STAT/ELEC 331 HW 6

Problems in addition to those from the book

1. Error-Correcting Codes

Consider a binary symmetric channel (BSC) and suppose we wish to transmit 4 information bits. In this problem we will consider the so-called Hamming (7,4) code, which consists of the codewords

0000000	0100101	1000011	1100110
0001111	0101010	1001100	1101001
0010110	0110011	1010101	1110000
0011001	0111100	1011010	1111111

Each of the 16 possible configurations of information bits is mapped to one of these codewords. The first four bits of a codeword correspond to the information bits. The codeword is then transmitted over the BSC, and decoding is based on the notion of Hamming distance.

The *Hamming distance* between two binary sequences is defined to be the number of bits where the two sequences differ. For example, if d denotes Hamming distance, then $d(011010001, 010011011) = 3$.

The codewords comprising the Hamming (7,4) code satisfy the following property: For any 7 bit sequence $\mathbf{z}$, there is a unique codeword $\mathbf{x}^k$ such that $d(\mathbf{z}, \mathbf{x}^k) \leq 1$. That is, there is a unique codeword whose Hamming distance from $\mathbf{z}$ is at most 1. You do not need to prove this, but it would be good to convince yourself.

This suggests the following decoding strategy: if a sequence $\mathbf{z}$ is received, let $\widehat{\mathbf{x}}$ denote the closest codeword (in terms of hamming distance). The decoder guesses that the transmitted information bits are the first four bits of $\widehat{\mathbf{x}}$.

a. If $p = 0.1$ (the parameter characterizing the BSC), what is the probability of error for the Hamming (7,4) code? I will get you started:

$$\begin{aligned}
\mathbf{P}(\widehat{\mathbf{x}} \neq \mathbf{x}) &= 1 - \mathbf{P}(\widehat{\mathbf{x}} = \mathbf{x}) \\
&= 1 - \sum_{k=1}^{16} \mathbf{P}(\{\widehat{\mathbf{x}} = \mathbf{x}^k\} \cap \{\mathbf{x} = \mathbf{x}^k\}) \\
&= 1 - \sum_{k=1}^{16} \mathbf{P}(\widehat{\mathbf{x}} = \mathbf{x}^k \mid \mathbf{x} = \mathbf{x}^k)\mathbf{P}(\mathbf{x} = \mathbf{x}^k).
\end{aligned}$$

This is simply an application of the law of total probability. Here $\mathbf{x}$ denotes the transmitted sequence and $\widehat{\mathbf{x}}$ the decoded sequence. Assume that all 16 configurations of information bits are equally likely.

b. If the Hamming (7,4) code is useful, it should have a lower probability of error than if we did no channel coding. Verify this for $p = 0.1$.

c. Consider the repitition code where each information bit is transmitted 3 times. What is the probability of error assuming $p = .1$ and $n = 4$ information bits are encoded?

d. Even though the answer in c is smaller than the answer in a, why might the Hamming (7,4) code be preferable to the triple-repetition code in practice?

2. Truncated Normal

Suppose $X \sim N(0, \sigma^2)$ and let f_X denote the pdf of X. Let $\gamma > 0$ and consider the continuous random variable Y whose pdf is

$$f_Y(y) = \begin{cases} 0 & \text{if } y < -\gamma \\ C f_X(y) & \text{if } -\gamma \le y \le \gamma \\ 0 & \text{if } y > \gamma \end{cases},$$

where C is such that f_Y integrates to 1. Y is called a *truncated normal* random variable. Such RVs are useful for modeling data with normal characteristics but a bounded range.

a. Express C in terms of Φ, the cdf of a standard normal.

b. Sketch the pdfs of X and Y on the same graph. On a separate graph, sketch the cdfs of X and Y. Label your graphs clearly. For this problem take $\sigma^2 = 1$ and $\gamma = 1$.

Optional: 2 points extra credit – Do not use a calculator or computer.

c. Determine the variance of Y as a function of γ and σ^2. Express the result as σ^2 times a factor between zero and one.

d. Fix $\sigma^2 > 0$. Using L'Hopital's rule if necessary, determine the limits of $var(Y)$ as $\gamma \to \infty$ and $\gamma \to 0$. Do your results make sense?