ECET 602: Homework 1

Spring, 2015

Due April 7 in class

1. Using a book of your choosing from the project Gutenberg website and MATLAB, calculate an estimated entropy of English text under the assumption that it is an independently and identically distributed sequence of characters (letters, spaces, numbers, and punctuation).

   (a) How does the entropy of this source compare with the entropy of a source with the same alphabet (letters, spaces, numbers, and punctuation) with maximum entropy?

   (b) Predict the number of bits necessary to represent this book by multiplying your estimated entropy by the number of symbols in the book.

   (c) Compare the size of the ASCII text file for the book with your predicted compressed length. Compress the ascii text file using your favorite compression program (e.g. pkzip, gzip, winzip, or power archiver). How does its size compare with your predicted size? If the compressed file is smaller than your expected size, can you give a reason why that is possible?

2. Find the capacity of the Binary erasure channel, which takes an input $X \in \{0, 1\}$ and returns an output Y that is equal to X with probability $1 - p$ or an "erasure" output $Y = e$ with probability p. Derive/prove your answer.

3. Given a particular probability mass function $p_X : \mathcal{X} \to [0, 1]$ for a discrete random variable taking values on a finite set $\mathcal{X}$, define the set of strongly typical sequences from this distribution as

$$\mathcal{T}_\epsilon^N(X) = \left\{ \mathbf{x} = (x_1, \ldots, x_N) \in \mathcal{X}^N \left| \left| \frac{1}{N}\mathcal{N}(a|\mathbf{x}) - p_X(a) \right| < \frac{\epsilon}{|\mathcal{X}|} \quad \forall a \in \mathcal{X} \right. \right\} \tag{1}$$

   where $\mathcal{N}(a|\mathbf{x})$ counts the number of occurrences of a in the vector $\mathbf{x}$.

   (a) Let $X_1, X_2, \ldots$ be a sequence of random variables that are independently and identically distributed according to p_X. For a fixed value of ϵ, does $\lim_{N\to\infty} \mathbb{P}[(X_1, X_2, \ldots, X_N) \in \mathcal{T}_\epsilon^N(X)]$ exist, and if so, what is its value. Prove your answer.

   (b) Provide and prove an upper and lower bound for the number of elements in the strongly typical set $\mathcal{T}_\epsilon^N(X)$.

Turn in all of your MATLAB code and a printout of your results. Include a brief written description of how you solved each of the problems.