Information Theory
Practice Set 11

Lakshmi Prasad Natarajan

Solutions are not to be returned

Notation: Bold letters $\boldsymbol{x}$, $\boldsymbol{y}$ denote sequences or vectors, whereas x, y are scalars. Random vectors are denoted as $\boldsymbol{X}, \boldsymbol{Y}$. Components of $\boldsymbol{x}$ are $x_1, \ldots, x_n$.

Review

- Recall the proof that the MAP rule minimizes the probability of error in hypothesis testing.
- In the MAP rule for decoding, suppose there are two codewords both of which yield the maximum value of a posteriori probability $p(\boldsymbol{x}(w)|\boldsymbol{y})$, what should the decoder do?

Practice Set

1. *AWGN Channel.* Consider the AWGN channel with $\mathcal{X} = \mathcal{Y} = \mathbb{R}$. The input X and output Y are related as $Y = X + Z$, where Z is independent of X and follows a Gaussian distribution with mean 0 and variance σ^2.

   (a) For a single use of this channel, what is the conditional distribution of output y given input x, i.e., what is the conditional probability density function $f(y|x)$?

   (b) If the channel is used n times, what is $f(y_1, \ldots, y_n | x_1, \ldots, x_n)$?

   (c) Suppose $\mathcal{C} = \{\boldsymbol{x}(1), \ldots, \boldsymbol{x}(M)\}$ is an (M, n) code for this channel. Which of the following is the optimal decoder $g^*(\boldsymbol{y})$ for this channel:

   $$\hat{w} = \arg\min_{w} \sum_{i=1}^{n} (y_i - x_i(w))^2, \text{ or}$$
   $$\hat{w} = \arg\min_{w} \sum_{i=1}^{n} |y_i - x_i(w)| .$$

   (d) Define the optimal decision regions $\mathcal{D}(w) = \{\boldsymbol{y} \mid g^*(\boldsymbol{y}) = w\}$ for $w = 1, \ldots, M$, that is, $\mathcal{D}(w)$ is the set of all possible channel outputs that are decoded to $\hat{w} = w$ by the optimal decoder. Draw the decision regions $\mathcal{D}(1), \ldots, \mathcal{D}(M)$ for the following codes in the AWGN channel:

      i. *Binary Phase Shift Keying (BPSK) Modulation.* Here, $n = 1$, $M = 2$, $\mathcal{C} = \{+a, -a\}$, where $a > 0$.

      ii. *Binary Frequency Shift Keying (BFSK) Modulation.* Here, $n = 2$, $M = 2$, $\mathcal{C} = \{(+a, 0), (0, +a)\}$, where $a > 0$.

      iii. *Bi-Orthogonal Modulation.* $n = 2$, $M = 4$, $\mathcal{C} = \{(+a, 0), (0, +a), (-a, 0), (0, -a)\}$.

      iv. *M-ary Phase Shift Keying (PSK) Modulation.* $n = 2$, $M \geq 3$, $\mathcal{C} = \left\{\left(\cos\left(\frac{2\pi w}{M}\right), \sin\left(\frac{2\pi w}{M}\right)\right) \;:\; w = 1, \ldots, M\right\}$.

2. What is the optimal decoding rule for the binary erasure channel? Explain why.

3. *Hamming Distance.* Let $\boldsymbol{x}, \boldsymbol{y}$ be two vectors in $\{0, 1\}^n$. The Hamming distance between them is defined as $d(\boldsymbol{x}, \boldsymbol{y}) = |\{i : x_i \neq y_i\}|$.

   (a) For a given $\boldsymbol{x}$, how many vectors $\boldsymbol{y}$ are there such that $d(\boldsymbol{x}, \boldsymbol{y}) \leq t$, where t is a positive integer.

   (b) Assume that a vector $\boldsymbol{x}$ is sent via BSC(δ). Let $\boldsymbol{Y}$ denote the random vector observed as the channel output when the input is $\boldsymbol{x}$. What is the probability that $\boldsymbol{Y}$ and $\boldsymbol{x}$ differ in at the most t positions, i.e., find $P\left[d(\boldsymbol{x}, \boldsymbol{Y}) \leq t\right]$.

`Remark:` : The Hamming distance satisfies triangle inequality, that is for any three vectors $\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z}$, we have

$$d(\boldsymbol{x}, \boldsymbol{y}) \leq d(\boldsymbol{x}, \boldsymbol{z}) + d(\boldsymbol{z}, \boldsymbol{y}).$$

The Hamming distance is also symmetric, $d(\boldsymbol{x}, \boldsymbol{y}) = d(\boldsymbol{y}, \boldsymbol{x})$. Further, $d(\boldsymbol{x}, \boldsymbol{y}) = 0$ if and only if $\boldsymbol{x} = \boldsymbol{y}$.

4. *Minimum distance of a code.* Assume $\mathcal{X} = \{0, 1\}$. An (M, n) code is a collection of M binary vectors of length n, $\mathcal{C} = \{\boldsymbol{x}(1), \ldots, \boldsymbol{x}(M)\}$.

   The minimum distance $d_{\min}$ of the code $\mathcal{C}$ is the smallest Hamming distance between any two codewords. That is

   $$d_{\min} = \min_{w \neq w'} d(\boldsymbol{x}(w), \boldsymbol{x}(w')).$$

   Find the minimum distance of the following codes:

   (a) $\mathcal{C} = \{(0,0,0,0), (1,1,1,1), (1,1,0,0), (0,0,1,1)\}$.

   (b) *Repetition Code.* $n \geq 2$, $M = 2$, $\mathcal{C} = \{(0, 0, \ldots, 0), (1, 1, \ldots, 1)\}$.

   (c) *Single Parity Check Code.* Let $n \geq 2$. $\mathcal{C}$ is the collection of all binary sequences of length n such that the number of 1's in it is even.
   `Hint`: Solve this first for $n = 3, 4$. Explicitly list all the codewords of single parity-check code for $n = 3, 4$. Later extend your argument to any value of n.

5. Consider the BSC(δ) channel with $\delta < 0.5$. Suppose we use a code $\mathcal{C}$ of length n and minimum distance $d_{\min}$ in this channel. We will derive an upper bound on P_e when the optimal decoder is used. For this, assume that t is an integer such that $d_{\min} \geq 2t + 1$. Usually, we choose t to be the largest integer such that $d_{\min} \geq 2t + 1$. For instance, if $d_{\min} = 5$ we have $t = 2$, or if $d_{\min} = 4$ we have $t = 1$.

   Note that $\mathcal{X} = \mathcal{Y} = \{0, 1\}$, and $\mathcal{C} = \{\boldsymbol{x}(1), \ldots, \boldsymbol{x}(M)\}$. If $\boldsymbol{x}(w)$ is transmitted and $\boldsymbol{y}$ is received, then the number of bit-flips introduced by the channel is exactly $d(\boldsymbol{x}(w), \boldsymbol{y})$. Also, recall that the optimal decoder for BSC is $g^*(\boldsymbol{y}) = \arg\min_w d(\boldsymbol{x}(w), \boldsymbol{y})$.

   (a) Argue that for any $w \neq w'$, we have $d(\boldsymbol{x}(w), \boldsymbol{x}(w')) \geq d_{\min}$.

   (b) Assume that $\boldsymbol{x}(w)$ is transmitted and the channel introduces t or fewer bit-flips, i.e., the channel output $\boldsymbol{y}$ satisfies $d(\boldsymbol{x}(w), \boldsymbol{y}) \leq t$. Use triangle inequality of Hamming distance to show that

   $$d(\boldsymbol{x}(w'), \boldsymbol{y}) \geq t + 1 > d(\boldsymbol{x}(w), \boldsymbol{y}).$$

   `Remark`: If the channel introduces t or fewer bit-flips, then the received vector $\boldsymbol{y}$ is closer to the transmitted codeword $\boldsymbol{x}(w)$ than any other codeword $\boldsymbol{x}(w')$, $w' \neq w$.

   (c) Argue that if the number of bit-flips in the channel is less than or equal to t, then the optimal decoder outputs the correct value, i.e., $d(\boldsymbol{X}(W), \boldsymbol{Y}) \leq t \Rightarrow \hat{W} = W$. Note that this is a sufficient condition for correct decoding.

   `Remark`: This implies that if the optimal decoder makes a mistake, then necessarily the number of bit-flips is $t + 1$ or more.

   (d) Prove that for the optimal decoder we have $P_e \leq \sum_{i=t+1}^{n} \binom{n}{i} \delta^i (1-\delta)^{n-i}$.

`Remark`: For sufficiently small values of δ, this upper bound on P_e is approximately equal to $\binom{n}{t+1}\delta^{t+1}$, which can be much smaller than δ. Thus, by using a code with large t (i.e., large $d_{\min}$), we can guarantee a small upper bound on P_e.

Such codes are called t-error correcting codes – since the decoder makes no mistake even if the channel introduces up to t 'errors' or bit-flips (hence, the code is able to 'correct' t errors). Achieving large values of $d_{\min}$ comes at the cost of reductions in code rate.

One of the main problems in *coding theory* is to design codes with large rate and simultaneously small P_e (either by having a large $d_{\min}$ or otherwise). This involves abstract algebra, combinatorics and graph theory.

Following are some examples of well-known families of codes:

- *Hamming Codes.* Length $n = 2^m - 1$ for some $m \geq 3$, $d_{\min} = 3$, Rate $R = \frac{2^m - m - 1}{2^m - 1}$.
- *Primitive BCH Codes.* Length $n = 2^m - 1$ for some $m \geq 3$, $d_{\min} \geq 2t + 1$, $R \geq \frac{2^m - 1 - mt}{2^m - 1}$.
- *Reed-Muller Codes.* $n = 2^m$ for $m \geq 1$, $d_{\min} = 2^{m-r}$, where $0 \leq r \leq m$, and $R = \frac{1}{2^m} \sum_{i=0}^{r} \binom{m}{i}$.