EE 231A  Handout #4, Problem Set 2
Information Theory  Thursday, October 7, 2004
Instructor: Rick Wesel  Due: Thursday, October 14, 2004

80 pts
Reading: Chapters 3 and 4 of *Elements of Information Theory*

Unless otherwise specified, points for a problem are evenly distributed over its parts.

1. (20 pts) *The AEP in action.*

In this problem we will compute the size and probability of the typical set for two different values of $n$ to see how $\Pr(A_\epsilon^{(n)})$ increases with $n$.

Specifically, consider a Bernoulli random variable $X$ with $p(1) = 3/4$ and $p(0) = 1/4$.

(a) (2 pts) Show that $H(X) = 2 - \frac{3}{4} \log 3$.

(b) (3 pts) Show that

$$\frac{1}{n} \log p(x_1, x_2, \ldots, x_n) = 2 - \frac{k}{n} \log 3,$$

where $k$ is the number of ones.

(c) (5 pts) Compute $\Pr(A_{\epsilon}^{(n)})$ and $|A_{\epsilon}^{(n)}|$ for $n = 8$ and $\epsilon = 0.2$. Use parts (a) and (b) along with Property 1 of Theorem 3.1.2 (more precisely, its converse, which is also true).

(d) (5 pts) Repeat for $n = 16$ and $\epsilon = 0.2$.

(e) (1 pt) Did $\Pr(A_{\epsilon}^{(n)})$ increase with $n$?

(f) (4 pts) Confirm that the inequality between $|A_{\epsilon}^{(n)}|$ and $2^{n[H+\epsilon]}$ is satisfied.

2. (12 pts) *The AEP and source coding.* A discrete memoryless source emits a sequence of statistically independent binary digits with probabilities $p(1) = 0.005$ and $p(0) = 0.995$. The digits are taken 100 at a time and a binary codeword is provided for every sequence of 100 digits containing three or fewer ones.

(a) Assuming that all codewords are the same length, find the minimum length required to provide codewords for all sequences with three or fewer ones.

(b) How does this compare to the 100 bit “brute-force” representation?

(c) Calculate the probability of observing a source sequence for which no codeword has been assigned.
3. (8 pts) Relative Entropy AEP. Let $X_1, X_2, \ldots$ be independent identically distributed random variables drawn according to the probability mass function $p(x)$, $x \in \{1, 2, \ldots, m\}$. Thus $p(x_1, x_2, \ldots, x_n) = \prod_{i=1}^{n} p(x_i)$. We know that $-\frac{1}{n} \log p(X_1, X_2, \ldots, X_n) \to H(X)$ in probability. Let $q(x_1, x_2, \ldots, x_n) = \prod_{i=1}^{n} q(x_i)$, where $q$ is another probability mass function on $\{1, 2, \ldots, m\}$.

(a) Evaluate $\lim -\frac{1}{n} \log q(X_1, X_2, \ldots, X_n)$, where $X_1, X_2, \ldots$ are i.i.d. $\sim p(x)$. Express your answer as a simple combination involving a subset of the terms $D(p||q)$, $H(p)$, and $H(q)$.

(b) Now evaluate the limit of the log likelihood ratio $-\frac{1}{n} \log \frac{q(x_1, \ldots, x_n)}{p(x_1, \ldots, x_n)}$ when $X_1, X_2, \ldots$ are i.i.d. $\sim p(x)$. Again express your answer in terms typical information theory quantities.

4. (8 pts) Take it to the limit.

The sequence pair $(x^n, y^n)$ is drawn i.i.d. according to the p.m.f.

$$p(x^n, y^n) = \prod_{i=1}^{n} p(x_i, y_i).$$

(2)
i.e. The pairs are independent of each other but the $x_i$ and $y_i$ within a pair $(x_i, y_i)$ are dependent according to the joint distribution $p(x, y)$.

What is the limit as $n \to \infty$ of

$$\frac{1}{n} \log \frac{p(x^n, y^n)}{p(x^n)p(y^n)}?$$

(3)

To get full credit you must express your answer in the simplest form and show your argument.

5. (8 pts) Time’s arrow. Let $\{X_t\}_{t=-\infty}^{\infty}$ be a stationary stochastic process. Prove that

$$H(X_0|X_{-1}, X_{-2}, \ldots, X_{-n}) = H(X_0|X_1, X_2, \ldots, X_n).$$

In other words, the present has a conditional entropy given the past equal to the conditional entropy given the future.

6. (8 pts) Average entropy per element vs. conditional entropy. For a stationary stochastic process $X_1, X_2, \ldots, X_n$, show that

$$\frac{H(X_1, X_2, \ldots, X_n)}{n} \geq H(X_n|X_{n-1}, \ldots, X_1).$$

(4)
7. (8 pts) *Random walk on chessboard.* Find the entropy rate of the Markov chain associated with a random walk of a king on the $3 \times 3$ chessboard shown below.

\[
\begin{array}{ccc}
1 & 2 & 3 \\
4 & 5 & 6 \\
7 & 8 & 9 \\
\end{array}
\]

Assume that the king must move at each step in the Markov process and that it is equally likely to choose any of the legal chess moves for a king that are available to it. i.e. It can’t stay in the same square and it can move to any neighboring square.

The distribution for the king’s initial position is the stationary distribution.

8. (8 pts) *Random Walk of a Spider.*

![Spider web diagram](image)

Figure 1: Spider web.

Compute the entropy rate for the random walk of a spider on the web shown in Figure 1. At each step in the random walk the spider must move to an adjacent node. The spider is equally likely to choose each of the adjacent nodes. Assume that the initial node of the spider is random with the stationary probability mass function.