MATH 4280 – Exam 2 – 4/2/08

Hurley

Name: Solutions.

All answers must be justified to receive full credit.

1. (10 points) Let $\mathcal{S}$ be a source, C_1 be a binary Shannon-Fano code for $\mathcal{S}$ and C_2 be an optimal code for $\mathcal{S}$.

   Put the following in non-decreasing order: $L(C_1)$, $L(C_2)$, $H_r(\mathcal{S})$ and $H_r(\mathcal{S})+1$.

$$H_r(\mathcal{S}) \leq L(C_2) \leq L(C_1) \leq H_r(\mathcal{S})+1 .$$

2. (25 points)

   (a) State Shannon's First Theorem for a source $\mathcal{S}$.

   There exists an r-ary encoding of $\mathcal{S}$ with average word-length arbitrarily close to the base r entropy of $\mathcal{S}$, $H_r(\mathcal{S})$.

   (b) Let $\mathcal{S}$ be a source with alphabet $S = \{s_1, s_2\}$ and probabilities $p_1 = \frac{11}{12}$ and $p_2 = \frac{1}{12}$.

   Construct (or describe how to construct) a binary encoding of $\mathcal{S}$ so that the difference between the average word-length of the encoding and the binary entropy of $\mathcal{S}$ is less than or equal to $\frac{1}{3}$.

   For any Shannon-Fano code (or optimal code) C_n for $\mathcal{S}^n$: $H_r(\mathcal{S}) \leq \frac{L(C_n)}{n} < H_r(\mathcal{S}) + \frac{1}{n}$.

   So for $n=3$, $\frac{L(C_3)}{3} - H_r(\mathcal{S}) < \frac{1}{3}$.

   So any Shannon-Fano code or optimal code C_3 for $\mathcal{S}^3$ will have the desired property.

3. (30 points) Let Γ be a binary channel with input probabilities $\{\frac{1}{4}, \frac{3}{4}\}$ and channel matrix $\begin{bmatrix} \frac{1}{2} & \frac{1}{2} \\ \frac{2}{3} & \frac{1}{3} \end{bmatrix}$.

(a) Compute the binary entropies of the input source $\mathcal{A}$, the output source $\mathcal{B}$ and $H(\mathcal{B}|\mathcal{A})$, the equivocation of $\mathcal{B}$ with respect to $\mathcal{A}$.

$$H_2(A) = H_2\left(\tfrac{1}{4}\right) = \tfrac{1}{4}\log_2 4 + \tfrac{3}{4}\log_2\left(\tfrac{4}{3}\right) \approx .8113.$$

To find $H(B)$ we need the output probabilities q and $\bar{q}$. Since $(p\ \bar{p})M = (q\ \bar{q})$,

$$(q\ \bar{q}) = \left(\tfrac{1}{4}\ \tfrac{3}{4}\right)\begin{bmatrix} \frac{1}{2} & \frac{1}{2} \\ \frac{2}{3} & \frac{1}{3} \end{bmatrix} = \left(\tfrac{1}{8}+\tfrac{1}{2}\quad \tfrac{1}{8}+\tfrac{1}{4}\right) = \left(\tfrac{5}{8}\ \tfrac{3}{8}\right)$$

So $H(B) = H\left(\tfrac{5}{8}\right) = \tfrac{5}{8}\log_2\left(\tfrac{8}{5}\right) + \tfrac{3}{8}\log_2\left(\tfrac{8}{3}\right) \approx .9544$

To find $H(B|A)$, we want the joint probabilities $R_{ij} = p_i P_{ij}$. $(R_{ij}) = \begin{pmatrix} \frac{1}{8} & \frac{1}{8} \\ \frac{1}{2} & \frac{1}{4} \end{pmatrix}$ and $H(B|A) = \sum_{i,j} R_{ij}\log_2\left(\frac{1}{P_{ij}}\right)$

$\approx .9387$

(b) Compute the mutual information of $\mathcal{A}$ and $\mathcal{B}$, $I(\mathcal{A}, \mathcal{B})$.

$$I(A,B) = H(B) - H(B|A)$$
$$\approx .9544 - .9387 = .0157$$

(c) Give three different formulas for the mutual information of $\mathcal{A}$ and $\mathcal{B}$ in terms of the system entropies of Γ.

(i) $I(A,B) = H(B) - H(B|A)$

(ii) $I(A,B) = H(A) - H(A|B)$

(iii) $I(A,B) = H(A) + H(B) - H(A,B)$.

(d) Compute $H(\mathcal{A}|\mathcal{B})$, the equivocation of $\mathcal{A}$ with respect to $\mathcal{B}$, and $H(\mathcal{A}, \mathcal{B})$, the joint entropy of $\mathcal{A}$ and $\mathcal{B}$.

From (ii) in part (c),

$$H(A|B) = H(A) - I(A,B) \approx .7956$$

From (iii) in part (c)

$$H(A,B) = H(A) + H(B) - I(A,B) \approx 1.75.$$

Alternately, $H(A,B) = \sum R_{ij}\log_2(R_{ij}) = 7/4 = 1.75$

Explictly,

$$\mathcal{S}^3 = \{ s_1^3, s_1^2 s_2, s_1 s_2 s_1, s_1 s_2^2, s_2 s_1^2, s_2 s_1 s_2, s_2^2 s_1, s_2^3 \}$$

with probabilities $\frac{1331}{1728}, \frac{121}{1728}, \frac{121}{1728}, \frac{11}{1728}, \frac{121}{1728}, \frac{11}{1728}, \frac{11}{1728}, \frac{1}{1728}$

So a binary Shannon-Fano code C_3 for $\mathcal{S}^3$ has word-lengths

$$\left\lceil \log_2 \frac{1728}{1331} \right\rceil = 1, \quad \left\lceil \log_2 \frac{1728}{121} \right\rceil = 4, \quad \left\lceil \log_2 \frac{1728}{11} \right\rceil = 8$$

and $\lceil \log_2 1728 \rceil = 11$.

So the average word-length of a binary Shannon-Fano code for $\mathcal{S}^3$ per element of $\mathcal{S}^3$ is

$$\frac{1331}{1728} \cdot 1 + 3 \cdot \frac{121}{1728} \cdot 4 + 3 \cdot \frac{11}{1728} \cdot 8 + \frac{1}{1728} \cdot 11$$

$$= \frac{3058}{1728} = \frac{1529}{864}.$$

And the average word-length per element of $\mathcal{S}$ is

$$\frac{1}{3} L(C_3) = \frac{1}{3} \cdot \frac{1529}{864} = \frac{1529}{2592} \approx .5899.$$

The binary entropy of $\mathcal{S}$ is

$$H_2(\mathcal{S}) = \frac{11}{12} \log_2 \left(\frac{12}{11}\right) + \frac{1}{12} \log_2 (12) = .4138$$

and

$$\frac{L(C_3)}{3} - H(\mathcal{S}) \approx .1761 \le \frac{1}{3}, \text{ as desired.}$$

Note one can also use a binary Huffman code. In that case a code for $\mathcal{S}^2$ will suffice, although $\mathcal{S}^3$ works as well.

4. (20 points) Let Γ be a channel with input source $\mathcal{A}$ and output source $\mathcal{B}$. Suppose that all the input probabilities are positive, i.e., $p_i > 0$ for all i, and suppose that $H(\mathcal{B}|\mathcal{A}) = 0$.

(a) What does $H(\mathcal{B}|\mathcal{A}) = 0$ mean from the point of view of the sender?

Equivalent statements, so either is O.K.

It means the sender has no uncertainty about the output of Γ.

It means knowing $\Gamma(a)$ in addition to a gives no additional information.

(b) Show that in this case Γ actually defines a function from the input alphabet A to the output alphabet B, and show how to define the function.

Since
$$0 = H(B|A)$$
$$= \sum_{i,j} R_{ij} \log_2\left(\frac{1}{P_{ij}}\right)$$
$$= \sum_i p_i \sum_j P_{ij} \log\left(\frac{1}{P_{ij}}\right).$$

Since each $p_i \sum_j P_{ij} \log\left(\frac{1}{P_{ij}}\right) \geq 0$ and they sum to 0,

$p_i \sum_j P_{ij} \log\left(\frac{1}{P_{ij}}\right) = 0$ for all i's.

Since each $p_i > 0$, For all i,
$$\sum_j P_{ij} \log\left(\frac{1}{P_{ij}}\right) = 0.$$

Again each $P_{ij} \log\left(\frac{1}{P_{ij}}\right) \geq 0$, and they sum to 0, so

$P_{ij} \log\left(\frac{1}{P_{ij}}\right) = 0 \Rightarrow P_{ij} = 0$ or 1 for all j.

Since $\sum_j P_{ij} = 1$, for each i, exactly one $P_{ij} = 1$ and the rest are 0. Denote the j s.t. $P_{ij} = 1$ by i^*.

So for a given i, $P_{ii^*} = 1$ and $P_{ij} = 0$ for $j \neq i^*$.

Define a function $f: A \to B$, by

$f(a_i) = b_{i*}$ for all $a_i \in A$.

Since,

$P_{ii*} = 1$, $pr\{b = b_{i*} \mid a = a_i\} = 1$.

that is if $a = a_i$, $\Gamma(a_i) = b_{i*}$, in all cases.

thus the channel Γ is the same as the function f.

The above proof is very much from first principles, and is not so intuitive. The answer to part (a), "The sender has no uncertainty about the output" makes part (b) at least intuitively clear. A more conceptual proof would be:

$$0 = H(B|A) = \sum_{i=1}^{r} p_i H(B|a_i).$$

Since $p_i H(B|a_i) \geq 0$, this implies $p_i H(B|a_i) = 0$ for all i, and since $p_i > 0$, $H(B|a_i) = 0$.

$H(B|a_i) = 0$ iff the probabilities $P_{i1}, P_{i2}, \cdots, P_{in}$, on B have one $P_{ij} = 1$ and the rest equal to 0, i.e., for each i, there exists an i^*, $1 \leq i^* \leq s$, s.t. $P_{ii*} = 1$ and $P_{ij} = 0$ for all $1 \leq j \leq s$, $j \neq i^*$.

Now procede as from the top of this page.

5. (15 points) Prove that for a binary symetric channel an ideal observer rule has $Pr_C = \max\{p, \bar{p}, P, \bar{P}\}$.

Since the channel is binary there are 4 possible decison rules. Using $\{0,1\}$ as the alphabet for the input and output sources, the 4 decision rules are:

$$\Delta_1(0) = 0, \quad \Delta_1(1) = 0 ; \qquad \Delta_2(0) = 0, \quad \Delta_2(1) = 1 ; \qquad \Delta_3(0) = 1, \quad \Delta_3(1) = 0 ; \qquad \Delta_4(0) = 1, \quad \Delta_4(0) = 1 .$$

For a binary symmetric channel the matrix of joint probabilities is

$$(R_{ij}) = \begin{bmatrix} pP & p\bar{P} \\ \bar{p}\bar{P} & \bar{p}P \end{bmatrix}.$$

For any decision rule $\Delta(b_j) = a_{j*}$, the probability of correct decoding $Pr_c = \sum_j R_{j*j}$.

So
$$Pr_c(\Delta_1) = pP + p\bar{P} = p$$
$$Pr_c(\Delta_2) = pP + \bar{p}P = P$$
$$Pr_c(\Delta_3) = \bar{p}\bar{P} + p\bar{P} = \bar{P}$$
$$Pr_c(\Delta_4) = \bar{p}\bar{P} + \bar{p}P = \bar{p}$$

The ideal observer rule Δ_{ideal} is defined to maximize Pr_c. Therefore $Pr_c(\Delta_{ideal}) = \max_{\Delta} \{Pr_c(\Delta)\}$, where the maximum is taken over the set of all decision rules.

So $Pr_c(\Delta_{ideal}) = \max\{Pr_c(\Delta_1), Pr_c(\Delta_2), Pr_c(\Delta_3), Pr_c(\Delta_4)\} = \max\{p, P, \bar{P}, \bar{p}\}$ ∎