EE/Ma 126a Information Theory - Final Exam

Ling Li, ling@cs.caltech.edu

December 6, 2000

F.1 *Instantaneous Codes.* We can use the Kraft's inequality to judge whether there exits an instantaneous code for a given set of description lengths.

(a) $\{l(x)\} = \{1, 3, 3, 4, 4, 4, 5, 5\}$. Thus

$$\sum_{x=1}^{8} 2^{-l(x)} = 2^{-1} + 2 \times 2^{-3} + 3 \times 2^{-4} + 2 \times 2^{-5} = 1.$$

So there exists an instantaneous code for this set of lengths. By setting the probability of each symbol x to $2^{-l(x)}$, i.e., $p(x) = 2^{-l(x)}$, we can use below Huffman code as an example:

$$\boxed{\{\alpha(x)\} = \{0, 110, 111, 1010, 1011, 1000, 10010, 10011\}}.$$

This is a prefix-free code thus an instantaneous code.

(b) In fact, the original code is suffix-free. So, just reversing all the codewords gives a prefix-free and instantaneous code with the same description lengths:

$$\boxed{\{\alpha(x)\} = \{1, 01, 001, 00001, 000001, 0000001, 00000001, 00000000\}}.$$

(c) $\{l(x)\} = \{2, 3, 3, 3, 4, 4, 4, 4\}$. Thus

$$\sum_{x=1}^{8} 2^{-l(x)} = 2^{-2} + 3 \times 2^{-3} + 4 \times 2^{-4} = \frac{7}{8} < 1.$$

So there exists an instantaneous code for this set of lengths. Note that only changing $\alpha(2)$ to 101 and $\alpha(3)$ to 100 gives a prefix-free and instantaneous code:

$$\boxed{\{\alpha(x)\} = \{00, 101, 100, 011, 1100, 1101, 1110, 1111\}}.$$

F.2 *Channel Capacity.* The transition matrix[1] of this channel is

$$p(y|x) = \begin{pmatrix} 1/4 & 1/16 & 7/16 & 1/4 \\ 1/16 & 1/4 & 1/4 & 7/16 \\ 7/16 & 1/16 & 1/4 & 1/4 \\ 1/4 & 7/16 & 1/16 & 1/4 \\ 1/4 & 7/16 & 1/4 & 1/16 \end{pmatrix}.$$

[1] See textbook p.190, defintions of symmetric channel and weakly symmetric channel.

Since 1) every row $p(\cdot|x)$ is a permutation of every other row, and 2) the summation over each column $\sum_x p(y|x)$ is the same as that on any other column, i.e., 5/4, this channel is weakly symmetric. So the capacity of the channel is

$$
\begin{aligned}
C &= \log|\mathcal{Y}| - H(1/4, 1/16, 7/16, 1/4) \\
&= \log 4 - \left(\frac{1}{4}\log 4 + \frac{1}{16}\log 16 + \frac{7}{16}\log\frac{16}{7} + \frac{1}{4}\log 4\right) \\
&= \boxed{\frac{7}{16}\log 7 - 1 \approx 0.2282},
\end{aligned}
$$

and the maximizing input distribution is

$$
\boxed{p^*(x) = \frac{1}{5}, x = 1..5},
$$

i.e., uniform distribution.

F.3 *Wandering robots in the old west.* From the description of that robot, X_n only depends on X_{n-1} and $p(X_n|X_{n-1})$ is determined by the parity of n. Thus $\{X_n\}$ is a Markov process and has a time variant probability transition matrix $P^{(n)}$:

$$
P^{(n)} = \left(P_{ij}^{(n)}\right) = (\Pr\{X_n = j | X_{n-1} = i\}) = \begin{cases} \begin{pmatrix} 0 & 0.4 & 0.6 \\ 0.6 & 0 & 0.4 \\ 0.4 & 0.6 & 0 \end{pmatrix}, & \text{when } n \text{ is odd;} \\ \begin{pmatrix} 0.4 & 0.2 & 0.4 \\ 0.4 & 0.4 & 0.2 \\ 0.2 & 0.4 & 0.4 \end{pmatrix}, & \text{when } n \text{ is even.} \end{cases}
$$

(a) When $n > 1$ and n is odd, this Markov process gives

$$
\boxed{H(X_n|X_1, \ldots, X_{n-1}) = H(X_n|X_{n-1}) = H(0.4, 0.6) \approx 0.9710}.
$$

When $n > 1$ and n is even,

$$
\boxed{H(X_n|X_1, \ldots, X_{n-1}) = H(X_n|X_{n-1}) = H(0.2, 0.4, 0.4) \approx 1.5219}.
$$

When $n = 1$, since X_0 is uniformly distributed, we have

$$
p(x_1) = \sum_{x_0} p(x_1|x_0)p(x_0) = \frac{1}{3}(0.4 + 0.6) = \frac{1}{3}.
$$

Thus

$$
\boxed{H(X_n|X_1, \ldots, X_{n-1}) = H(X_1) = \log 3 \approx 1.5850}.
$$

(b) When n is odd,

$$
\begin{aligned}
H(X_1, X_2, \ldots, X_n) &= \sum_{i=1}^{n} H(X_i|X_1, \ldots, X_{i-1}) \\
&= \frac{n-1}{2}H(0.4, 0.6) + \frac{n-1}{2}H(0.2, 0.4, 0.4) + H(X_1).
\end{aligned}
$$

When n is even,

$$\begin{aligned} H(X_1, X_2, \ldots, X_n) &= \sum_{i=1}^{n} H(X_i | X_1, \ldots, X_{i-1}) \\ &= \frac{n-2}{2} H(0.4, 0.6) + \frac{n}{2} H(0.2, 0.4, 0.4) + H(X_1). \end{aligned}$$

Thus

$$\begin{aligned} \lim_{n\to\infty, n \text{ odd}} \frac{H(X_1, X_2, \ldots, X_n)}{n} &= \lim_{n\to\infty, n \text{ even}} \frac{H(X_1, X_2, \ldots, X_n)}{n} \\ &= \frac{1}{2} \left[H(0.4, 0.6) + H(0.2, 0.4, 0.4) \right]. \end{aligned}$$

So the entropy rate is

$$\boxed{H(\mathcal{X}) = \lim_{n\to\infty} \frac{H(X_1, X_2, \ldots, X_n)}{n} = \frac{1}{2} \left[H(0.4, 0.6) + H(0.2, 0.4, 0.4) \right] \approx 1.2464}.$$

(c) As described before, $\boxed{\text{the robot's movement is a Markov process}}$.

However, $\boxed{\text{it is not stationary}}$, since a stationary Markov process should have a time invariant probability transition matrix P.

Consider $\{X_{2n}\}$ which is a sequence of the robot positions after even numbers of seconds. This can be regarded as a stationary Markov process with probability transition matrix

$$P = P^{(2n-1)} P^{(2n)} = \begin{pmatrix} 0 & 0.4 & 0.6 \\ 0.6 & 0 & 0.4 \\ 0.4 & 0.6 & 0 \end{pmatrix} \begin{pmatrix} 0.4 & 0.2 & 0.4 \\ 0.4 & 0.4 & 0.2 \\ 0.2 & 0.4 & 0.4 \end{pmatrix} = \begin{pmatrix} 0.28 & 0.4 & 0.32 \\ 0.32 & 0.28 & 0.4 \\ 0.4 & 0.32 & 0.28 \end{pmatrix}.$$

And this Markov process is also irreducible and aperiodic. So the distribution of X_{2n} tends to the stationary distribution

$$\left(\frac{1}{3}, \frac{1}{3}, \frac{1}{3} \right) \tag{1}$$

as $n \to \infty$. The stationary distribution is the (real) eigenvector of P.

For $\{X_{2n+1}\}$ which is a sequence of the robot positions after odd numbers of seconds, similar analysis gives the distribution of X_{2n+1} also tends to (1) as $n \to \infty$.

So $\boxed{\text{the distribution of } X_n \text{ tends to be uniform as } n \to \infty}$.

F.4 *Mutual information.*

(a) $I(X;Y) \geq 0$ with the equality iff X and Y are independent. We can set

$$p(y|x) = p(y) = P_y$$

to achieve the equality $I(X;Y) = 0$, where P_y is constant and $P_{b_0} + P_{b_1} = 1$. Thus $\boxed{p_*(y|x) = P_y}$ and $\boxed{I_*(X;Y) = 0}$. Note that $p_*(y|x)$ is not unique here.

(b) $I(X;Y) = H(Y) - H(Y|X) \leq H(Y) \leq \log|\mathcal{Y}| = 1$ with the equality iff 1) Y is a function of X, and 2) Y is uniformly distributed. We can verify that for function

$$f(x) = \begin{cases} b_0, & x = a_0; \\ b_1, & x = a_1, a_2, a_3. \end{cases}$$

and $Y = f(X)$, we have

$$p(y = b_0) = p(x = a_0) = \frac{1}{2}, \quad p(y = b_1) = 1 - p(y = b_0) = \frac{1}{2}.$$

Thus with

$$\boxed{p^*(y|x) = \begin{cases} 1, & \text{if } y = f(x); \\ 0, & \text{otherwise.} \end{cases}}$$

we get the maximum of $I(X;Y)$, i.e., $\boxed{I^*(X;Y) = 1}$.

(c) From Theorem 2.74 of the textbook, $I(X;Y)$ is a convex function of $p(y|x)$ for fixed $p(x)$. Thus

$$\boxed{\lambda = \lambda I^*(X;Y) + (1-\lambda) I_*(X;Y) \geq I(X;Y)|_{\lambda p^*(y|x) + (1-\lambda) p_*(y|x)}}$$

for each $\lambda \in [0, 1]$.

F.5 *Predicting the Future.*

(a) Since $p = \Pr\{Y = X\}$, $1 - p = \Pr\{Y \neq X\}$, the Fano's inequality says

$$\boxed{H(X|Y) \leq H(p) + (1-p)\log(|\mathcal{X}| - 1) = H(p) + (1-p)\log 12}. \tag{2}$$

X is randomly chosen from a *fair* deck of cards, so X is uniformly distributed, $p(x) = \frac{1}{13}$. For

$$p(y|x) = \begin{cases} p, & \text{if } y = x; \\ \frac{1-p}{12}, & \text{otherwise.} \end{cases}$$

we have $p(y) = \sum_x p(y|x)p(x) = \frac{1}{13}$. Thus

$$p(x|y) = \frac{p(y|x)p(x)}{p(y)} = p(y|x).$$

Thus

$$\begin{aligned} H(X|Y) &= H(p, \frac{1-p}{12}, \ldots, \frac{1-p}{12}) = p\log\frac{1}{p} + 12 \times \frac{1-p}{12}\log\frac{12}{1-p} \\ &= p\log\frac{1}{p} + (1-p)\log\frac{1}{1-p} + (1-p)\log 12 \\ &= H(p) + (1-p)\log 12. \end{aligned}$$

So the upper bound (2) is sharp.

(b) Denote the set of all sequences x^n such that

$$\left| \frac{1}{n} \sum_{i=1}^{n} x_i^2 - E \right| \leq \epsilon$$

by $B_\epsilon^{(n)}$. Thus the number we want to bound is $\left|B_\epsilon^{(n)}\right|$. By WLLN, we get

$$\Pr\left\{\left(B_\epsilon^{(n)}\right)^c\right\} = \Pr\left\{\left|\frac{1}{n}\sum_{i=1}^{n} X_i^2 - E\right| > \epsilon\right\} \le \frac{\sigma^2}{\epsilon^2 n},$$

where $\sigma^2 = \sum_{x=1}^{m} p(x)(x^2 - E)^2 = \sum_{x=1}^{m} p(x)x^4 - E^2$.

For typical set $A_\epsilon^{(n)}$ defined in the textbook p.51, we have

$$\Pr\left\{\left(A_\epsilon^{(n)}\right)^c\right\} = \Pr\left\{\left|-\frac{1}{n}\sum_{i=1}^{n} \log p(X_i) - H(X_1)\right| > \epsilon\right\} \le \frac{\hat{\sigma}^2}{\epsilon^2 n},$$

where $\hat{\sigma}^2 = \sum_{x=1}^{m} p(x)\left(-\log p(x) - H(X_1)\right)^2 = \sum_{x=1}^{m} p(x)\log^2 p(x) - H^2(X_1)$.

Thus we have

$$\begin{aligned}\Pr\left\{A_\epsilon^{(n)} \cap B_\epsilon^{(n)}\right\} &= 1 - \Pr\left\{\left(A_\epsilon^{(n)}\right)^c \cup \left(B_\epsilon^{(n)}\right)^c\right\} \\ &\ge 1 - \Pr\left\{\left(A_\epsilon^{(n)}\right)^c\right\} - \Pr\left\{\left(B_\epsilon^{(n)}\right)^c\right\} \ge 1 - \frac{\sigma^2 + \hat{\sigma}^2}{\epsilon^2 n}. \qquad (3)\end{aligned}$$

Since for any $x^n \in A_\epsilon^{(n)}$, $p(x^n) \le 2^{-n(H(X_1)-\epsilon)}$, we get

$$\Pr\left\{A_\epsilon^{(n)} \cap B_\epsilon^{(n)}\right\} \le \left|A_\epsilon^{(n)} \cap B_\epsilon^{(n)}\right| \cdot 2^{-n(H(X_1)-\epsilon)}. \qquad (4)$$

Combining (3) and (4), we can bound $\left|B_\epsilon^{(n)}\right|$ from below:

$$\boxed{\left|B_\epsilon^{(n)}\right| \ge \left|A_\epsilon^{(n)} \cap B_\epsilon^{(n)}\right| \ge 2^{n(H(X_1)-\epsilon)}\left(1 - \frac{\sigma^2 + \hat{\sigma}^2}{\epsilon^2 n}\right)}.$$

F.6 *Huffman codes.* The Huffman procedure will construct a tree with each leaf is a symbol.

(a) Let α denote one symbol with Huffman description of length 2. Without loss of generality, the Huffman tree containing α as a second level leaf is shown in Fig. 1(a). From the

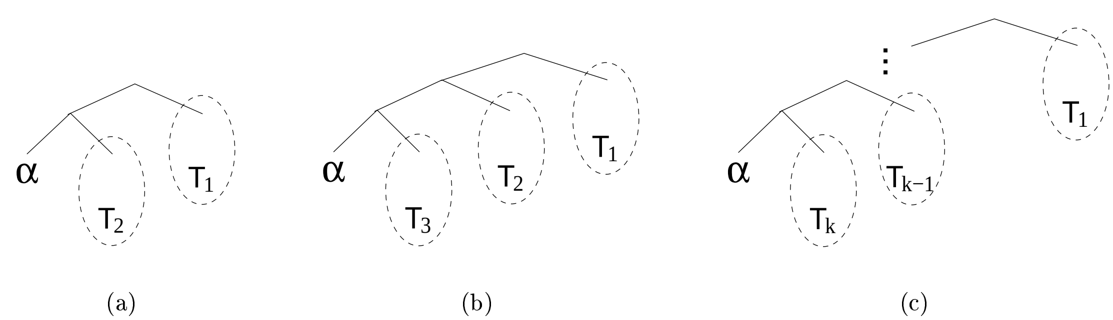

Figure 1: Huffman trees. $T_1, T_2, \ldots, T_k$ denote the subtrees.

Huffman procedure, we know the total probability of subtree T_1 is no less than that of α, i.e., $p(T_1) \geq p(\alpha) = p$; And the total probability of subtree T_2 is bigger than 0, i.e., $p(T_2) > 0$. Thus

$$\boxed{1 = p(T_1) + p(T_2) + p(\alpha) > 2p}.$$

(b) Now α is a third level leaf and the Huffman tree is in Fig. 1(b). $p(T_3) > 0$ and $p(T_1) \geq p(T_2) \geq p(\alpha) = p$. Thus

$$\boxed{1 = p(T_1) + p(T_2) + p(T_3) + p(\alpha) > 3}.$$

(c) Fig. 1(c) is the general Huffman tree with symbol α at the k-th level (thus has a Huffman description of length k). Here I claim

Lemma: For $i = 1, 2, \ldots, k$,

$$p(\alpha) + \sum_{j=0}^{i-1} p(T_{k-j}) > a_i \cdot p, \tag{5}$$

where a_i is the Fibonacci sequence defined in the problem.

Proof: Since $p(T_k) > 0$ and $p(T_{k-1}) \geq p(\alpha) = p$, we have

$$p(\alpha) + p(T_k) > p(\alpha) = a_1 \cdot p,$$

$$p(\alpha) + p(T_k) + p(T_{k-1}) > 2p(\alpha) = a_2 \cdot p.$$

Thus (5) holds for $i = 1, 2$.
Suppose for $i = t - 1$ and $i = t$ $(2 \leq t \leq k - 1)$, (5) holds. For $i = t + 1$, we have

$$p(T_{k-t}) \geq p(\alpha) + \sum_{j=0}^{t-2} p(T_{k-j}) > a_{t-1} \cdot p.$$

Thus

$$\begin{aligned} p(\alpha) + \sum_{j=0}^{t} p(T_{k-j}) &= p(\alpha) + \sum_{j=0}^{t-1} p(T_{k-j}) + p(T_{k-t}) \\ &> a_t \cdot p + a_{t-1} \cdot p \\ &= a_{t+1} \cdot p, \end{aligned}$$

i.e., (5) also holds for $i = t + 1$. Thus by induction, the lemma holds. □

Since $p(\alpha) + \sum_{j=0}^{k-1} p(T_{k-j}) = p(\alpha) + \sum_{j=1}^{k} p(T_k) = 1$, we have (let $i = k$ in the lemma)

$$\boxed{1 > a_k \cdot p}.$$

(d) If the Huffman description for that symbol is longer than k bits long, say, $(k+t)$ bits long, where $t > 0$, from (c) we have

$$1 > a_{k+t} \cdot p \geq a_{k+1} \cdot p,$$

or $1/p > a_{k+1}$, which contradicts with $a_k < 1/p \leq a_{k+1}$. So a symbol with probability p should have a Huffman description no longer than k bits long.