Homework 2

(Total 150 pts)

Due 5:00 pm, February 18, 2025 (Tuesday)

Note: Submit electronically to Canvas two files ("HW2.pdf", and "Q3.m").

1. (20 pts) A fair coin is flipped until the first tail occurs. Let X denote the number of flips required. Find the entropy $H(X)$ in bits.

2. (40 pts) X and Y are two discrete random variables drawing symbols from the alphabet A = {0, 1}. The joint distribution of $P(X, Y)$ is given as

X \ Y	0	1
0	1/3	1/3
1	0	1/3

Determine the following values:
(a) $H(X)$, and $H(Y)$; (b) $H(X|Y)$, and $H(Y|X)$; (c) $H(X,Y)$; (d) $H(Y) - H(Y|X)$;
(e) $I(X;Y)$; (f) K-L Distance $D_{KL}(X||Y)$; (g) K-L Distance $D_{KL}(Y||X)$;
(h) Cross Entropy $H_C(X,Y)$; (i) Cross Entropy $H_C(Y,X)$.
Fill in the table below with your answers.

$H(X)$	$H(Y)$	$H(X\|Y)$	$H(Y\|X)$	$H(X,Y)$	$H(Y) - H(Y\|X)$

$I(X;Y)$	$D_{KL}(X\|\|Y)$	$D_{KL}(Y\|\|X)$	$H_C(X,Y)$	$H_C(Y,X)$

3. (90 pts) Bit-plane processing and run-length coding of binary source.
Consider again input sequence in HW 1:
http://www.ece.uah.edu/~dwpan/course/ee614/data/coins_1d.mat
Since each symbol in the sequence is represented by 8 bits, we can extract eight bit-plane sequences from the input sequence. Assume that the first bit is the least significant bit (LSB), and the eighth bit is the most significant bit (MSB). As an illustrating example, let us consider a very short input sequence below, which consists of only 4 symbols (49, 50, 48, 49). The extracted bit-plane sequences are shown in the rightmost column in the table. We can see that the bit-plane sequences consist of two distinct symbols: 0 and 1.

Bit Position	**Symbol 1: 49**	**Symbol 2: 50**	**Symbol 3: 48**	**Symbol 4: 49**	**Bit-plane Sequence**
1 (LSB)	1	0	0	1	1001
2	0	1	0	0	0100
3	0	0	0	0	0000
4	0	0	0	0	0000
5	1	1	1	1	1111
6	1	1	1	1	1111
7	0	0	0	0	0000
8 (MSB)	0	0	0	0	0000

(a) From the input sequence:
http://www.ece.uah.edu/~dwpan/course/ee614/data/coins_1d.mat
extract the fourth bit-plane sequence, denoted as B4. Answer the following questions:
What is the probability of 0 (P0), and the probability of 1 (P1), respectively in the B4 sequence?
What is the entropy of the B4 sequence, denoted as H(B4)?
What is the value of the product, H(B4) × Length (B4), in bits?

(b) Next, let us consider run-length coding the bit-plane sequence. Since there are only two distinct symbols (either '0' or '1'), we can focus on determining merely the run-lengths of '0's. More specifically, the run-length of a stream of '0's is defined as the number of consecutive '0's followed by the first '1'. As an illustrative example, consider the following binary sequence consisting of 15 binary symbols:
"$\boxed{0\ 1}$ $\boxed{0\ 0\ 0\ 0\ 0\ 1}$ $\boxed{1}$ $\boxed{1}$ $\boxed{1}$ $\boxed{0\ 0\ 0\ 1}$". Scanning through the binary sequence will generate the following sequence of run-lengths of '0's:
"1, 5, 0, 0, 0, 3".
Note that if there is no '0' right before a single '1', then the run-length of '0's for this '1' will be zero (0). Besides, we do NOT consider run-lengths of '1's.

Now, scan through the B4 sequence obtained in (a) to generate the run-length sequence, denoted as "n_seq", where n stands for the run-lengths of '0's (before one '1').
Answer the following questions:
What is the average run-length, denoted as 'E[n]'?
What is the entropy of 'n_seq', denoted as H(n_seq)? Note, you might need to write your own script instead of using the built-in function "entropy" here.
What is the length of 'n_seq'? Denote this length as 'Length (n_seq)'.
What is the value of the product, H(n_seq) × Length (n_seq), in bits?

(c) Fill in the summary table below with your answers (failure to do so will result in points being deducted).

P0	P1	H(B4)	H(B4) × Length (B4)	E[n]	H(n_seq)	Length (n_seq)	H(n_seq) × Length (n_seq)

(d) Based on the results given in (c), which scheme provides higher compression in theory: direct coding on the B4 sequence, or coding on the run-length sequence 'n_seq'? What do you think would be the reason?

(e) Extract the eighth (MSB) bit-plane sequence, denoted as B8. Repeat (a) through (d), including populating the summary table for the B8 sequence.

(f) Which bit-plane sequence, B4 or B8, leads to higher compression (in theory) using run-length coding? What do you think would be the reason?

(g) Submit on Canvas the Matlab script ("Q3.m") you used to obtain the above results.