Information theory - Exercises

Exercises from Cover and Thomas, 2nd edition

2.1, 2.29
4.2, 4.6
5.6, 5.8, 5.12, 5.15, 5.44
7.12, 7.16, 7.18, 7.20, 7.23, 7.27, 7.34(a,b), 7.35, 7.36
10.2, 10.5
15.1, 15.2, 15.3, 15.20, 15.23, 15.25, 15.31, 15.33

Other exercises

Exercise 1. Design a Huffman code for the outcome of a fair die. Evaluate the average length, and compare it against the source entropy. For which source distribution this would be an optimal code?

Exercise 2. Let \( p \) be the probability of head in a sequence of tosses of a biased coin. A sequence of messages is parsed using the alphabet \( X = \{ h, th, tt \} \). Evaluate \( H(X) \) in bits/message and also in bits per coin toss. Verify that the latter entropy is \( H(p) \), as expected.

Exercise 3. Consider the following codes for the source alphabet \( X = \{ a, b, c, d, e \} \):

- a) 110, 1110, 0, 100, 1111
- b) 111, 100, 0, 101, 110
- c) 10, 110, 01, 111, 00
- d) 10, 0, 110, 111, 101

Which codes are Huffman codes? Decode the binary sequence 11000101000111 using all the above codes. Explain the anomalies.

Exercise 4. Let the codewords of a Huffman code have length \( L_i = 3, 2, 4, 1, 4 \). Find the code. Devise a method to choose the codewords uniquely (i.e., such that the decoder knows the codewords once it knows their lengths).

Exercise 5. A pair of bits \( XY \) is chosen as follows. First, \( X = 0 \) with probability \( p \). Then, if \( X = 0 \), \( Y = 0 \) with probability \( q \); if \( X = 1 \), \( Y = 0 \) with probability \( r \). Use the properties of joint and conditional entropies to find the entropy of the pair. Evaluate the probabilities of the four source messages, and check that the entropy is correct.
Exercise 6. A ternary source with alphabet $X = \{0, 1, 2\}$ produces a message as follows. With probability $p$ the message is 0. Otherwise, 1 and 2 are chosen with probabilities $q$ and $1 - q$. Use the properties of joint and conditional entropies to find the source entropy. Evaluate the probabilities of the source messages, and check that the entropy is correct.

Exercise 7. For a memoryless source, write a computer program that evaluates the average length of a Huffman code for blocks of $N$ symbols. You can use, for instance, the Matlab function `huffmandict.m`. Verify the correctness of your program checking the examples in the lecture notes.

Exercise 8. A memoryless binary source produces zeros and ones with probabilities 0.9 and 0.1. Encode a block of six zeros with 0, and any other block of six bits $xyztuv$ with $1xyztuv$. Find the average length of this code, and compare with the source entropy.

Exercise 9. For the above source, use Huffman codes for blocks of size $N = 1, 2, 3, 4, 5, 6$. Find the average length. A computer is needed, for large $N$.

Exercise 10. A memoryless binary source produces zeros and ones with probabilities $1/3$ and $2/3$. Find source codes for blocks of $N = 5$ symbols, evaluate the average code length, and compare with the source entropy.

Exercise 11. We have a quaternary first-order Markov source. Given the message $X_{k-1}$, $X_k = X_{k-1}$ with probability $1/2$, and is equal to any other symbol with probability $1/6$. Note that the four symbols are equiprobable (why?). Evaluate the source entropy.

Exercise 12. Write a computer program that evaluates the entropy of a (first-order) Markov source, given the state transition matrix $Q = P(x_k|x_{k-1})$. To evaluate the stationary message probabilities you can

- write the equations you would use to solve the problem by hand, and find the solution numerically; or
- take any row of $Q^n$, where $n$ is a large integer (why?); or
- find the eigenvalues and eigenvectors of $Q^T$, take the eigenvector whose eigenvalue is 1, and divide it by the sum of the components (why?); or
- use any other method you can devise.

Verify the correctness of your program with simple examples that can be solved by hand.

Exercise 13. Find the entropy of a Markov source with binary alphabet $X = \{0, 1\}$ and the following transition matrix

$P(x_k|x_{k-1}) = \begin{bmatrix} 1/2 & 1/2 \\ 1/4 & 3/4 \end{bmatrix}$

For this source, evaluate the average length of a Huffman code for blocks of $N = 2$ and $N = 3$ symbols. Be careful, when you determine the probability of a block of symbols.
Exercise 14. Verify numerically that \( \frac{1}{L}H(X_1, \ldots, X_L) \geq H(X_L|X_1, \ldots, X_{L-1}) \) for \( L = 2, \ldots, 10 \) and the following Markov source: \( X = \{0,1\}; P(X_k = X_{k-1}) = 0.95 \). Remark: Do not solve this problem by hand, before getting your numerical results. Compare your results afterwards.

Exercise 15. Given the transition matrix of a (first-order) Markov source, write a computer program that evaluates the average length of a Huffman code for blocks of \( N \) symbols. Be careful, when you determine the probability of a block of symbols.

Exercise 16. In a channel with quaternary input and output, the output differs from the input with probability \( \varepsilon \). Conditioned on an error, each wrong symbol has probability \( \frac{1}{3} \). Evaluate the channel capacity.

Exercise 17. A channel with binary input \( \{0,1\} \) has ternary output \( \{0,e,1\} \), where \( e \) means that the input symbol is erased. The probability of error is \( \varepsilon \) and the probability of erasure is \( \rho \). Evaluate the channel capacity. Verify that your result agrees with the BEC and the BSC, setting \( \varepsilon = 0 \) and \( \rho = 0 \), respectively.

Exercise 18. The channel input is \( X = \pm 1 \) and the channel output is \( Y = X + N \), where \( N \) is independent of \( X \) and uniformly distributed in \((-2,2)\). Evaluate the channel capacity. Give an elementary explanation of this result. Hint: If \(-1 \leq Y \leq 1, \ldots;\) otherwise . . .

Exercise 19. A BSC with crossover probability \( \varepsilon \) is used to transmit the same bit twice. Thus we obtain a channel with binary input and quaternary output. Evaluate the capacity of this channel, and the capacity per channel use, and finally show that it is always less than the capacity of the BSC (but for \( \varepsilon \approx 1/2 \)).

Exercise 20. Evaluate the capacity of two BSC channels, with the same crossover probability \( \varepsilon \), in cascade. Why are we losing some capacity? Could we break the cascade in some way, and achieve the BSC capacity? (You must know the channel coding theorem, to answer the latter question).

Exercise 21. Find examples that show that \( I(X, Y|Z) \) can be greater or smaller than \( I(X, Y) \). For instance, consider various joint distributions of \( X = \{0,1\}, Y = \{0,1\}, \) and \( Z = \{0,1\} \).

Exercise 22. An asymmetric binary channel has crossover probability (i.e., probability of error) \( \varepsilon_0 = 10^{-4} \) when \( X = 0 \) and \( \varepsilon_1 = 10^{-2} \) when \( X = 1 \). Find (numerically) the optimal input distribution \( P(x) \) and the channel capacity \( C \). Then, let \( X \) have a uniform distribution, and evaluate the mutual information \( I(X,Y) \) (which must be less than \( C \), of course).

Exercise 23. The transition probabilities of a channel are

\[
P(y|x) = \begin{bmatrix}
1 & 0 & 0 & 0 \\
0 & 1 & 0 & 0 \\
0 & 0 & 3/4 & 1/4 \\
0 & 0 & 1/4 & 3/4
\end{bmatrix}
\]

Find the optimal input distribution and the channel capacity. Verify your results numerically. Evaluate the mutual information if the input distribution is uniform.
Exercise 24. Let \( Z = g(Y) \) be an invertible function. Show that \( I(X, Z) = I(X, Y) \).

Exercise 25. Let \( U = f(X) \) and \( Z = g(Y) \) be invertible functions. Show that \( I(U, Z) = I(X, Y) \).

Exercise 26. Reproduce Fig. 4.7 of the lecture notes.

Exercise 27. Let \( X = \{ \pm A, \pm 3A \} \) be the input of an AWGN channel with noise variance \( \sigma^2 \). Given \( \sigma_X^2 \), determine (numerically) \( A \), the optimal input distribution, and the channel capacity \( C \), i.e., the maximum of the constrained mutual information \( I(X, Y) \). Plot \( C \) versus the signal-to-noise ratio \( \sigma_X^2 / \sigma^2 \) (in dB). You should reproduce the results of the lecture notes (Fig. 4.2). You can assume \( \sigma^2 = 1 \) (convince yourselves that you can).

Exercise 28. Let \( X = \{ \pm A, \pm B \} \) be the input of an AWGN channel with noise variance \( \sigma^2 = 1 \). Let the probability distribution of \( X \) be uniform. Given \( \sigma_X^2 \), determine (numerically) the optimum levels \( A \) and \( B \) and the channel capacity \( C \), i.e., the maximum of the constrained mutual information \( I(X, Y) \). Plot \( C \) versus the signal-to-noise ratio (in dB).

Exercise 29. Given a BSC with crossover probability \( \varepsilon = 0.1 \), evaluate \( E_0(\rho) \) (analytically) and then the error exponent \( E(R) \) (numerically). Plot \( E(R) \) versus \( R \). By symmetry, you may assume that the input distribution is symmetric.

Exercise 30. \( X = \{ 0, 1 \} \) is the alphabet of a symmetric source, \( Y = \{ 0, 0?, 1?, 1 \} \) be the reproduction alphabet, and

\[
d(x, y) = \begin{bmatrix} 0 & 1 \\ 2 & 0 \end{bmatrix}
\]

is the distortion matrix. Find numerically the rate-distortion function \( R(D) \) and plot it. Compare your results with the rate-distortion function for the Hamming distortion.

Exercise 31. Let \( X = \{ 0, 1 \} \) be the alphabet of a symmetric source, let \( Y = \{ 0, 0?, 1?, 1 \} \) be the reproduction alphabet, and let

\[
d(x, y) = \begin{bmatrix} 0 & 1 & 5 & \infty \\ \infty & 5 & 1 & 0 \end{bmatrix}
\]

be the distortion matrix. Find numerically the rate-distortion function \( R(D) \) and plot it.

Exercise 32. Let \( X = \{ 0, 1, 2, 3 \} \) (with uniform distribution), \( Y = \{ 0, 3 \} \), and the distortion matrix be

\[
d(x, y) = |x - y| = \begin{bmatrix} 0 & 3 \\ 1 & 2 \\ 2 & 1 \\ 3 & 0 \end{bmatrix}
\]

Find (numerically) the rate-distortion function.

Exercise 33. A symmetric binary source is encoded as follows: 000, 100, 010, 001 are encoded into 000; 110, 101,011, 111 are encoded into 111. Thus, there are two codewords and the rate is 1/3. Evaluate the average distortion. Evaluate the inverse transition probability \( P(x = 0|y = 0) \). Show that \( P(x = 000|y = 000) \neq P(x = 0|y = 0)^3 \) and \( P(x = 100|y = 000) \neq P(x = 1|y = 0) \times P(x = 0|y = 0)^2 \). This means that the lossy encoder is not optimal. Finally, evaluate the minimum distortion corresponding to \( R = 1/3 \).
Exercise 34. (Taken from the 1959 Shannon paper; only if you know Hamming codes) Use Hamming codes with blocklength $N=7$, 15, 31, 63, and 127 to design lossy encoders for a symmetric binary source with rates $R=4/7$, 11/15, 26/31, 57/63, and 120/127, respectively. Evaluate the distortion, and compare it with the minimum distortion.

Exercise 35. Let a Gaussian random variable with variance $\sigma^2$ be quantized with just one bit. Find the optimal restitution levels, and show that the distortion is $0.363 \sigma^2$. You can assume $\sigma^2 = 1$ (convince yourselves that you can).

Exercise 36. Let a Gaussian random variable with variance $\sigma^2 = 1$ be quantized with two bits. Find the optimal quantization intervals and the optimal restitution levels. Then, evaluate the distortion. This problem can be solved only numerically.

Exercise 37. The output $Y$ of an AWGN channel with input $X = \pm 1$ and noise variance $\sigma^2$ is quantized with three bits. Let the eight intervals be $(-\infty, -3\Delta)$, $(-3\Delta, -2\Delta)$, $\ldots$, $(2\Delta, 3\Delta)$, $(3\Delta, \infty)$. Let $Z$ be the quantized output. Find (numerically) the value of $\Delta$ that maximizes the mutual information $I(X, Z)$. You need $P(z|x)$ and $P(z)$. Plot $I(X, Z)$ versus the signal-to-noise ratio $1/\sigma^2$ (in dB) and compare with the channel capacity. Repeat the evaluation for 16 quantization intervals.

Exercise 38. For channels with binary input and continuous output, coding theorists almost always use the channel Log-Likelihood-Ratios (LLRs) defined as

$$L_c = \log_e \frac{p(y|X = 0)}{p(y|X = 1)}$$

As to the input, if the source is not symmetric one can define the a priori LLR

$$L_a = \log_e \frac{P(X = 0)}{P(X = 1)}$$

so that, for instance, the a posteriori LLR is

$$L = \log_e \frac{P(X = 0|y)}{P(X = 1|y)} = \log_e \frac{p(y|X = 0)P(X = 0)}{p(y|X = 1)P(X = 1)} = L_c + L_a$$

For an AWGN channel with input $X = \pm 1$ and noise variance $\sigma^2$ show that $L_c = \frac{2}{\sigma^2} Y$.

Show that LLRs coming from independent sources of information can be added. For instance, let the same input $X$ be transmitted through two independent AWGN channels with different SNR. Show that the corresponding LLRs can be summed. Remark: Summing the channel outputs would be wrong; the sum of the LLRs is a weighted sum of the channel outputs.