ELEC 535 Homework 7

Due date: In class on Friday, March 28, 2003

Instructor: Rudolf Riedi Rice University, Spring 2003

Problem 7.1 (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 $\{1, 2, \ldots, m\}$.

1. Evaluate $\lim -\frac{1}{n} \log q(X_1, X_2, \ldots, X_n)$, where $X_1, X_2, \ldots$ are i.i.d. $\sim p(x)$. (Hint: use the LLN and pay attention to the fact that the distribution of the X_i is p.)

2. Now evaluate the limit of the log likelihood ratio $\frac{1}{n} \log \frac{q(X_1, X_2, \ldots, X_n)}{p(X_1, X_2, \ldots, X_n)}$ when $X_1, X_2, \ldots$ are i.i.d. $\sim p(x)$. Thus the odds favoring q are exponentially small when p is true.

Problem 7.2 (Random box size)

An n-dimensional rectangular box with sides $X_1, X_2, X_3, \ldots, X_n$ is to be constructed. The volume is $V_n = \prod_{i=1}^n X_i$. We define the edge length l of an n-cube with volume V_n as $l = V_n^{1/n}$.

Let $X_1, X_2, \ldots$ be i.i.d. uniform random variables over the unit interval $[0, 1]$. Find $\lim_{n\to\infty} V_n^{1/n}$, and compare to $(EV_n)^{1/n}$. Clearly the expected edge length does not capture the idea of the volume of the box.

Problem 7.3 (Preprocessing the output.)

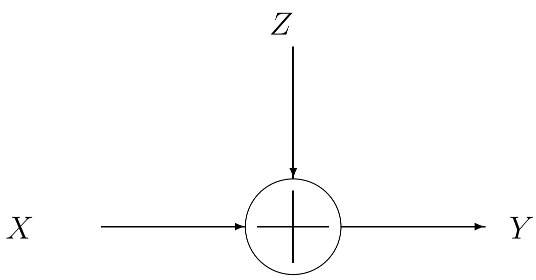
One is given a communication channel with transition probabilities $p(y|x)$ and channel capacity $C = \max_{p(x)} I(X;Y)$. A helpful statistician preprocesses the output by forming $\tilde{Y} = g(Y)$. He claims that this will strictly improve the capacity.

1. Show that he is wrong.

2. Under what conditions does he not strictly decrease the capacity?

Problem 7.4 (An additive noise channel.)

Given is the following discrete memoryless channel:



where $\Pr\{Z = 0\} = \Pr\{Z = a\} = 1/2$. This means that we model the output as $Y = X + Z$. The alphabet for x is $\mathcal{X} = \{0, 1\}$. Compute the alphabet for Y. Find the channel capacity assuming that Z is independent of X .

Observe that the channel capacity depends on the value of a.

Problem 7.5 (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 assigned for every sequence of 100 digits containing three or fewer ones. (Thus, most transmission errors can easily be detected since they will typically lead to sequences with too many ones.)

1. Assuming that all codewords are the same lengths, find the minimum length required to provide codewords for all sequences with three or fewer ones. (Hint: compute first the number of 0-1 sequences of length 100 with three or fewer ones.)

2. Calculate the probability of observing a source sequence for which no codeword has been assigned. (Detectable transmission error.)

3. Consider the random variable S_n that is the sum of the n i.i.d. random variables $X_1, X_2, \ldots, X_n$. Then $n\mu$ and $n\sigma^2$ are the mean and variance of S_n, where μ and σ^2 are the mean and variance of X_i. Chebychev's inequality states that

$$\Pr(|S_n - n\mu| \geq \epsilon) \leq \frac{n\sigma^2}{\epsilon^2},$$

   Use Chebyshev's inequality to bound the probability of observing a source sequence for which no codeword has been assigned. Compare this bound with the actual probability computed in part (2).