Exercise 7: Entropy, Mutual Information, and Capacity

© 2024
Christoph Studer (studer@ethz.ch)
Stefan M. Moser (moser@isi.ee.ethz.ch)

- **Exercise: Thursday, 21 November 2024**
- This exercise will be assisted in CAB G11 from 8:15 h to 10:00 h
- Problems 1, 3, and 4 require a computer with MATLAB (or Octave) installed
- **Due date:** Monday, 2 December 2024, 08:00 h
- **Submission instructions:** Upload your solutions (answers and MATLAB code) to Moodle as a *single* zip-file with a file name `⟨last-name⟩-⟨first-name⟩-e7.zip`. Your homework should consist of a single pdf file for the written answers and one MATLAB script per problem part. The MATLAB file must be named according to the syntax `E7_P1_3.m` (for "Exercise 7 Problem 1 Part 3").
- **Rules:** You may collaborate with other students on this exercise, but each student must turn in their own assignment. You must specify the names of your collaborators! Copying answers from previous years' exercises is strictly forbidden.

Problem	**Maximum Points**	**Points Received**
1	20	
2	10	
3	35	
4	35	
Total	**100**	

Problem 1: Capacity of the Binary Symmetric Channel (20pts)

In class, we have discussed the channel capacity of discrete memoryless channels (DMCs). The goal of this problem is to compute the capacity of the binary symmetric channel (BSC). Remember that the channel capacity for DMCs is given by

$$C = \max_{p_{\mathsf{x}}(\cdot)} \mathcal{I}(\mathsf{x};\mathsf{y}), \tag{1}$$

where $\mathsf{x} \in \mathcal{X}$ are the channel input symbols, $\mathsf{y} \in \mathcal{Y}$ the channel output symbols, and $p_{\mathsf{x}}(x)$ is the prior probability that we are trying to optimize over (i.e., we have to find the optimal probability mass function at the channel's input). For the BSC, we have $\mathcal{X} = \mathcal{Y} = \{0, 1\}$ and the following channel law:

$$p_{\mathsf{y}|\mathsf{x}}(1,1) = p_{\mathsf{y}|\mathsf{x}}(0,0) = 1 - \varepsilon, \tag{2}$$

$$p_{\mathsf{y}|\mathsf{x}}(1,0) = p_{\mathsf{y}|\mathsf{x}}(0,1) = \varepsilon, \tag{3}$$

with the cross-over probability $\varepsilon \in [0, 1]$.

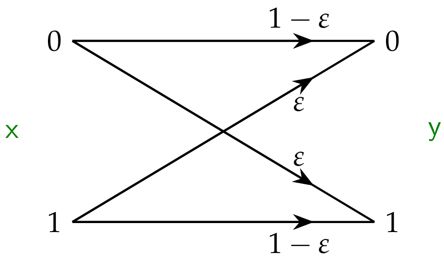

Figure 1: Graphical illustration of the binary symmetric channel (BSC).

Part 1 (8pts). Expand the mutual information expression $\mathcal{I}(\mathsf{x};\mathsf{y})$ in (1), i.e., write down its explicit form in terms of $\varepsilon \in [0, 1]$ and the input probability $p_{\mathsf{x}}(0) = 1 - p_{\mathsf{x}}(1) = p$.
Hint: You need the definitions of mutual information, entropy, and conditional entropy.

Part 2 (4pts). Compute the optimal input distribution $p_{\mathsf{x}}(0) = 1 - p_{\mathsf{x}}(1) = p$ that maximizes the mutual information derived in Part 1. What is the channel capacity for the BSC? Provide a closed-form expression.

Part 3 (8pts). Plot the channel capacity in MATLAB as a function of the cross-over probability $\varepsilon \in [0, 1]$. Which crossover probability minimizes the channel capacity? Explain why the channel capacity of the BSC is the same for $\varepsilon = 0$ and $\varepsilon = 1$.

Problem 2: Entropy of a Random Vector (10pts)

Suppose that the random vector $\mathbf{x} = (x_1, x_2, x_3)^\mathsf{T}$ takes on the values

$$\begin{pmatrix}0\\0\\0\end{pmatrix}, \quad \begin{pmatrix}0\\1\\1\end{pmatrix}, \quad \begin{pmatrix}1\\0\\1\end{pmatrix}, \quad \text{and} \quad \begin{pmatrix}1\\1\\0\end{pmatrix},$$

each with probability $\frac{1}{4}$.

Part 1 (2pts). Compute $\mathcal{H}(x_1, x_2, x_3)$.

Part 2 (6pts). Calculate the following uncertainties:

1. $\mathcal{H}(x_1)$, $\mathcal{H}(x_2)$, and $\mathcal{H}(x_3)$.
2. $\mathcal{H}(x_2|x_1 = 0)$, $\mathcal{H}(x_2|x_1 = 1)$, and $\mathcal{H}(x_2|x_1)$.
3. $\mathcal{H}\big(x_3\big|(x_1, x_2) = (0,0)\big)$, $\mathcal{H}\big(x_3\big|(x_1, x_2) = (0,1)\big)$, $\mathcal{H}\big(x_3\big|(x_1, x_2) = (1,0)\big)$, $\mathcal{H}\big(x_3\big|(x_1, x_2) = (1,1)\big)$, and $\mathcal{H}\big(x_3\big|x_1, x_2\big)$.

Part 3 (2pts). Verify from Part 1 and 2 that the chain rule is satisfied.

Problem 3: Capacity of a SISO Additive White Gaussian Noise Channel (35pts)

In this problem, you will numerically evaluate the channel capacity of a complex-valued SISO AWGN channel, and the maximum possible transmission rates for different QAM and PSK modulation schemes. We consider the complex-valued AWGN channel:

$$y = x + n, \tag{4}$$

where $y \in \mathbb{C}$ is the received signal, $x \in \mathcal{X}$ is the transmit signal with $\mathbb{E}_x\left[|x|^2\right] = E_s$, and $n \sim \mathcal{CN}(0, N_0)$ is AWGN. In class, you learned that the capacity of this channel is given by

$$C_{\text{AWGN}} = \max_{f_x} \mathcal{I}(x; y) = \log_2(1 + SNR) \quad [\text{bit/channel use}], \tag{5}$$

where f_x turns out to be a Gaussian distribution with variance E_s, and $SNR = E_s/N_0$. This implies that the input distribution that achieves capacity in this channel is Gaussian and no rate R above C_{AWGN} allows for reliable communication.

What if we are forced to use BPSK, QPSK, or even higher-order modulation schemes? What will the maximum transmission rate be at a certain SNR value, given we have an optimal forward error correction scheme? This is what we are going to simulate next.

Part 1 (4pts). Use MATLAB to plot the channel capacity (5) in terms of bits/channel use dependent on the SNR (in decibel). Plot the SNR in the range from -10 dB to $+20$ dB, and use a linear scale for the y-axis (for the channel capacity).

Hint: Do not forget the correct units.

Part 2 (4pts). Imagine you are forced to use QPSK transmission, but you can come up with a perfect forward error correction scheme (e.g., something much more powerful than a Hamming code). What is the maximum possible rate you can achieve? Provide a brief explanation.

Hint: Do not try to come up with a coding scheme. Just think about what the maximum possible rate could be.

Part 3 (8pts). In the following parts of this problem, we will simulate the so-called *maximum achievable rate* given a modulation scheme, such as BPSK or others. The maximum achievable rate is simply given by

$$C(SNR, \mathcal{X}) = \mathcal{I}(\mathsf{x}; \mathsf{y}), \tag{6}$$

i.e., the mutual information between the input signal $\mathsf{x} \in \mathcal{X}$, where $\mathcal{X}$ is a fixed modulation scheme (e.g., QPSK), and the output signal $\mathsf{y} \in \mathbb{C}$, in the presence of AWGN. Similarly to the channel capacity, it is impossible to communicate reliably at a higher rate R than the achievable rate $C(SNR, \mathcal{X})$, given a particular modulation scheme $\mathcal{X}$.

In principle, one could analytically compute the expression in (6). For higher-order modulation schemes (such as 16-QAM), however, the resulting expressions would be extremely complicated. We will therefore perform Monte–Carlo simulations to compute the maximum achievable rate. It is important to realize that the mutual information $\mathcal{I}(\mathsf{x}; \mathsf{y}) = \mathfrak{h}(\mathsf{y}) - \mathfrak{h}(\mathsf{y}|\mathsf{x})$ can be rewritten as

$$\mathcal{I}(\mathsf{x}; \mathsf{y}) = \int_{y \in \mathbb{C}} \sum_{x \in \mathcal{X}} p(x, y) \log_2\left(\frac{p(x, y)}{p(x)\, p(y)}\right) \mathrm{d}y \tag{7}$$

or, equivalently,

$$\mathcal{I}(\mathsf{x}; \mathsf{y}) = \mathbb{E}_{\mathsf{y},\mathsf{x}}\left[\log_2\left(\frac{p(\mathsf{x}, \mathsf{y})}{p(\mathsf{x})\, p(\mathsf{y})}\right)\right]. \tag{8}$$

(Actually, note that we have been a bit sloppy here regarding the fact that x is discrete, while y is continuous.) Hence, to approximate $\mathcal{I}(\mathsf{x}; \mathsf{y})$, we can simply compute the *empirical* expected value of (8). More specifically, we can randomly generate input signals $\mathsf{x} \in \mathcal{X}$ and noise realizations $\mathsf{n} \sim \mathcal{CN}(0, N_0)$, and compute $\mathsf{y} = \mathsf{x} + \mathsf{n}$. We can then evaluate the $\log_2$-term in (8). By repeating this procedure thousands of times and by averaging the results, we can compute an empirical estimate of the expectation operation in (8).

To solve this part, do not perform any Monte–Carlo simulations just yet. Instead, rewrite the $\log_2$-term in (8) using the PMF p_{x}, the conditional PDF $f_{\mathsf{y}|\mathsf{x}}$ and the PDF f_{y}, and simplify the expression so that the prior term p_{x} vanishes. Furthermore, write out the distribution f_{y} analytically.

Hint: The distribution f_{y} will contain a sum over all constellation points in $\mathcal{X}$.

Part 4 (15pts). Simulate the maximum achievable rate in (6) using the procedure outlined in Part 3 and your simplified $\log_2$-expression. Plot the maximum achievable rate for BPSK, QPSK, 8-PSK, and 16-QAM in terms of bits/channel use dependent on the SNR (in decibel). Perform a sufficiently large number of Monte–Carlo trials to get smooth curves for the achievable rates. Also include the channel capacity curve from Part 1. Plot the SNR in the range from $-10\,\text{dB}$ to $+20\,\text{dB}$, and use a linear scale for the y-axis.

Part 5 (4pts). In Part 2, you were asked to guess the maximum achievable rate for QPSK transmission. Does the simulation result from Part 4 confirm your answer in Part 2?

Remark: As you will see, at sufficiently low SNR values, the use of BPSK, QPSK, 8-PSK, and 16-QAM are near-optimal, i.e., they achieve basically the same rates as the channel capacity C_{AWGN}, which would require a Gaussian codebook. Hence, the use of strong codes in combination with these modulation schemes will perform as good as AWGN codebooks, which we know are not practical.

Problem 4: Capacity of Hard-Output vs. Soft-Output Channel (35pts)

The goal of this problem is to analyze the performance of a hard-output channel and compare it to the performance of a soft-output channel. Therefore, we consider the complex-valued AWGN channel from Problem 3. In Exercise 6 Problem 2, you have learned how to compute LLR values from an AWGN channel output y. Throughout this problem, we consider a QPSK constellation with unit energy, as depicted in Figure 2.

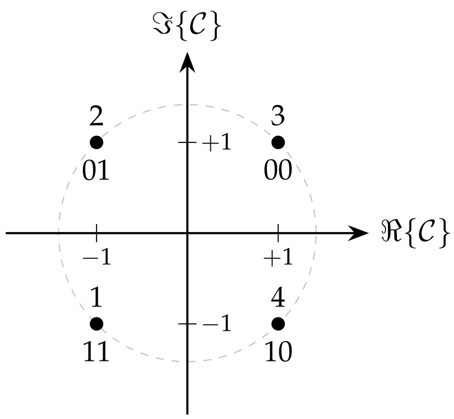

Figure 2: Unit-energy QPSK constellation with Gray labeling (shown below).

Part 1 (5pts). Compute the LLRs corresponding to bit 1 and 2 for the complex channel output y.

Part 2 (10pts). Simulate the mutual information $\mathcal{I}(x; \mathsf{LLR}_1, \mathsf{LLR}_2)$ and plot the results for varying SNR values.

Hint: Are the LLRs independent? Note that LLR_1 *and* LLR_2 *are continuous random variables. Thus, apply the chain rule to decompose* $\mathcal{I}(x; \mathsf{LLR}_1, \mathsf{LLR}_2)$ *and implement a histogram-based method to obtain a binned (discrete) estimate of the probability density functions required in* (8).

Part 3 (10pts). Apply the MAP rule to obtain bit estimates for $\breve{b}_1$ and $\breve{b}_2$. Now, simulate the mutual information for the resulting hard-output channel, i.e., $\mathcal{I}(x; \breve{b}_1, \breve{b}_2)$ and plot the result for varying SNR values.

Hint: Again try to simplify the expression for the mutual information and implement a histogram-based method to estimate the probability mass functions required in (8).

Part 4 (10pts). Do you observe any difference in the mutual information between the soft- and hard-output channels? Quantify the performance gap with respect to SNR [dB] at 1 bit per channel use. If you were to design a practical forward error correction decoder, would you prefer to take LLRs, or bit estimates as input?