) - \mathcal{H}_2(p) \leq 1 - \mathcal{H}_2(p)$

achieved by $P(X_k = 0) = 1/2$, $P(X_k = 1) = 0$, $P(X_k = 2) = 1/2$.

$$\Rightarrow C = 1 - \mathcal{H}_2(p)$$

(b) $H(Y \mid X = x) = \mathcal{H}_2(p), \quad \forall x$

$P(Y = 0) = p$ (regardless of $p_X(x)$)

$P(Y = 1) = (1-p)$ " " "

$$\Rightarrow C = \mathcal{H}_2(p) - \mathcal{H}_2(p) = 0$$

(c) $H(Y \mid X = x) = \log_2 3, \quad \forall x$

$$\Rightarrow I(X;Y) = H(Y) - \log_2 3$$

$$\leq 2 - \log_2 3$$

which is achieved when $p_X(x) = 1/4, \ \forall x$

$$\Rightarrow C = 2 - \log_2 3$$

$$= 0.415$$

10)

(a)

Choose suboptimal decision rule:

$$\hat{X}(y) = \arg\max_{x} p_X(x)$$

$$
\begin{aligned}
P_e &= P\left(X \neq \arg\max_{x} p_X(x)\right) \\
&= 1 - \max_{x} p_X(x) \\
&\leq 1 - \frac{1}{|\mathcal{X}|} \\
&= \frac{|\mathcal{X}|-1}{|\mathcal{X}|}
\end{aligned}
$$

The optimal decision rule must be at least as good.

(b)

$$F_{|\mathcal{X}|}(p) = H_2(p) + p\log_2(|\mathcal{X}|-1)$$

$$\frac{d}{dp} F_{|\mathcal{X}|}(p) = -\log_2\frac{p}{(1-p)} + \log_2(|\mathcal{X}|-1) = 0$$

$$\Rightarrow \frac{p}{(1-p)} = |\mathcal{X}|-1$$

$$\Rightarrow p = \frac{|\mathcal{X}|-1}{|\mathcal{X}|}$$

(second derivative is indeed negative, also).

(c) I will give the rule,

$$I(X;Y) = H(X)$$

$$\Rightarrow H(X|Y) = 0$$

$$\Rightarrow H(X|Y=y) = 0 \qquad \forall y$$

$$\Rightarrow \exists\, \tilde{x}(y) \text{ s. th. } p_{X|Y}(\tilde{x}(y)|y) = 1 \qquad \forall y$$

$\Rightarrow$ choose $\hat{X}(y) = \tilde{x}(y)$

$$\Rightarrow P_e = 0.$$