Huffman Codes

1. Imagine you are given a random variable $X$ to compress, where $X = \{A, C, G, T\}$. $X$ is distributed according to:

$$X = \begin{cases} 
  A, & \text{w.p. } p_A = \frac{1}{2} \\
  C, & \text{w.p. } p_C = \frac{1}{4} \\
  G, & \text{w.p. } p_G = \frac{1}{8} \\
  T, & \text{w.p. } p_T = \frac{1}{8} 
\end{cases}$$

You decide to use a Huffman code to compress the sequence.

(a) Specify a binary Huffman code for $X$.  

(b) Compute the average length $E[L]$ in bits per symbol of the compressed sequence and compare it to the entropy of the source $H(X)$.  

(c) Let $X = X_1, X_2, \cdots$ be a memoryless source with $X_i \sim X$. Let $C(X_i)$ denote the binary Huffman codeword corresponding to $X_i$ using the Huffman code from Part (a). Let $B = C(X_1)C(X_2)\cdots$ denote the binary process formed by concatenating the codewords corresponding to the source $X$. What is the distribution of the resulting binary process $B$?

Solution:

(a) Create a binary Huffman code.

One possible Huffman code is:

<table>
<thead>
<tr>
<th>$x_i$</th>
<th>$p_i$</th>
<th>$c_i$</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>1/2</td>
<td>0</td>
</tr>
<tr>
<td>C</td>
<td>1/4</td>
<td>10</td>
</tr>
<tr>
<td>G</td>
<td>1/8</td>
<td>110</td>
</tr>
<tr>
<td>T</td>
<td>1/8</td>
<td>111</td>
</tr>
</tbody>
</table>

(b) The average length in bits per symbol is given by:

$$E[L] = 1 \times (1/2) + 2 \times (1/4) + 3 \times (1/8) + 3 \times (1/8) = 7/4$$

By noticing that the probabilities of the source are dyadic, you can conclude without any computation that $E[L]$ equals to the entropy of the source $H(X)$.

$$H(X) = (1/2) \log 2 + (1/4) \log 4 + (1/8) \log 8 + (1/8) \log 8 = 7/4$$
(c) First, note that the source $X$ is i.i.d. Now, let $B = B_1B_2 \cdots$ where $B_i \in \{0, 1\}$. Consider the probability $P(B_{i+1}|B_1, B_2 \cdots B_i)$. Since the decoder is following the Huffman tree described in (a), knowing the first $i$ bits, will determine the exact position in the Huffman tree (since the Huffman code is instantaneous). Now there are two possibilities

- You are at the end of a codeword. The next codeword will begin with a 0 or a 1 with equal probability.
- You are at an internal node in the decoding tree. In this case (from the tree construction) it is easy to see that next bit in the codeword is equally likely to be a 0 or a 1.

Thus, regardless of what the first $i$ bits were, the $(i + 1)^{th}$ bit can be 0 or 1 with equal probability. Thus,

$$P(B_{i+1} = 1|B_1, B_2 \cdots B_i) = P(B_{i+1} = 1) = \frac{1}{2}.$$

Thus, the $B$ process is in fact i.i.d. Bernoulli($\frac{1}{2}$). The codeword bits appear to be fair coin flips!

2. The following claim is sometimes found in the literature:

*It can be shown that the length of the binary Huffman codeword of a symbol $a_i$ with probability $p_i$ is always less than or equal to $\lceil -\log_2 p_i \rceil$*

Is this claim true? Justify your answer \textbf{5 points}

**Solution:**
This claim is not true in general! To prove it, we shall construct a counterexample. Choose a small number $\epsilon$. Consider the following source $X$ with 4 symbols:

$$X = \begin{cases} 
  x_1, & \text{w.p. } p_1 = \epsilon \\
  x_2, & \text{w.p. } p_2 = \frac{1}{3} - \epsilon \\
  x_3, & \text{w.p. } p_3 = \frac{1}{3} - \epsilon \\
  x_4, & \text{w.p. } p_4 = \frac{1}{3} + \epsilon 
\end{cases}$$

Now consider the Huffman tree corresponding to the above source. If we combine $x_1$ and $x_2$ in the first step, the code will have length functions: $l(x_1) = l(x_2) = 3$, $l(x_3) = 2$ and $l(x_4) = 1$. Now for $x_2$, the Shannon code length would be $\lceil -\log_2 \frac{1}{3} \rceil = 2$, for $\epsilon < \frac{1}{12}$. Thus, the binary Huffman codeword length can exceed the Shannon codeword length in general for a symbol.

3. Consider a source of $K$ symbols, with $p_1 \geq p_2 \geq \ldots \geq p_K$. Find the largest $q$ s.t. $p_1 < q$ implies $l_i > 1$, where $l_i$ is the length of the binary Huffman codeword associated with symbol $i$. \textbf{10 points}

**Solution:**
First notice that this question makes sense only for $K \geq 4$, since for other choices of
K we have $l_1 = 1$.
For $K = 4$, the first super symbol is created by merging $p_3$ and $p_4$. $l_1 > 1$ if $p_3 + p_4 > p_1$, since in this case the next super symbol will be created by joining $p_1$ and $p_2$, ensuring $l_1 > 1$. Thus, we want to make $p_3, p_4$ as small as possible, while satisfying $p_3 + p_4 > p_1$.
To minimize $p_3 + p_4$, we maximize $p_2$, which gives $p_2 = p_1$ and $p_3 + p_4 = 1 - 2p_1$. Solving the inequality we get

$$p_1 < 1 - 2p_1 \Rightarrow p_1 < \frac{1}{3}$$

We generalize the result for arbitrary $K$ using induction. Assume that $p_1 < \frac{1}{3}$ implies $l_1 > 1$ for a source having $K'$ symbols. Now consider a source with $(K' + 1)$ source symbols. In the first step, we merge $p_k$ and $p_{k+1}$. Now there are two cases. If $p_k + p_{k+1} > p_1$, then $l_1 > 1$ automatically. On the other hand, if $p_k + p_{k+1} < p_1$, we are in the scenario of $K'$ symbols, where we know that $p_1 < \frac{1}{3}$ implies $l_1 > 1$ by the induction hypothesis.

### Jensen Shannon Divergence

30 points

Let $P$ and $Q$ be discrete probabilities on the alphabet $\{1, 2, \cdots, K\}$. Relative entropy $D(P||Q)$, can be viewed as a distance between probability distributions. In general it is not symmetric, i.e., there exist $P$ and $Q$ for which $D(P||Q) \neq D(Q||P)$. We can construct symmetric distance measures between two distributions based on relative entropy; one such measure is called the Jensen-Shannon Divergence, which is defined as

$$JS(P||Q) = \frac{1}{2} D \left( P \mid\mid \frac{P + Q}{2} \right) + \frac{1}{2} D \left( Q \mid\mid \frac{P + Q}{2} \right).$$

It is easy to check that $JS(P||Q) = JS(Q||P)$.

1. Let $Z$ be a Bernoulli($\frac{1}{2}$) random variable and let $X$ be a random variable drawn according to $P$ if $Z = 1$, and according to $Q$ if $Z = 0$. Express $JS(P||Q)$ in terms of $I(X; Z)$.

   20 points

2. What is the maximum value of $JS(P||Q)$ among all possible distributions $P, Q$. [Hint: Use your answer to 1.]

   Solution:

   1. Correct Answer is $I(X; Z) = JS(P||Q)$.
Let $M = \frac{P+Q}{2}$. Note that $X$ is distributed according to $M$ by definition.

\[
I(X; Z) = H(X) - H(X|Z)
\]

\[
= \sum -M_i \log M_i - \frac{1}{2} \left( \sum -P_i \log P_i + \sum -Q_i \log Q_i \right)
\]

\[
= \sum -\frac{P_i}{2} \log M_i + \sum -\frac{Q_i}{2} \log M_i - \frac{1}{2} \left( \sum -P_i \log P_i + \sum -Q_i \log Q_i \right)
\]

\[
= \sum \frac{P_i}{2} (\log P_i - \log M_i) + \sum -\frac{Q_i}{2} (\log Q_i - \log M_i)
\]

\[
= \frac{1}{2} \sum P_i \log \frac{P_i}{M_i} + \frac{1}{2} \sum Q_i \log \frac{Q_i}{M_i}
\]

\[
= \frac{1}{2} D(P||M) + \frac{1}{2} D(Q||M)
\]

\[
= JS(P||Q)
\]

An alternate proof, which exploits the representation of mutual information in terms of relative entropy is as follows:

\[
I(X; Z) \overset{(a)}{=} D(P_{X,Z}||P_X P_Z)
\]

\[
= D(P_{X|Z}||P_X)
\]

\[
= \mathbb{E} \left[ \log \frac{P_{X|Z}}{P_X} \right]
\]

\[
\overset{(b)}{=} \mathbb{E} \left[ \log \frac{P_{X|Z=z}}{P_X} | Z = z \right]
\]

\[
\overset{(c)}{=} \frac{1}{2} D(P||M) + \frac{1}{2} D(Q||M)
\]

\[
= JS(P||Q).
\]

In the above, (a) follows by definition of mutual information. (b) follows by tower property of expectation. (c) follows from the fact that $P_{X|Z=1} = P$ and $P_{X|Z=0} = Q$, and $P_Z(0) = P_Z(1) = \frac{1}{2}$.

2. Correct answer is 1.

Recall from part 1, that

\[
JS(P||Q) = I(X; Z) = H(Z) - H(Z|X) \leq H(Z) = 1.
\]

**Entropic Sources**

35 points

Consider a (not necessarily memoryless) source $Z$. Let $H(Z^n)$ denote the entropy of an $n$-tuple from this source.

**Definition 1.** A source $Z$ is said to be **subentropic** if the total mass of its most likely $\left\lfloor 2^{(1+\delta)H(Z^n)} \right\rfloor$ outcomes goes to 1 as $n \to \infty$ for any $\delta > 0$.  
Definition 2. A source is **superentropic** if the total mass of its most likely $\lfloor 2^{(1+\delta)H(Z^n)} \rfloor$ outcomes does not go to 1 as $n \to \infty$ for any $\delta > 0$.

Definition 3. A source is **entropic** if it is both **subentropic** and **superentropic**.

Answer the following questions:

I. Any source is either subentropic or superentropic or both.  **True** or **False**?  5 points

II. For each of the following sources, classify the source as
   (i) subentropic but not superentropic
   (ii) superentropic but not subentropic
   (iii) entropic
   (iv) none of the above.

   1. $Z^n = (X_1, \cdots, X_n)$ is an i.i.d. source with finite alphabet.  10 points

   2. For $0 < q < 1$,  10 points

      $$P[Z^n = z^n] = \begin{cases} 1 - q & \text{if } z^n = (0, \cdots 0) \\ q/(2^n - 1) & \text{if } z^n \neq (0, \cdots 0). \end{cases}$$

   3. $Z^n$ has four types of outcomes  10 points
      - one mass with probability $\frac{1}{2}$;
      - $2^n$ masses each with probability $(\frac{1}{2} - \frac{1}{n})2^{-n}$;
      - $\lfloor \frac{1}{n}2^{n^2/2} \rfloor$ masses each with probability $2^{-n^2/2}$;
      - one mass with the remaining probability if $\frac{1}{n}2^{n^2/2}$ is not integer valued.

Solution

I. Note that if a source is not subentropic, then the probability of a set of size $\lfloor 2^{(1+\delta)H(Z^n)} \rfloor$ is not converging to one. Then, the mass of an exponentially smaller set of size $\lfloor 2^{(1-\delta)H(Z^n)} \rfloor$ cannot go to one either. Thus the source has to be superentropic by definition. Vice versa is also true i.e. a source that is not superentropic has to be subentropic. Therefore, the correct answer is **True**.

II. 1. Correct answer is (c).

   Note that since the $Z_i$'s are i.i.d. on a finite alphabet, the Asymptotic Equipartition Law applies directly. First, we note that $H(Z^n) = nh(Z_1)$. Now we verify that the source is subentropic. Fix a $\delta > 0$. Choose any $\epsilon > 0$ such that $nH(Z_1)(1 + \delta) > n(H(Z_1) + \epsilon)$. Consider the typical set $A^{(n)}$. Clearly,

   $$P(\text{most likely } 2^{(1+\delta)H(Z^n)} \text{outcomes}) \geq P(A^{(n)}).$$

   Since $P(A^{(n)}) \to 1$, we observe that the source is subentropic by Definition 1.

   Similarly, we have from the AEP that any set that is exponentially smaller than the
typical set will have vanishing probability as \( n \to \infty \). Since the probability does not converge to 1, the source is superentropic by Definition 2. In fact, in this case we have the stronger result that

\[
P(\text{most likely } \lceil 2^{(1-\delta)H(Z^n)} \rceil \text{ outcomes}) \to 0.
\]

Therefore, the source is entropic.

2. The correct answer is (b).

Note for the given source that,

\[
H(Z^n) = h(q) + qn + q\log(1 - 2^{-n}),
\]

where \( h(\cdot) \) is the binary entropy function. Let us consider an event whose size grows as \( \alpha n \) where \( \alpha < 2 \). Then the probability of this event will not converge to 1 as \( n \to \infty \). To see this note that the probability of such a set will behave like

\[
1 - q + \frac{a^n - 1}{2^n - 1} q \text{ which converges to } 1 - q.
\]

In this case we have \( \alpha = 2^q < 2 \).

Thus, the source is not subentropic (and hence is superentropic), making (b) the correct choice.

3. The correct answer is (c).

Recall that the source has four outcomes:

(a) one mass with probability \( \frac{1}{2} \);
(b) \( 2^n \) masses each with probability \( \left(\frac{1}{2} - \frac{1}{n}\right)2^{-n} \);
(c) \( \lfloor \frac{1}{n}2^{n^2/2} \rfloor \) masses each with probability \( 2^{-n^2/2} \);
(d) one mass with the remaining probability if \( \frac{1}{n}2^{n^2/2} \) is not integer valued.

We can compute the entropy \( H(Z^n) \), for large \( n \) to be

\[
H(Z^n) = \frac{1}{2} \log 2 + \left(\frac{1}{2} - \frac{1}{n}\right) \log(2^n(1/2 - 1/n)^{-1}) + \frac{1}{n} \log \frac{1}{2^{n^2/2}}
\]

\[
= \frac{1}{2} + \left(\frac{1}{2} - \frac{1}{n}\right)n + \left(\frac{1}{2} - \frac{1}{n}\right)\log((1/2 - 1/n)^{-1}) + \frac{n}{2} - \frac{1}{n} \log n
\]

\[
= n + o(n).
\]

Note that the rest of the \( o(n) \) terms are sublinear, and do not contribute to the largest exponent in the size of the set, which in this case is \( n \). Thus, a set of size \( 2^{n(1+\delta)} \), consisting of the most likely outcomes of \( Z^n \), will include the masses in (a) and (b). The total mass of this set will exceed \( \frac{1}{2} + \left(\frac{1}{2} - \frac{1}{n}\right) = 1 - \frac{1}{n} \). This clearly goes to 1 as \( n \to \infty \). Thus, the source is subentropic by Definition 1.

Now consider a set of size \( 2^{n(1-\delta)} \). It will include the mass in (a) and a fraction of the masses in (b). In particular, the total probability of such a set would not exceed \( \frac{1}{2} + \left(\frac{1}{2} - \frac{1}{n}\right)2^{-n\delta} \) which converges to \( \frac{1}{2} \) as \( n \to \infty \). Thus the source is superentropic by Definition 2. Hence the answer is (c) by Definition 3.