Homework 3

(Total 230 pts)

Due 5:00 pm, February 27, 2025 (Thursday)

Note: Submit electronically to Canvas two files ("HW3.pdf", and "Q1.m", "Q2.m", "Q3.m").

1. (80 pts) Load the following MAT file in Matlab,
   http://www.ece.uah.edu/~dwpan/course/ee614/data/data.mat,
   and estimate the entropies of the binary sequence "data".

   (a) Estimate the first-order entropy $H(X) = -\sum_{i=0}^{1} P(X=i)\log_2 P(X=i)$.

   (b) Estimate the second-order entropy of the sequence of symbol pairs (X_1, X_2):
   $$H(X_1, X_2) = -\sum_{i_1=0}^{1}\sum_{i_2=0}^{1} P(X_1 = i_1, X_2 = i_2)\log_2 P(X_1 = i_1, X_2 = i_2).$$

   (c) Suppose the binary sequence can be modeled by a discrete-time first-order Markov chain consisting of two states (S_0 and S_1). Estimate the following probabilities: $P(S_0)$, $P(S_1)$, $P(0|0)$, $P(1|0)$, $P(0|1)$, $P(1|1)$, based on which calculate the entropy of the Markov model H_M and the conditional entropy $H(X_2|X_1)$.

   (d) Fill in the following table with the results of (a), (b) and (c).

$H(X)$	
$H(X_1, X_2)$	
H_M	
$H(X_2\|X_1)$	
$P(S_0)$	
$P(S_1)$	
$P(0\|0)$	
$P(1\|0)$	
$P(0\|1)$	
$P(1\|1)$	

   (e) Submit on Canvas the Matlab script ("Q1.m") you used to obtain the above results.

2. (60 pts) Suppose that we have a source that emits symbols from an alphabet X that contains five distinct symbols {A, B, C, D, E}, with the following symbol probabilities:

P[A]	**P[B]**	**P[C]**	**P[D]**	**P[E]**
0.3	0.3	0.2	0.1	0.1

   (a) What is the entropy, $H(X)$, of the source X?

   (b) If we encode X using a minimum-variance Huffman code, what is the average codeword length? Sketch the Huffman code tree you constructed.

   (c) Suppose we encode a block of two symbols at a time. We assume that there is no dependence between the two symbols in a block. Then we have a new source Y that contains 25 symbols {AA, AB, AC, AD, AE, …, EA, EB, EC, ED, EE}. Calculate the entropy, $H(Y)$, of the source Y.

   (d) Write a Matlab script to code Y using a minimum-variance Huffman code. What is the average codeword length?

(e) Compare the results obtained in (b) and (d). What is your conclusion?

(f) Submit on Canvas the Matlab scripts ("Q2.m") you used to obtain the results.

(g) Fill in the following table with the results.

(a) $H(X)$	(b) Avg. codeword length of X	(c) $H(Y)$	(d) Avg. codeword length of Y

3. (90 pts) Huffman coding of digital images.
Use the *imread* function to read in the built-in color image 'landOcean.jpg' on Matlab. Next, extract the red (R), green (G) and blue (B) component images, respectively.
   (a) Display and attach the red component image R. What is the first-order entropy of the image R, $H(R)$?
   (b) Write a Matlab M file to code the image R using Huffman code. What is the size (bits) of the coded bitstream? And on average how many bits/pixel is needed to code the image?
   (c) Decode the bitstream obtained in (b). Use *tic* and *toc* functions to measure the elapsed time of decoding the coded bitstream. What is the running time (in seconds)? What is the version number of the Matlab you used?
   (d) Perform lossless check. Can your decoder recover the original image?
   (e) Repeat steps (a) through (d) using the green and blue component images, respectively.
   (f) Instead of Huffman coding the R, G, and B component images separately, we can apply Huffman coding on the entire color image (I) as a whole. What is the first-order entropy of the entire color image, $H(I)$? What is the size (bits) of the Huffman coded bitstream? And on average how many bits/pixel is needed to code the entire color image?
   (g) Which scheme offers higher overall compression on the entire color image, the separate R/G/B coding scheme, or the one-time Huffman coding scheme in (f)? Justify your answer.
   (h) Submit on Canvas the Matlab scripts ("Q3.m") you used to obtain the results.
   (i) Fill in the following table with your results.

(a) $H(R)$	(b) Size of the coded bitstream (bits)	(b) Avg. number of bits/pixel	(c) Matlab version	(c) Decoding time (sec)	(d) Lossless check passed?

$H(G)$	Size of the coded bitstream for G (bits)	Avg. number of bits/pixel for G	$H(B)$	Size of the coded bitstream for B	Avg. number of bits/pixel for B

$H(I)$	Size of the coded bitstream for I (bits)	Avg. number of bits/pixel for I	Which scheme offers higher overall compression? Check only one.	
			R/G/B ( )	One-time ( )