ECE 776 - Information theory
Midterm

Q1 (1 point). Given the channel \( p(y|x) \)

\[
p(y|x) = \begin{bmatrix}
1/3 & 2/3 & 0 \\
0 & 1/3 & 2/3 \\
2/3 & 0 & 1/3
\end{bmatrix},
\]
calculate the capacity.

**Sol.** The channel is symmetric since each row is a permutation of the other rows and the sums on each column are the same. Therefore, we have

\[
C = \log |Y| - H(1/3, 2/3, 0) = \log 3 - H(1/3) = 0.67 \text{ bits}.
\]

Q2 (1 point). Given two random variables \( X \) and \( Y \), assume that \( X \) is uniformly distributed in the set \( \mathcal{X} = \{1, \ldots, M\} \). Prove the following inequality that relates the mutual information \( I(X;Y) \) to the probability \( P[X = Y] \)

\[
I(X;Y) \geq P[X = Y] \log M - H(P[X = Y]).
\]

(Hint: use the Fano inequality)

**Sol.** The Fano inequality reads

\[
P[X \neq Y] \log M + H(P[X \neq Y]) = (1 - P[X = Y]) \log M + H(P[X = Y]) \\
\geq H(X|Y) = H(X) - I(X;Y)
\]

Thus

\[
I(X;Y) \geq H(X) - (1 - P[X = Y]) \log M - H(P[X = Y]) = H(X) - \log M + P[X = Y] \log M - H(P[X = Y]) \\
\geq P[X = Y] \log M - H(P[X = Y]),
\]

where the last inequality follows from the fact that \( H(X) - \log M \geq 0 \).

Q3 (1 point). You are given a random vector \( \mathbf{X} = (X_1, \ldots, X_n) \) of binary random variables \( X_i \in \{0, 1\} \). From this vector, we construct a new vector \( \mathbf{Y} \) which measures the run lengths of the symbols \( X_i \) as they occur. As an example, if \( \mathbf{X} \) reads \( \mathbf{X} = (0, 0, 0, 1, 1, 0, 1, 0, 0) \), we have \( \mathbf{Y} = (3, 2, 2, 1, 2) \), which provides the number of consecutive instances of the same symbol (0 or 1) in \( \mathbf{X} \). What is the relationship between \( H(\mathbf{X}) \) and \( H(\mathbf{Y}) \)? And between \( H(\mathbf{X}) \) and \( H(\mathbf{Y}, X_1) \)?

**Sol.** Since \( \mathbf{Y} \) is a function of \( \mathbf{X} \), we have

\[
H(\mathbf{X}) \geq H(\mathbf{Y}).
\]
Moreover, the mapping between $X$ and $(Y,X_1)$ is one-to-one, in fact knowing the run-lengths and the initial symbol fully specifies $X$. It follows that

$$H(X) = H(Y,X_1).$$

**Q4** (1 point). Given an iid sequence $X_i, i = 1, 2, ..., $ it is known that a rate of $H(X)$ bits per symbol is enough to describe the source. Say now that a second iid source $Y_i$, "correlated" with $X_i$, is available at the decoder. It can be proved that in this case only $H(X|Y)$ bits per symbol are necessary. Assuming that $X_i$ and $Y_i$ are $Ber(0.5)$ and that $P[X_i \neq Y_i] = p$, find $H(X)$ and $H(X|Y)$. What happens if $p = 0$? If $p = 1$? If $p = 0.5$? Interpret the results.

**Sol.:** The entropy is easily calculated as $H(X) = 1$. The conditional entropy reads

$$H(X|Y) = p_Y(0)H(X|Y = 0) + p_Y(1)H(X|Y = 1) = 0.5 \cdot H(p) + 0.5 \cdot H(p) = H(p).$$

If $p = 0$ or $p = 1$, based on the knowledge of $Y_i$, the decoder can immediately reconstruct $X_i$ without any further information, thus we have $H(X|Y) = 0$. On the other hand, if $p = 0.5$, the information about $Y_i$ is useless in reconstructing $X_i$ and $H(X|Y) = H(X) = 1$.

**P1** (2 points). A small radar transmits an electromagnetic pulse and listens for a possible echo returned by a target object. The small radar simply sets a threshold on the power of the received signal: if the received power is above the threshold, then an echo is detected and the radar outputs 1 (that is, target detected); otherwise it outputs 0 (target not present). Because of noise, the radar can make erroneous decisions. In order to obtain a more accurate detection, a series of $n$ pulses is transmitted.

Mathematically, let us assume that the target is present or not with probability 50%. If the target is present, the $n$ outputs of the radar are given by the iid random process $X_{11}, X_{12}, ..., X_{1n}$ with $X_{1i} \sim Ber(0.8)$. This means that with probability 0.8 the detector is able to correctly detect the target for each transmitted pulse. Conversely, if the target is not present, the $n$ outputs of the radar are given by the iid random process $X_{21}, X_{22}, ..., X_{2n}$ with $X_{1i} \sim Ber(0.2)$, meaning that the probability of false alarm is 0.2.

We want to study the random process given by the radar outputs, say $Y_i$.

(a) Is $Y_i$ stationary? Is it iid?

**Sol.:** Yes, the statistics of the process do not change with time (see also class notes).

(b) Evaluate the entropy rate $H(Y)$ of $Y_i$.

**Sol.:**

$$H(Y) = \lim_{n \to \infty} \frac{H(Y^n)}{n} = \frac{1}{2}H(0.8) + \frac{1}{2}H(0.2),$$

see class notes for details.

(c) Can we apply the AEP to $Y_i$ (that is, is it true that $-\frac{1}{n}\log p(Y^n) \to H(Y)$)? Explain.
Sol.: No, we cannot apply the AEP because the sequence is not ergodic. To see this, let us calculate
\[ -\frac{1}{n} \log p(Y^n) \rightarrow \begin{cases} H(0.8) & \text{w.p. 0.5} \\ H(0.2) & \text{w.p. 0.5} \end{cases}. \]

(d) Assume that we need to send sequence \( Y \) to a controller. Is there a code that achieves an average description length \( L_n \) such that \( \frac{L_n}{n} \rightarrow H(Y) \)? If yes, find such code.

Sol.: Let us use Huffman or Shannon-Fano coding on sequence \( Y^n \). We obtain an average code length \( L_n \) that satisfies
\[ \frac{H(Y^n)}{n} \leq \frac{L_n}{n} < \frac{H(Y^n) + 1}{n}, \]
so that for \( n \rightarrow \infty \) \( \frac{L_n}{n} \rightarrow H(Y) \).

P2 (2 points). A source \( X \) is characterized by the pmf
\[ p(x) = \begin{cases} 0.7 & x = 1 \\ 0.2 & x = 2 \\ 0.1 & x = 3 \end{cases}. \]

(a) Find the codeword lengths for the binary Huffman code and for the Shannon-Fano code.

Sol.: It is easy to see that for Huffman, we have lengths \( (\ell_1, \ell_2, \ell_3) = (1, 2, 2) \) while for Shannon-Fano \( (\ell_1, \ell_2, \ell_3) = ([− \log 0.7] , [− \log 0.2] , [− \log 0.1]) = (1, 3, 4). \)

(b) Calculate the entropy of the source and compare it with the average lengths of the two codes at the previous point.

Sol.: The entropy reads
\[ H(X) = -0.7 \log 0.7 - 0.2 \log 0.2 - 0.1 \log 0.1 = 1.157 \text{ bits}, \]
while the average codeword lengths are
\[ L(C_{Huffman}) = 0.7 \cdot 1 + 0.2 \cdot 2 + 0.1 \cdot 2 = 1.3 \text{ bits} \]
\[ L(C_{Shannon}) = 0.7 \cdot 1 + 0.2 \cdot 3 + 0.1 \cdot 4 = 1.7 \text{ bits}. \]

(c) Can the average length of any code be smaller than 1.15 bits per symbol? Can the average length of a symbol-by-symbol code be smaller than 1.3 bits per symbol?

Sol.: Since \( H(X) = 1.157 \text{ bits} \), it is not possible for any code to achieve \( L(C) = 1.15 \text{ bits/symbol} \). Moreover, among symbol-by-symbol codes, we know that Huffman codes minimize the average length, therefore it is not possible to achieve \( L(C) < L(C_{Huffman}) = 1.3 \text{ bits} \).
(b) Considering alphabets other than binary, say $D$-ary alphabets, what is the smallest integer $D$ such that the average length for the Shannon-Fano code equals the average length for the Huffman code?

**Sol.**: With $D > 2$, $L(\mathcal{C}_{\text{Huffman}}) = 1$. However, this is not true for Shannon-Fano codes. Imposing that the longest codeword for Shannon codes is 1, we obtain the desired condition: $[−\log_D0.1] = 1$. Therefore, we get that the smallest integer $D$ such that the average length for the Shannon-Fano code equals the average length for the Huffman code is $D = 10$.

**P3** (2 points). Consider a memoryless channel that takes pairs of bits as input and produces two bits as output as follows: \(00 \rightarrow 01\), \(01 \rightarrow 10\), \(10 \rightarrow 11\), \(11 \rightarrow 00\) (to read: input $\rightarrow$ output).

Let $X_1X_2$ denote the two input bits and $Y_1Y_2$ the two output bits.

(a) Calculate the mutual information $I(X_1X_2; Y_1Y_2)$ for a given joint pmf of the four pairs of input bits (define: $p_1 = P[X_1 = 0, X_2 = 0]$, $p_2 = P[X_1 = 0, X_2 = 1]$, $p_3 = P[X_1 = 1, X_2 = 0]$ and $p_4 = P[X_1 = 1, X_2 = 1]$).

**Sol.**: With the definitions, we have

$$I(X_1X_2; Y_1Y_2) = H(X_1X_2) = H(p_1, p_2, p_3, p_4).$$

(b) Show that the capacity is 2 bits.

**Sol.**: The capacity is (define $p = (p_1, p_2, p_3, p_4)$)

$$C = \max_p I(X_1X_2; Y_1Y_2) = 2,$$

with optimal pmf $p = u = (1/4, 1/4, 1/4, 1/4)$.

(c) Show that, surprisingly, $I(X_1, Y_1) = 0$ for the capacity-maximizing distribution of the input (that is, information is transferred by considering both bits). (*Hint*: Find the joint PMF of $X_1$ and $Y_1$)

**Sol.**: We need to obtain the joint pmf of $X_1$ and $Y_1$ under the assumption that $p = u = (1/4, 1/4, 1/4, 1/4)$. We start from the joint pmf of pairs of bits:

\[
\begin{array}{c|ccccc}
X_1X_2 \backslash Y_1Y_2 & 00 & 01 & 10 & 11 \\
00 & 0 & 1/4 & 0 & 0 \\
01 & 0 & 0 & 1/4 & 0 \\
10 & 0 & 0 & 0 & 1/4 \\
11 & 1/4 & 0 & 0 & 0
\end{array}
\]

From this we easily obtain:

\[
\begin{array}{c|c}
X_1 \backslash Y_1 & 0 & 1 \\
0 & 1/4 & 1/4 \\
1 & 1/4 & 1/4
\end{array}
\]

so that $X_1$ and $Y_1$ are independent and $I(X_1, Y_1) = 0$. 
