Information Theory
Practice Set 12

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 $x_1, \ldots, x_n$.

Reading Exercise

- *Capacity of Weakly-Symmetric Channels:* Please read Section 7.2 of Cover & Thomas (this is a part of the exam syllabus).

  Both the BEC and the BSC exhibit symmetry in terms of their transition probabilities. This is the reason why the optimal input distribution is uniform. Section 7.2 generalizes this result a broader class of channels.

- *Source-Channel Separation Theorem:* Section 7.13 of Cover & Thomas (not part of the exam syllabus).

  This theorem is important from a practical perspective. Suppose you must transmit a source V with some entropy $H(V)$ over a channel $p(y|x)$, under what conditions can you achieve reliable communication? This theorem tells us that reliable communication is possible if and only if $H(V) < C$, where C is the capacity of the channel. Further, this theorem tells us that whenever $H(V) < C$, we can achieve reliable communication through a modular approach: first perform source coding on V to obtain a binary string, then encode this binary string using a capacity-achieving channel code. This two-step approach separates the channel coding problem from the source coding problem, and this is how most digital communications is conducted in practice today.

Practice Set

1. *The maximum probability of error.*

   We used $P_e = P[W \neq \hat{W}]$ as the performance metric in the lectures. A slightly more stringent metric is $\lambda = \max_w P[\hat{W} \neq w \mid W = w]$. We will now relate λ to P_e.

   Consider a code $\mathcal{C}$, with encoding and decoding functions f, g, respectively. For each $w \in \{1, \ldots, M\}$, define $\lambda_w = P[\hat{W} \neq w \mid W = w]$. Then $\lambda = \max_w \lambda_w$.

   (a) Argue that P_e is the arithmetic mean of λ_w, $w \in \{1, \ldots, M\}$. Conclude that $P_e \leq \lambda$.

   We next want to show that if $\mathcal{C}$ is an (M, n) code with $P_e \leq \epsilon$, then there exists an $(M/2, n)$ code $\mathcal{C}'$ with $\lambda \leq 2\epsilon$. The rates of these two codes are $\log M/n$ and $\log M/n \ - \ 1/n$. Thus the second code has a rate that is only negligibly smaller than the first code (when n is large), but it satisfies a more stringent reliability condition: the error probability is small ($\leq 2\epsilon$), no matter which message is transmitted.

   We will obtain $\mathcal{C}'$ from $\mathcal{C}$ via *expurgation*: start from $\mathcal{C}$, and throw away the worst $M/2$ of its codewords.

   (b) Consider the (M, n) code $\mathcal{C}$, and suppose $\lambda_1 \leq \lambda_2 \leq \cdots \leq \lambda_M$. Prove that if $P_e \leq \epsilon$, then $\lambda_{M/2} \leq 2\epsilon$.
   `Hint`: You can use Markov inequality to prove this. Or, you can prove via contradiction. Assume that $\lambda_{M/2} > 2\epsilon$, and show that this assumption implies $P_e > \epsilon$, which is a contradiction.

   (c) Let $\mathcal{C}'$ be the code consisting of the first $M/2$ codewords in $\mathcal{C}$. This is an $(M/2, n)$ code. To decode $\mathcal{C}'$, suppose we use g, which is the same decoder as that of $\mathcal{C}$. Thus, the output $\hat{w}$ of the decoder g can be any integer from $\{1, \ldots, M\}$, while the input at the encoder $w \in \{1, \ldots, M/2\}$. Argue that the maximal error probability λ for $\mathcal{C}'$ is at the most 2ϵ.

Remark: When we defined achievability of rate R, we did so using P_e: we required $P_e \to 0$. We can equivalently define achievability in terms of λ, by requiring $\lambda \to 0$. The above problems show that both the definitions are equivalent. Thus, capacity defined in terms P_e (as done in our lectures) and capacity defined in terms of λ (as done in the textbook) are equal.

2. We discussed in Lecture 38 that the capacity $C_{\text{AWGN}}(P)$ of the AWGN channel with input power constraint P is the maximum value of $I(X;Y)$ over all input distributions $f(x)$ with $\mathbb{E}(X^2) \leq P$.

   (a) Show that the capacity is equal to $\frac{1}{2}\log\left(1+\frac{P}{\sigma^2}\right)$.

   (b) Plot the capacity as a function of the signal to noise ratio $\text{SNR} = \frac{P}{\sigma^2}$, and observe that this function is concave.

   (c) Is it true that reliable communication is possible only when the signal power P is larger than the noise power σ^2?

3. Exercise Problems from Chapter 7 of Cover & Thomas:

   7.1, 7.2, 7.4 (a and b), 7.5 (this is applied in practice when multiple frequency bands are used for wireless communication, such as in a technology called OFDM), 7.7, 7.12, 7.13, 7.22, 7.23(b), 7.25, 7.27.

4. Based on 7.16(d). Any (M,n) code (with the encoder and decoder) together with a DMC can be thought of as yielding a new channel that maps input W to output $\hat{W}$. What we have done is augment the DMC with a preprocessing step (the encoder f) and a post processing step (the decoder g), and thereby create a 'channel' between W and $\hat{W}$. This channel is characterized by the transition probabilities

$$p(\hat{w}|w) = P[\hat{W} = \hat{w}|W = w] = P[g(\boldsymbol{Y}) = \hat{w} \mid \boldsymbol{X} = \boldsymbol{x}(w)].$$

   Argue that the capacity of this new channel $\max_{p(w)} I(\hat{W};W)$ is at the most nC where C is the capacity of the DMC.

5. *Denoising using Regenerative Repeaters.*

   Suppose we need to communicate messages from point A to point C. A and C are not connected directly, but only via another point B as follows: the communication link between A and B is $\text{BSC}(p)$ and the link from B to C is also a $\text{BSC}(p)$ channel, where $0 < p < 0.5$. The noise in these two communication links are independent of each other. In practice, B could be a router or a wireless access point. In communication engineering we say that B is a *relay.*

   (a) Suppose we plan to achieve reliable communication as follows: connect the output of the channel $A \to B$ to the input of channel $B \to C$ to form an overall channel $A \to C$, and then build a capacity-achieving code for this overall cascaded channel.

   What are the input and output alphabets of this overall channel? What is its capacity? Can you show that the capacity of this $A \to C$ channel is less than the capacity of $\text{BSC}(p)$?

   Remark: In part (a), we decided to NOT process the signals at point B. As a result we ended up with a system whose capacity is strictly less than the capacities of the two individual communication links in the system. Can we improve the communication rate if we are allowed to process the signals at point B?

   (b) Suppose we do the following: decode the transmitted message at point B, re-encode this message and transmit the codeword in the channel $A \to C$. For simplicity, ignore the delay incurred due to this additional processing. Can you argue that we can reliably communicate through this system at any rate $R < \text{BSC}(p)$?

Remark: This is a major advantage of digital communications over analog communications. As long as we communicate at rates less than the capacities of the individual links, we can perform denoising and regenerate the input message at any point in the network. The noise encountered so far in the network is completely removed. This relaying strategy is also called *decode and forward.* This is standard practice in Internet protocols.