Information Theory: Exercises

Mathias Winther Madsen
mathias.winther@gmail.com

January 12, 2015

Probability threshold sets (Cover and Thomas, Exercise 3.5) Let $X_1, X_2, X_3, \ldots$ be a series of independent random variables drawn from a distribution X with entropy $H(X)$. Let further $C_n(\tau)$ be the set of all high-probability sequences $(x_1, x_2, \ldots, x_n)$ for which

$$\Pr(X_1 = x_1, X_2 = x_2, X_3 = x_3, \ldots, X_n = x_n) \geq 2^{-n\tau}.$$

1. What's the highest number of elements such a set can contain?

2. Sketch a graph of $\Pr C_n(\tau)$ as a function of τ for a large value of n, and for an extremely large value of n.

Source coding (Cover and Thomas, Exercise 3.7) An information source produces produces pixels $X_1, X_2, X_3, \ldots$ with $\Pr(X_i = \texttt{WHITE}) = 0.995$ and $\Pr(X_i = \texttt{BLACK}) = 0.005$.

You decide to brute-force encode outputs from this source, 100 pixels at a time, by means of a table of equally long codewords. You include all sequences with three or fewer black pixels in the table and accept there will be an error in the remaining cases.

1. Compute or estimate the number of codewords you will need for this encoding scheme.

2. What are your options for reducing these space requirements?

3. Bound the probability that this encoding scheme will encounter an untabulated sequence.