# Homework Set #4

Due 23 October 2008, 6pm, INR 031

## Problem 1 (Entropy Rates of Markov Chains)

The *entropy rate* of a stochastic process $\{X_i\}$ is

$$H(\mathcal{X}) = \lim_{n\to\infty} \frac{1}{n} H(X_1, X_2, \ldots, X_n)$$

when the limit exists.

(a) Find the entropy rate of the two-state Markov chain with transition matrix

$$P = \begin{bmatrix} 1-p_{01} & p_{01} \\ p_{10} & 1-p_{10} \end{bmatrix}.$$

(b) What values of $p_{01}$, $p_{10}$ maximize the entropy rate?

(c) Find the entropy rate of the two-state Markov chain with transition matrix

$$P = \begin{bmatrix} 1-p & p \\ 1 & 0 \end{bmatrix}.$$

(d) Find the maximum value of the entropy rate of the Markov chain of part (c). We expect that the maximizing value of $p$ should be less than $\frac{1}{2}$, since the state 0 permits more information to be generated than the 1 state.

(e) Let $N(t)$ be the number of allowable state sequences of length $t$ for the Markov chain of part (c). Find $N(t)$ and calculate

$$H_0 = \lim_{t\to\infty} \frac{1}{t} \log N(t).$$

[*H*int: Find a linear recurrence that expresses $N(t)$ in terms of $N(t-1)$ and $N(t-2)$. Why is $H_0$ an upper bound on the entropy rate of the Markov chain? Compare $H(0)$ with the maximum entropy found in part (d).]

## Problem 2 (Description Encoding)

We consider the case of encoding a binary sequence $x^n \in \{0,1\}^n$. We assume that the members of the sequence $x_1, x_2, \ldots, x_n$ are generated independently from Bernoulli distribution with probability $p$, where $p$ is unknown.

We will encode the sequence in two steps. In the first step, we estimate the distribution $p$. We first observe the entire sequence, count the number of ones (i.e. $k = \sum_{i=1}^{n} x_i$), and then describe this number.

(a) How many bits need to be reserved for the binary description of $k$? How many different sequences of length $n$ exist with $k$ ones? Label this number $N$.

(b) In the second stage of our algorithm, we encode one of the possible $N$ sequences. How many bits are needed for this description?

(c) Find a good upper bound on the total length of the description $l(x^n)$ for our procedure. You may use the following bound:

$$\sqrt{\frac{n}{8k(n-k)}} \leq \binom{n}{k} 2^{-nH(k/n)} \leq \sqrt{\frac{n}{\pi k(n-k)}}.$$

(d) If the length of the optimal code for the Bernoulli distribution corresponding to $\frac{k}{n}$ is $l^*(x^n)$, what is the cost of describing the sequence statistics (i.e. calculate $\frac{l(x^n)-l^*(x^n)}{l^*(x^n)}$). How does this quantity behave as $n \to \infty$?

## Problem 3 (Arithmetic Coding)

Consider the random variables $X_i$ with a ternary alphabet $\{A, B, C\}$, having probabilities $\{.2, .3, .5\}$. The source produces a sequence of $X_i$'s independently and identically distributed. As $X_i$'s are i.i.d., let's call the sequence $X^n$ from now on. Imagine that the source emits $ACCB$... and this sequence is to be encoded using arithmetic coding.

(a) What is the cumulative distribution function $F(X^n)$ for $n = 1$, i.e., the cumulative distribution function after the first symbol? What is the interval corresponding to the first symbol of the sequence ($A$)?

(b) What is the cumulative distribution function after the second symbol? What is the interval corresponding to $AC$?

(c) Find the binary representations of the corresponding intervals for (a) and (b) (Remember Shannon-Fano-Elias coding).

(d) Find the binary code representing $ACCB$ similarly.

(e) How many bits can be known for sure if it is not known how $ACCB$ continues?

## Problem 4 (Lempel-Ziv Algorithm)

(a) Give the parsing and encoding of 00000011010100000110101 using the LZ algortihm.

(b) Give a sequence for which the number of phrases in the LZ parsing grows as fast as possible

(c) Give a sequence for which the number of phrases in the LZ parsing grows as slowly as possible

(d) Let $X_i$ be a binary stationary Markov process with the transition matrix $\begin{bmatrix} \frac{1}{3} & \frac{2}{3} \\ \frac{1}{2} & \frac{1}{2} \end{bmatrix}$.

1. Find the stationary distribution of this Markov process ($[p_0, p_1]$).
2. Imagine that the Markov process is in the state 0. How many steps does it take on average for the process to return to the state 0 again? (Verfiy that it is equal to $\frac{1}{p_0}$)

3. Find the stationary distribution of the extended Markov process formed by considering blocks of length $n$ of $X_i$'s ($X_0^{n-1}$) instead of $X_i$'s. So the states of this extended Markov process ($X_0^{n-1}$) are $x_0^{n-1} = x_0 x_1 \cdots x_{n-1}$ where $x_i$'s are 0 or 1.
4. How many steps does it take on average for the extended Markov process to return to the state $x_0^{n-1}$ starting from the state $x_0^{n-1}$? (Use (d2) to at least guess the answer in order to continue if you didn't prove it.)
5. Consider each sequence which is to be encoded as a state of an extended Markov process and assume a LZ algorithm with infinite-length sliding window. Then to encode the block $x_0 x_1 \cdots x_{n-1}$, the last time we have seen these $n$ symbols should be communicated. Call it $R_n(x_0 x_1 \cdots x_{n-1})$. Explain that the requested average number of steps in (d4) is indeed $\mathbf{E}\{R_n(X_0 X_1 \cdots X_{n-1}) | (X_0 X_1 \cdots X_{n-1}) = x_0 x_1 \cdots x_{n-1}\}$.
6. Compute the entropy rate of this Markov process.
7. Verify the following inequalities and equalities.

$$\begin{aligned}
\lim_{n\to\infty} \frac{1}{n}\mathbf{E}l(X_0^{n-1}) &= \lim_{n\to\infty} \frac{1}{n}(\log R_n + 2\log\log R_n + O(1)) \\
&= \lim_{n\to\infty} \frac{1}{n} sum_{x_0^{n-1}} p(x_0^{n-1})\mathbf{E}(\log R_n(X_0^{n-1})|X_0^{n-1} = x_0^{n-1}) \\
&\le \lim_{n\to\infty} \frac{1}{n} sum_{x_0^{n-1}} p(x_0^{n-1}) \log \mathbf{E}(R_n(X_0^{n-1})|X_0^{n-1} = x_0^{n-1}) \\
&= H(\mathcal{X})
\end{aligned}$$

*Hint: As you do not have the maximum value of $R_n$, $\log R_n$ is not enough to encode $R_n$. you might need to encode and send the length of the encoded $R_n$ as well as the encoded $R_n$ itself.*