The Cooper Union Department of Electrical Engineering
Prof. Fred L. Fontaine
ECE300 Communication Theory
Problem Set III: Information Theory
September 26, 2023

In this assignment, feel free to use MATLAB to help you with computations. Be sure to include the code with your submission.

1. The codeword lengths of a VLC are $\{2, 2, 3, 3, 5, 5, m\}$.
   (a) Use the Kraft inequality to find the smallest possible value for m. *Note:* Do not assume that the answer is ≥ 5 necessarily.
   (b) Assign the probability $p = 2^{-\ell}$ to each of the symbols where ℓ is the respective codeword length, *except* for one symbol with the LONGEST code length; assign that last value so that the probabilities sum to 1. With these values, compute the entropy $H(X)$ and compare with the average code length $\bar{L}$, and verify that $H(X) \leq \bar{L} < H(X) + 1$.

2. Consider a DMS with symbol probabilities:
$$\{0.25, 0.22, 0.18, 0.15, 0.06, 0.05, 0.05, 0.04\}$$
   (a) Construct a scalar Huffman code. You must show the code tree as well as the code table.
   (b) Compute $H(X)$ and $\bar{L}$.
   (c) Suppose the DMS output is transmitted using the Huffman code. Let $Y_1Y_2Y_3\cdots$ denote the sequence of the *binits* (not symbols) in the *coded* sequence.
      1. Find $P(Y_1Y_2)$ for all four cases $00, 01, 10, 11$.
      2. Find $P(Y_1Y_2Y_3)$ for all eight cases $000, 001, \cdots, 111$.
      3. Find $P(Y_3|Y_1Y_2)$ for all eight cases $0|00, 1|00, ..., 0|11, 1|11$.
   (d) Use your results to find $H(Y_3|Y_2Y_1)$.

3. Find the capacity of an AWGN channel with bandwidth $2MHz$, power $10W$, and noise power spectral density $N_0 = 10^{-8}W/Hz$.

4. The *binary erasure channel (BEC),* discussed in class, is similar to the BSC, except that if the received symbol is deemed "ambiguous" it is assigned an erasure symbol (E), rather than forcing a commitment to either 0 or 1. For purposes of decoding blocks of data bits, the erasure symbols can be thought of as "don't care" bits (i.e., the decoder has to work around these positions, labelled as unknown). Assume p_E is the probability of 0 or 1 being received as an erasure symbol, and p' the probability of $0 \to 1$ or $1 \to 0$.

   In many cases, $p' \ll p_E$, so we can obtain a simplified form of the BEC by having $0 \to E$ and $1 \to E$ each with probability p_E, and $0 \to 0$, $1 \to 1$ with probability $1 - p_E$ (the errors $0 \to 1, 1 \to 0$ never occur).

(a) Write the probability transition matrix for the simplified BEC.

(b) Assuming the input has $P(X = 0) = \pi_0$, so the input distribution is $\vec{p}_X = \begin{bmatrix} \pi_0 & 1 - \pi_0 \end{bmatrix}$, compute the output distribution $\vec{p}_Y = [P(Y = 0) \quad P(Y = 1)]$.

(c) Obtain a formula for $I(X, Y)$ for the simplified BEC. Simplify your answer to the form:

$$I(X, Y) = g_1(\pi_0)\, g_2(p_E)$$

i.e., a product of functions of each of the parameters, separately.

(d) Find the channel capacity, and the input distribution that achieves this capacity limit.

(e) Draw a graph of the channel capacity (in bits) as a function of p_E. [Using MATLAB]. How does this differ, if at all, with the channel capacity for the BSC?

5. **Water Filling Algorithm**

A Gaussian random vector with independent components has respective variances:

$$\sigma^2 = \begin{bmatrix} 4 & 3 & 0.8 & 0.1 \end{bmatrix}$$

Following the notation in the lectures, choice of a level λ for the water-filling algorithm determines a total distortion D (mean-square) and total rate R (in bits).

(a) Write code in MATLAB that, given the vector of variances and a level λ computes D and R for a Gaussian vector.

(b) Our goal is to achieve the minimum distortion subject to the constraint that $R \leq 2$. First, by trial and error, find levels λ_1, λ_2 that are integer multiples of 0.2 (they don't have to be consecutive, any spacing is fine, just don't make the spacing finer than 0.2; like 2.2 and 2.8 would be fine, if they work) that yield two rates R_1, R_2, one above 2 and one below 2. Then, by trial and error (you can try a dense set of λ between these values), find a proposed level λ that achieves $1.99 < R \leq 2$. I am *not* asking for a formal or systematic algorithm that iteratively converges to optimal λ; as I said, trial and error / trying many points in a loop is fine. You can "visually" observe your results to solve this part of the problem. The one part I do want you to code formally is part **a.**

(c) Graph the $R(D)$ curve traced out by points for levels chosen between λ_1, λ_2, and the line segment connecting the points $(D(\lambda_1), R(\lambda_1))$ and $(D(\lambda_2), R(\lambda_2))$ to confirm the convexity of the rate-distortion curve.

(d) Does increasing λ increase or decrease D? Increase or decrease R?

(e) For this λ, report not only the total values of R and D, but also the R_i, D_i for each of the vector components. Identify the components (if any) that are not encoded at all (i.e., have $R_i = 0$).